Hoogsteen versus Reversed-Hoogsteen Base Pairing: DNA Triple Helices

Yuen-Kit Cheng and B. Montgomery Pettitt*,†

Contribution from the Chemistry Department, University of Houston, Houston, Texas 77204-5641. *Received October 28, 1991*

Abstract: Recently, oligonucleotides have been shown to inhibit transcription in genes by triple helix or triplex formation in vitro and in vivo. A better understanding of the forces that stabilize triplex structures will be important in developing applications of this method of genetic medication to arbitrary sequences. Therefore, base pairings and strand orientations for homogeneous $d(T\cdot A\cdot T)_{27}$ and $d(C\cdot G\cdot G)_{27}$ triplexes were examined. The method was extended to triplex models formed by c-*myc* gene promoter region and complementary oligonucleotides. Templates of a single plane with three bases were constructed and used in a simple geometric replication based on experimental geometric parameters. Minimizations and quenched molecular dynamics simulations were performed on the model systems. The estimated accessibility of the major groove for countercation coordination was obtained by calculating the effective accessible surface areas of backbone phosphate oxygen atoms. Free energy calculations of the solvation/desolvation penalty on single strands, duplexes, and the stereoisomeric triplexes have been performed. They were combined with corresponding enthalpic terms so that the results could be discussed in a more realistic aspect. For $d(T\cdot A\cdot T)_{27}$ triplexes, results from both the internal potential energy and solvation free energy calculations contribute to the experimentally known base pairing and strand orientation. Solvation is found to determine the strand orientation for $d(C\cdot G\cdot G)_{27}$ triplexes with either Hoogsteen or reversed-Hoogsteen base pairing between the two purine strands being possible. We have compared the $d(C\cdot G\cdot G)_{27}$ triplexes by computing the hydrogen–hydrogen distances which may be useful in verifying these models by future NMR/NOE studies. Using these homopolymers the orientation of oligonucleotides bound to the c-*myc* gene promoter site is shown to also be dominated by the forces of solvation.

I. Introduction

The association between complementary deoxyribonucleic acid (DNA) strands in aqueous solution yields a double helix (duplex) with additional hydrogen-bonding donors and acceptors directed along the major groove. These donors and acceptors are exposed to the surrounding environment and are able to interact with specific binding molecules, such as proteins, to form a specific complex or other DNA molecules to form a triple helix (triplex). The discovery of a triple ribonucleic acid helix dates to 1957.[1] Recently, triplex formation has been studied intensively.[2-32] Numerous experimental techniques have been employed to study triplexes, including gel electrophoresis,[2-21] nuclear magnetic resonance (NMR) spectroscopy,[2,22-26] and spectrophotometry.[2,4,23,24,27-30] These experiments have provided data on binding constants, specific base-pairing patterns, and thermodynamic properties. However, unlike the large amount of single-crystal X-ray structural studies on duplexes [see ref 33 and references therein], only fiber diffraction data of the triplexes poly(U·A·U),[34] poly(I·A·I),[35] and poly(dT·dA·dT)[36] have been obtained and refined. (A, T, G, C, U, and I denote the nucleic acid bases adenine, thymine, guanine, cytosine, uridine, and inosine, respectively. Prefix "d" stands for deoxyribo in nucleotide or nucleoside nomenclature.)

Many of the recent theoretical structural studies on nucleic acids deal with DNA–drug interactions[25,37] and local properties due to heterogeneity and complexity of sequences, such as bending caused by pyrimidine–purine or purine–pyrimidine junctions.[38-40] The global structural features of triplex formation have attracted much less attention.[32] Although intramolecular DNA triple helices have been known for some time,[4,8,9,16,17,19] inhibition of gene promoter sites by triplex formation in vitro has only been observed recently.[3,18]

While a large number of designed triplexes exist,[7,11-15,28,31] their exact three-dimensional structures are certainly dependent on the particular sequences studied. The earliest fiber diffraction studies on poly(dT·dA·dT), denoted by TAT throughout, showed the third or triplex-forming strand is oriented parallel to the purine strand.[36] In experimental studies, oligomers predominantly with protonated cytosines (C^+) in acidic pH solutions have been shown to also bind parallel to the G-rich strand of CG duplexes.[11,13,21] However, it has recently been shown that anti-parallel is the preferred orientation when G-rich synthetic oligonucleotides bind to the pu-

* Address correspondence to this author.
† Alfred P. Sloan Fellow 1989-1991.

(1) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.
(2) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081-6087.
(3) Durland, R. H.; Kessler, D. J.; Duvic, M.; Hogan, M. E. In *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*; Pullman, A., Ed.; 1990; 565-578.
(4) Chen, F.-M. *Biochemistry* **1991**, *30*, 4472-4479.
(5) Lyamichev, V. I.; Voloshin, O. N.; Frank-Kamenetskii, M. D.; Soyfer, V. N. *Nucl. Acids Res.* **1991**, *19*, 1633-1638.
(6) Orson, F. M.; Thomas, D. W.; McShan, W. M.; Kessler, D. J.; Hogan, M. E. *Nucl. Acids Res.* **1991**, *19*, 3435-3441.
(7) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363.
(8) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucl. Acids Res.* **1990**, *18*, 6621-6624.
(9) Fox, K. R. *Nucl. Acids Res.* **1990**, *18*, 5387-5391.
(10) Bianchi, A.; Wells, R. D.; Heintz, N. H.; Caddle, M. S. *J. Biol. Chem.* **1990**, *265*, 21789-217.
(11) Strobel, S. A.; Derven, P. B. *Science* **1990**, *249*, 73-75.
(12) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435-2437.
(13) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971.
(14) Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 7286-7287.
(15) Francois, J.-C.; Saison-Behmoaras, T.; Hélène, C. *Nucl. Acids Res.* **1988**, *16*, 11431-114.
(16) Hanvey, J. C.; Shimizu, M.; Wells, R. D. *J. Biol. Chem.* **1989**, *264*, 5950-5956.
(17) Hanvey, J. C.; Shimizu, M.; Wells, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6292-6296.
(18) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456-459.
(19) Kohwi, Y.; Kohwi-Shigematsu, T. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3781-3785.
(20) Lyamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D.; Cantor, C. R. *Nucl. Acids Res.* **1988**, *16*, 2165-2178.
(21) Moser, H.; Dervan, P. B. *Science* **1987**, *238*, 645-650.
(22) Mooren, M. M. W.; Pulleyblank, D. E.; Wijmenga, S. S.; Blommers, M. J. J.; Hilbers, C. *Nucl. Acids Res.* **1990**, *18*, 6523-6529.
(23) Kan, L.-S.; Calahan, D. E.; Trapane, T. L.; Miller, P. S.; Ts'o, P. O. P.; Huang, D. H. *J. Biomol. Struct. Dyn.* **1991**, *5*, 911-933.
(24) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1942-1946.
(25) Umemoto, K.; Sarma, M. H.; Gupta, G.; Luo, J.; Sarma, R. H. *J. Am. Chem. Soc.* **1990**, *112*, 4539-4545.

0002-7863/92/1514-4465$03.00/0

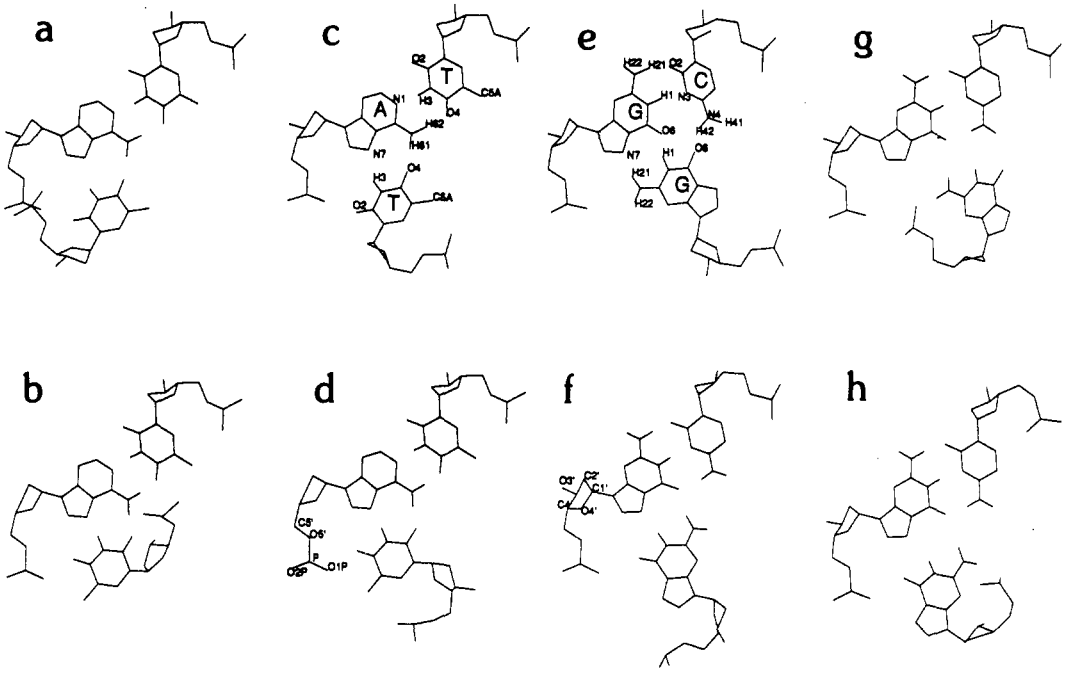

Figure 1. The eight possible TAT and CGG triplets. All triplets have identical relative position of strands I, II, and III. In each panel, the top right, middle, and bottom nucleotides belong to strands I, II and III, respectively: (a–d) for TAT and (e–h) for CGG triplets. All triplets except those of (a) and (b) are extracted from minimized triplexes. Base types are illustrated in (c) and (e). Standard Watson–Crick pairs are formed between strands I and II for all triplets. Top panels are triplets with H pairs between strands II and III, while those of the bottom panels are **R** pairs. Panels (a) and (b) give s dihedrals for strand III of TAT, which render the backbone atoms of strand III either overlapping with the backbone of strand II [as in (a)] or too close to the major groove [as in (b)]. These models are probably not stable and are omitted in the energetic calculations. Panels (c) and (d) depict the t isomers of TAT triplets which were adopted in our calculations. Panels (e) and (f), with t dihedrals, give CGG **pHt** and CGG **aRt**, respectively. Panels (g) and (h) are the s isomers [(g) gives CGG **aHs**]. Panels (f) and (g) depict triplets with strands II and III running **a**, which agrees with experiments in (c-*myc*)-TFO triplexes.[3,7] Relevant atom labels are given in panels (c)–(f); see also refs 43 and 44 for details of nucleic acid nomenclature.

rine-rich strand of several duplex gene promoter regions.[3,7,41] The highly-biased CG composition of these promoter regions suggests that the preference in the orientation is probably a property of recognition, due to the specific base composition of both the target duplex and the triplex forming oligonucleotide (TFO). The second pyrimidine strand in the TAT complex binds parallel to the purine strand of the Watson–Crick (W–C) pair through a Hoogsteen base-pairing scheme.[36,42] Because of the multiplicity of hydrogen-bonding groups, reversed-Hoogsteen patterns also exist. No refined X-ray structural data on poly(dC·dG·dG) [CGG], which is dominant in the sequence of the c-*myc* gene promoter domain, has been published. Without empirical structural data, we shall use a modeling study to help interpret the existing experimental observations.

A recent theoretical study in our laboratory proposed a possible homogeneous CGG triplex structure,[32] and discussed the need for permutations of strand orientations and base-pairing schemes. The present modeling study is aimed at the base-pairing schemes of triplex formation. Hereafter, the first pyrimidine-rich strand is denoted by I, its W–C base-pairing purine-rich strand partner by II, and the TFO by III. The common nomenclature for nucleotide

Table I. The Triplex Sequences Considered in the Present Study[a]

Triplex	Strand	Sequence
TAT		
1 **pHt**	I	5′-(TTTTTTTTTTTTTTTTTTTTTTTTTTT)-3′
	II	3′-(AAAAAAAAAAAAAAAAAAAAAAAAAAA)-5′
	III	3′-(TTTTTTTTTTTTTTTTTTTTTTTTTTT)-5′
2 **aRt**		as TAT pHt, but with III anti-parallel to II
CGG		
3 **aHs**	I	5′-(CCCCCCCCCCCCCCCCCCCCCCCCCCC)-3′
	II	3′-(GGGGGGGGGGGGGGGGGGGGGGGGGGG)-5′
	III	5′-(GGGGGGGGGGGGGGGGGGGGGGGGGGG)-3′
4 **pHt**		as CGG aHs, but with III parallel to II
5 aRt		as CGG aHs
Hetero		
6 **aHs**	I	5′-(CCTTCCCCACCCTCCCCACCCTCCCCA)-3′
	II	3′-(GGAAGGGGTGGGAGGGGTGGGAGGGGT)-5′
	III	5′-(GGTTGGGGTGGGTGGGGTGGGTGGGGT)-3′
7 **aRt**		as hetero aHs

[a] All residues are deoxyribonucleotides. Strands I and II of hetero **aHs** and **aRt** correspond to the c-myc gene promoter domain.

(26) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859–7870.
(27) Johnson, K. H.; Gray, D. M.; Sutherland, J. C. *Nucl. Acids Res.* **1991**, *19*, 2275–2280.
(28) Ono, A.; Ts'o, P. O. P.; Kan, L.-S. *J. Am. Chem. Soc.* **1991**, *113*, 4032–4033.
(29) Pilch, D. S.; Brousseau, R.; Shafer, R. H. *Nucl. Acids Res.* **1990**, *18*, 5743–5750.
(30) Xodo, L. E.; Manzini, G.; Quadrifoglio, F. *Nucl. Acids Res.* **1990**, *18*, 3557–3564.
(31) Strobel, S. A.; Dervan, P. B. *Nature* **1991**, *350*, 172–174.
(32) van Vlijmen, H. W. Th.; Rame, G. L.; Pettitt, B. M. *Biopolymers* **1990**, *30*, 517–532.
(33) Kennard, O.; Hunter, W. N. *Q. Rev. Biophys.* **1989**, *22*, 327–379.
(34) Arnott, S.; Bond, P. J. *Nature New Biol.* **1973**, *244*, 99–101.
(35) Arnott, S.; Bond, P. J. *Science* **1973**, *181*, 68–69.
(36) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 509–521.
(37) Langley, D. R.; Doyle, T. W.; Beveridge, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 4395–4403.
(38) Lavery, R. In *Structure & Expression. Volume 3: DNA Bending & Curvature*; Olson, W. K., Sarma, M. H., Sarma, R. H., Sundaralingam, M., Eds.; Adenine Press: Guilderland, 1988; pp 191–211.
(39) Sundaralingam, M.; Sekharudu, Y. C. In *Structure & Expression. Volume 3: DNA Bending & Curvature*; Olson, W. K., Sarma, M. H., Sarma, R. H., Sundaralingam, M., Eds.; Adenine Press: Guilderland, 1988; pp 9–23.
(40) Fratini, A. V.; Kopka, M. L.; Drew, H. R.; Dickerson, R. E. *J. Biol. Chem.* **1982**, *257*, 14686–14705.
(41) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246–9255.
(42) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907–916.

atom types has been used throughout.[43,44]

The method for the construction of the model, the scheme for the energy analysis, and the approach adopted for the geometric study are discussed. Results are presented for the homopolymeric series TAT, CGG, and the c-*myc* gene promoter sequence. In addition, we report the results of molecular dynamics simulations, hydrogen–hydrogen (H–H) distances in the structures, and solvation energy calculations. Discussion of the results and conclusions concerning strand orientation and base pairings follow.

II. Model Structures

Within the c-*myc* gene promoter sequence, twenty residues are CG W–C base pairs and the remainder are composed of either isolated TA or AT pairs (Table I). We use the results from a study of homopolymers, which are more easily interpreted, to form the basis for studies that involve compositional change in a gene sequence. The c-*myc* gene promoter sequence structure is constructed and compared with the homogeneous CGG sequence.

For strands II and III, GG and AT base pairs, respectively, can interact through Hoogsteen (**H**) or reversed-Hoogsteen (**R**) schemes. Two nucleic acid strands can be oriented in two ways, parallel (**p**) or antiparallel (**a**), which refer to the relative placement of the O3′ and O5′ oxygen atoms. The strand orientation forces the sugar to be either syn (**s**) or anti (**t**) to the base in **H** or **R** orientation. Therefore, eight different combinations are evaluated (Figure 1). For comparison, both TAT and CGG homopolymers have been constructed. Since both the base pairing and strand orientation of TAT are known, we use the various models of TAT as a control for our calculations.

Figure 1 reveals a basic geometric difference between dT·dA·dT and dC·dG·dG triplets. In the former case, since thymine is monocyclic rather than bicyclic, as is guanine, the phosphodiester backbone of strand III is on the average much closer to the W–C pair. Therefore, because of the geometric requirements, the backbone atoms of strands II and III will strongly repel each other, and the **H** pairing and **s** glycosyl, with an **a** orientation between strands II and III, is an improbable arrangement (Figure 1a), denoted by **aHs**. (This three-letter mnemonic code represents the orientation, base pairing between strands II and III and the glycosyl conformation of strand III of a triplex, respectively.) For the **p** and **R** model of TAT, by forcing the glycosyl dihedral to **s**, the backbone of strand III overlaps with atoms of strands I and II in the major groove (Figure 1b). From this simple analysis, we selected for further consideration only two isomers for a TAT-type triplex: **pHt** (which was the X-ray fiber conformation) and **aRt**.

In the homogeneous CGG case, from similar considerations, the TFO which is **a** to the purine-rich strand II can adopt **s** or **t** for each individual residue intuitively. For the Hoogsteen case, the **s** configuration will be the only probable arrangement (**aHs**). The **t** isomer is also possible through a perturbation of the backbone dihedrals, which renders the average furanose plane oriented more nearly parallel to the base. This, however, induces a strain on the backbone and the resulting unlikely geometry has never been observed. Therefore, we considered further only the **aHs** conformation of this isomer exclusively (Figure 1g). Similarly, for the **R** models, the **t** arrangement is favored (**aRt**, see Figure 1f). Analogous to the **H** case, the **aRs** isomer model is energetically quite unfavorable and will not be further considered in this work. If the TFO is **p** to the purine-rich strand II, the probable triplet schemes for **H** and **R** triplexes are shown as in Figure 1, panels e and h, respectively. Their isomeric **s** and **t** arrangements, respectively, are not considered because of the large perturbations to the furanose rings as mentioned above. Therefore, **aHs** was compared with **aRt**. However, the most stable conformation of the glycosyl dihedral is **t** for a single poly(dG) strand. This implies for **aHs** triplex formation that the single strand TFO has to sacrifice the favored **t** conformation. Thus, the **pHt** will further be compared with the **a** models. Notice that the physically reasonable forms not eliminated by this analysis are in the center of Figure 1 with those on the far left or right being eliminated from further consideration.

III. Methods

The parameters for the molecular mechanics models are primarily taken from ref 51. These authors advocated use of a reduced phosphate charge model for all calculations done without explicit salt counterions. Our solvation energy estimations did not contain explicit counterions (see below). For comparison, the solvation energy calculations above were also performed with a full charge set of parameters taken from ref 50. For minimizations, the distance-dependent dielectric model of solvent screening prescribed[51] was used in our chosen model. Both TA and CG A-form Watson and Crick base-pairing coordinates were taken from X-ray data which included the sugars and phosphate backbones. For TAT triplex base plane templates, the coordinates for the bases of strand III were taken from Arnott's data.[36,45] For the homogeneous CGG series, the initial positions of the bases of strand III were obtained by inspection,[32] ensuring both the base-pairing patterns and steric requirements were approximately met. Next, triple homopolymers were generated with the experimental twist angle of 32.7° and a rise per base pair of 3.04 Å derived from TAT. A short minimization then followed to remove irregular backbone connections. During this minimization, the W–C strands were kept fixed in order to avoid large distortions. A triplet plane around the middle region of the triplex was then extracted and used as a template for further constructions. This procedure was repeated for the four different base-pairing patterns of CGG (Figure 1e–h) discussed above. For the transformation of the homogeneous CGG 27-mer into the c-*myc* gene promoter sequence, the backbone of the guanine residue obtained as above was used for either the thymine or adenine of strand III.

In order to prevent model-built irregularities from producing artifacts in the geometry optimizations, we used the numerically conservative steepest descent (SD) minimizations before employing the more robust adopted basis Newton–Raphson (ABNR) iterations.[46]

By cutting out a representative plane from the minimized sequence as a template for further refinements in construction, the uniformity of the resultant triplex can be improved considerably. The triplex sequences considered are listed in Table I.

Apart from the base interactions between strands,[32] base-specific stacking energies were also computed.[47] Intrastrand base-stacking energies can be computed as the interactions between adjacent base planes within the same strand. The interstrand component included the interactions from residues in other strands in the planes immediately above and below. In both cases, all furanose and phosphate backbone atoms were excluded. The intrastrand base-stacking energy defined in this way will highlight irregular stacking of bases along a particular helical chain, which is especially useful for base mismatch or mutation studies. Extensive bifurcated hydrogen bonding, due to a large twist, will be reflected by the interstrand term.

The interactions between backbone chains (including P, O1P, O2P, O3′, O5′, C3′, C4′, and C5′ type atoms) were also computed separately. The backbone interaction calculations will provide information on the conformation of the TFO in the major groove. Finally, the partition of energy components for the whole triplex into the usual classical force field terms was performed, viz., the potential energy (ener), the intermolecular van der Waals (vdw), electrostatic (elec), and hydrogen-bonding (hbon) components, and the intramolecular bond-stretching, bending, dihedral, and improper dihedral components.

One aim of the geometric analysis was to study how the accommodation of a TFO affected the spatial arrangement of the duplex. The analysis followed the convention used by Dickerson.[40] The structural flexibilities of various triplexes are discussed in terms of those computed geometric parameters.

Accessible surface areas were calculated with a spherical probe (1.4 Å in radius) "water molecule" according to the method of Handschumacher and Richards.[48] Specifically, the accessible surfaces of phosphate oxygen atoms along the helical backbone chains were investigated. This particular consideration was intended to reflect how the TFO occupies the major groove and how the phosphate oxygen atoms of strand III are arranged in the major groove.

(43) Saenger, W. *Principles of Nucleic Acid*; Springer-Verlag: New York, 1988.

(44) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry. Part I: The Conformation of Biological Macromolecules*; W. H. Freeman and Company: New York, 1980.

(45) Arnott, S.; Bond, P. J.; Selsing, E.; Smith, P. J. C. *Nucl. Acids Res.* **1976**, *3*, 2459–2470.

(46) States, D. J. Ph.D. Thesis; Harvard, 1983.

(47) Aida, M. *J. Theo. Biol.* **1988**, *130*, 327–335.

(48) Handschumacher, M.; Richards, F. M. Program ACCESS, 1983.

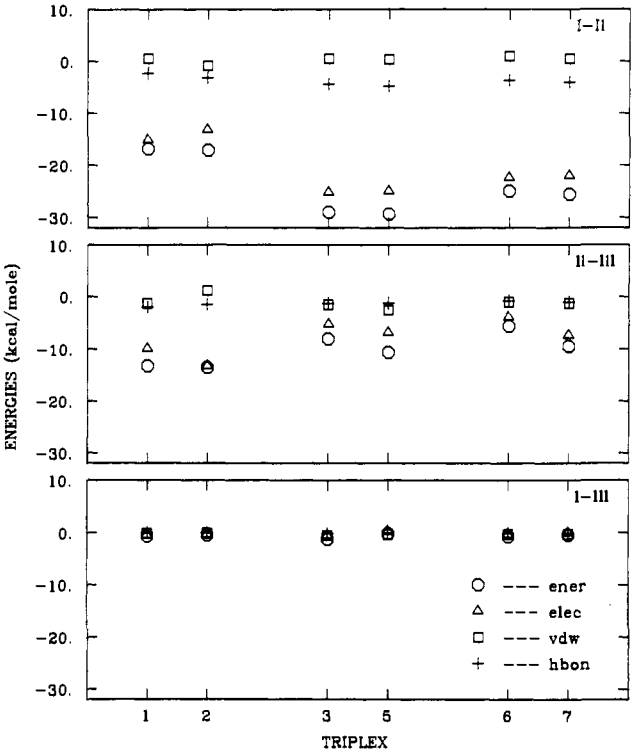

Figure 2. Averages of base-plane interaction energies of triplexes. The intermolecular potential energy (ener) was partitioned into van der Waals (vdw), hydrogen-bonding (hbon), and electrostatic (elec) terms. The legend is given in the bottom frame. The top frame is for the interactions between strands I and II (Watson–Crick pair), the middle frame for the interactions between strands II and III (H or **R**), and the bottom frame for the interactions between strands I and III. See Table I for the reference numbers of the triplexes.

Differences in the free energies of solvation are expected to be a major driving force in the process of association. Therefore, the free energies of solvation were calculated by a linearized Poisson–Boltzmann method.[49] The molecular system (ϵ = 2.0) was mapped onto a grid. Next, calculations were performed with the system surrounded by vacuum (ϵ = 1.0) and in an aqueous solvent (ϵ = 78.0 and zero ionic strength) environment together with the appropriate dielectric boundary conditions. The difference between the two free energy values gives an estimate of the free energy of solvation (transfer from gas to water). The partial charges and radii on atoms were adopted from the literature.[50,51]

The H–H distances for future NOE (nuclear Overhauser effect) comparisons were calculated. Geminal proton distances were omitted in the analysis. The upper-bound used for H–H calculations was 5 Å.[52] The relative intensities were computed by using an inverse sixth-power distance law;[52] unit intensity was set at the upper distance bound.

IV. Results

In this section we apply the above-mentioned methods to two distinct classes of molecules. First, we consider the homopolymers and the forces stabilizing the strand orientations for both TAT and CGG triplexes. Next, the c-*myc* gene promoter site is considered. Results from the homopolymers are used to interpret the heteropolymeric case.

(a) Homopolymers. (i) TAT. As a test case and control for our methods, we examine model TAT triplexes in **p** and **a** orientations. Model minimization results slightly favored **pHt** in terms of total internal potential energy. The energetic decompositions are given in Table II. The van der Waals and intramolecular components were stronger in **pHt**, but **aRt** had larger

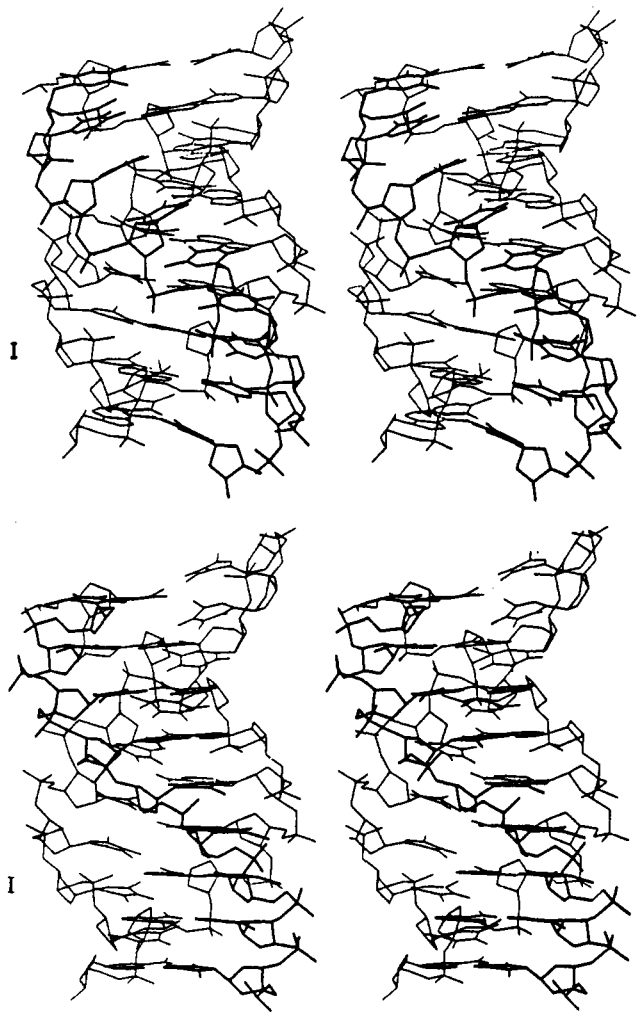

Figure 3. Stereoviews for (a, top) TAT **pHt** and (b, bottom) TAT **aRt**, truncated to 9-mers around the middle portions of the full 27-mers. Strand III's are drawn in thick lines for clarity and strand I's are labeled. All glycosyl dihedral angles are t.

electrostatic interactions (especially between backbones). Partial cancellation of these opposing factors caused **pHt** to be somewhat more stable than **aRt**. We observed an interesting difference in the partitioning of the total base interaction energy. **aRt** has strand III bound more strongly with the W–C pair of strands. While there was a weaker affinity of strand III in **pHt**, the duplex had larger W–C base pairing or was less perturbed by the TFO occupying the major groove than in **aRt**.

At the base-specific level, the magnitudes of W–C and **H** (or **R**) base-paired interactions are similar (Figure 2). This trend is expected as the directed hydrogen bondings are pseudosymmetric about the adenine base in both cases (see Figure 1c,d). Also, the atoms of the two pyrimidines are far enough apart to avoid any significant base-specific-level interaction.

The stereopairs of both **pHt** and **aRt**, truncated to nine residues for each segment around the middle of the structures for clarity, are depicted in Figure 3. The basic difference between these two orientations of strand III is seen in the opposite orientations of its phosphate oxygens with respect to the major groove. Bases of strand III in **pHt** attained the same sense and about the same extent of inclination (relative to the average helical axis) as the W–C pairs. For **aRt**, the adenine and thymine **R** base pairs of strands II and III, respectively, are oriented almost perpendicular to the vertical helical axis, and the individual thymine bases of strand I are forced to shift vertically and off the mean triplet planes slightly. This observation partially explains the energetic results already discussed. The radius of gyration along the helical axis is uniform and essentially identical for both triplexes. Other geometric parameters (roll, tilt, twist, rise per base pair, and minor

(49) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187–197.
(50) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
(51) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591–616.
(52) Thomas, P. D.; Basus, V. J.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1237–1241.

Table II. The Interaction Potential Energy Components for Triplexes after Minimization (500 steps of SD followed by 1000 more steps of ABNR)[a]

	I–II	II–III	I–III	III–(I–II)	total
TAT pHt					
Base Interactions					
ener	−655	−566	−129	−695	−4669
vdw	−114	−165	−28	−193	−2510
elec	−473	−338	−101	−438	−3855
hbon	−69	−63	−1	−64	−143
intramolecular					1838
Backbone Interactions					
ener	0	−118	−206	−325	
vdw	0	−17	−119	−135	
elec	0	−102	−82	−184	
TAT aRt					
Base Interactions					
ener	−615	−591	−151	−743	−4533
vdw	−121	−105	−22	−126	−2342
elec	−398	−439	−129	−568	−4138
hbon	−96	−48	0	−49	−152
intramolecular					2099
Backbone Interactions					
ener	0	−201	−203	−404	
vdw	0	−58	−40	−97	
elec	0	−144	−162	−306	
CGG aHs					
Base Interactions					
ener	−1019	−459	−406	−865	−5948
vdw	−136	−81	−43	−124	−2718
elec	−757	−339	−342	−681	−4964
hbon	−126	−40	−20	−60	−253
intramolecular					1987
Backbone Interactions					
ener	0	−228	−11	−240	
vdw	0	−228	−1	−228	
elec	0	4	−11	−6	
CGG aRt					
Base Interactions					
ener	−1071	−586	−120	−705	−5383
vdw	−125	−176	−27	−203	−2330
elec	−810	−374	−86	−460	−4932
hbon	−136	−36	−6	−43	−204
intramolecular					2084
Backbone Interactions					
ener	0	−224	−112	−335	
vdw	0	−59	−19	−78	
elec	0	−165	−93	−257	
Hetero aHs					
Base Interactions					
ener	−917	−358	−331	−689	−5375
vdw	−128	−64	−65	−129	−2563
elec	−683	−260	−245	−505	−4644
hbon	−107	−34	−21	−54	−227
intramolecular					2059
Backbone Interactions					
ener	0	−239	−4	−244	
vdw	0	−208	−1	−208	
elec	0	−28	−4	−31	
Hetero aRt					
Base Interactions					
ener	−949	−513	−139	−651	−5142
vdw	−135	−165	−25	−190	−2320
elec	−694	−319	−107	−426	−4730
hbon	−119	−28	−7	−35	−177
intramolecular					2084
Backbone Interactions					
ener	0	−188	−138	−326	
vdw	0	−48	−22	−70	
elec	0	−140	−116	−256	

[a] See the Methods section for the abbreviations used for the components. Columns 2–5 (from the left) give the interaction energies of various combinations of strands. The III–(I–II) column represents the affinity of strand III with duplex (I–II). The last column gives the total internal potential energy components which include both intermolecular and intramolecular terms. See the Methods section for the atoms used in the calculation of backbone interactions. All energies are given in kcal/mol.

Table III. The Effective Accessible Surface Areas of Phosphate Oxygen Atoms[a]

	accessible surface		
triplex	I	II	III
TAT			
pHt	26	30	16
aRt	33	29	41
CGG			
aHs	28	34	3
aRt	25	23	46
hetero			
aHs	33	37	2
aRt	34	27	49

[a] The residue average sums of both oxygen atoms of individual strands are given. All values are expressed in Å^2.

Table IV. The Base-Stacking Energies of Triplexes CGG **aHs** and **aRt**[a]

		base-stacking energy	
triplex	strand	intra-	inter-
aHs	I	−4.7	−21.4
	II	−9.4	−16.2
	III	−7.7	−22.6
aRt	I	−4.5	−13.3
	II	−7.7	−19.6
	III	−8.5	−14.9

[a] The values given are residue averages of individual strands. All values are expressed in kcal/mol.

groove width) possessed by the triplexes are similar.

The orientations of the phosphate groups in the major groove, with respect to the third strand direction, led one of the phosphate oxygen atoms on strand III of **pHt** to be completely screened from the solvent (water) probe molecule; the effective accessible surface area of this oxygen atom vanishes. Table III gives the residue average sums of effective accessible surface area from phosphate oxygen atoms of individual strands.

(ii) CGG. We now consider homogeneous CGG. The total potential energy (Table II) of the **aHs** isomer is about 10% more favorable than that of the **aRt** counterpart. This is largely caused by the difference in the van der Waals component. Also, the hydrogen-bonding component for the former is larger by approximately 25%. For **aHs**, in order to adopt a syn dihedral for strand III, the backbone atoms for strand III are essentially in contact with the backbone of strand II (Chart I). This results in a larger magnitude for van der Waals and hydrogen-bonding interactions.

In both model-built triplexes, base-plane interactions and specific intrastrand base-stacking energies are reasonably uniform. Figure 2 shows the relative magnitudes of the three possible interstrand base-plane interactions. As expected, the CG W–C pairs (Figure 2, top frame) have the strongest interactions. The GG pairs (Figure 2, middle frame) also possess base-plane specific interactions, though at about 30% of those of the W–C pairs. In both cases, the electrostatic component dominates. Bases on the same plane of strands I and III, respectively, interact with each other insignificantly (Figure 2, bottom frame). The intrastrand base-stacking energies of **aHs** and **aRt** are similar (Table IV), except a periodic fluctuation of stacking along strand II of **aRt** is observed (not shown). Strand II of **aHs** is more rigid and leads

Table V. Comparison of the Internal Potential Energies of CGG aHs and aRt with the Assumed Single Strand and Duplex Precursors[a]

partition	single strand (G)		duplex (CG)		total (G)+(CG)	aHs	aRt
vdw	(-725)	+	(-1413)	=	-2138	-2718	-2330
elec	(-1287)	+	(-2963)	=	-4250	-4964	-4932
hbon	(-5)	+	(-142)	=	-147	-253	-204

[a] Single strand and duplex were minimized as for triplexes (see caption of Table II). All energies are expressed in kcal/mol.

to a relatively more uniform base stacking. The specific interstrand stacking energies for strands I and III of **aHs** are higher than those of **aRt**. This is a consequence of the bifurcated hydrogen-bonding pattern. For **aHs**, atoms O6 of the guanines on strand III are situated in the middle of the major groove so that extensive interactions with atoms H41 and H42 of the cytosines on strand I, above and below the specific plane considered, are likely (see Figure 1g). The W–C CG pair interactions are stronger than those for the **H** or **R** GG pairs. This means, at the base-plane level, base-pairing patterns of the W–C duplex yield more rigidity. However, the combined interaction of strand III with both strands I and II is comparable to the W–C interactions in both triplexes. This is consistent with a large affinity between the TFO and the underlying duplex that was experimentally observed by both a binding constant study[3] and thermodynamic analysis.[2] Table V gives the sum of the total internal potential energies of single- and double-strand helices and compares it with that of **aHs** and **aRt**.

We now consider the problem of strand orientation in a homopolymeric series. The underlying differences among the four stereoisomers of CGG type may be seen by examining the corresponding base triplet arrangements. In Figure 4, the templates for an **Ht** triplet and an **Rt** isomer are superimposed. If only guanosine III of either triplet is considered, the structures differ by flipping this nucleotide. However, this causes the phosphate oxygen atoms of III to point inwards or outwards with respect to the major groove. More important is the consequence of configuration in the major groove. For the **Ht** case, strand III is biased toward strand I. In contrast, by a slight adjustment of the glycosyl dihedral to less anti, the position in the major groove is more centered in the **Rt** case.

To further elaborate this family of structures, a minimization on CGG with **H** base pairing between strands II and III in p orientation (**pHt**) was performed. The total internal potential energy of this triplex is comparable to and even slightly more negative than **aHs** (-5981 versus -5947 kcal/mol). However, the W–C pair interactions of this triplex are the least among the CGG triplex schemes. From Figure 4, we can see both the base and the extended backbone phosphate group (except the furanose ring, for which the center of geometry remains fixed) of **pHt** are closer to the cytosine of strand I. For **aRt**, both of these moieties are relatively removed from the cytosine of strand I. Similarly, for the remaining two triplex schemes, only one of these moieties is close to the cytosine. Therefore, guanine of strand III in the **pHt** scheme distorts the W–C pairs the most in order to attain a geometric optimum configuration. In fact, the hydrogen bonding between W–C atoms O6 (guanine) and H42 (cytosine) is weakened significantly (see Figure 1e) and the hydrogen-bonding pattern is disturbed. Instead, H41 of the cytosine residue is found to have a more favorable interaction with atom O6 of the guanine base on strand III. Overall, the W–C pair opens up slightly toward the major groove to use the possible guanine of strand III to obtain an energy balance. This point is illustrated through Table VI, in which the W–C interactions of the CGG **aHs**, **pHt**, and **aRt** are listed.

TAT triplexes present a different case. Thymine is a monocyclic base rather than bicyclic as in guanine. The center of the five-membered ring for the thymine base is more or less coincident for either **H** or **R** base pairs (Figure 1). Therefore, the difference in **H** versus **R** base pairings would be less pronounced than in the CGG triplexes. The almost identical values in radius of gyration

Table VI. Comparison of Watson–Crick (I–II) Base-Paired Interactions of CGG aHs, pHt, and aRt[a]

triplex	I–II base interactions			
	ener	vdw	elec	hbon
aHs	-1019	-136	-757	-126
pHt	-867	-128	-617	-123
aRt	-1071	-125	-810	-136

[a] All energies are expressed in kcal/mol.

Table VII. The Free Energy of Solvation (A_{sol}, kcal/mol) of Triplexes Using a Linearized Poisson–Boltzmann Approach[57] for Two Different Electrostatic Charge Sets[a]

systems	1st charge set[b]		2nd charge set[c]	
	A_{sol}	q_{sol}	A_{sol}	q_{sol}
TAT **pHt**	-794	(-6.720)	-5009	(-17.710)
TAT **aRt**	-770	(-6.720)	-5007	(-17.710)
single helix (G)	-227	(-2.240)	-913	(-5.880)
duplex (CG)	-428	(-4.480)	-2514	(-11.760)
CGG **aHs**	-813	(-6.720)	-5118	(-17.640)
CGG **pHt**	-781	(-6.720)	-5029	(-17.640)
CGG **aRt**	-831	(-6.720)	-4941	(-17.640)
hetero **aHs**	-829	(-6.720)	-5255	(-17.660)
hetero **aRt**	-811	(-6.720)	-4928	(-17.660)

[a] Grid size is 0.6 Å and dielectric constant is set to 1.0 and 78.0 for vacuum and aqueous solvent environments, respectively, and 2.0 inside the molecules. Smoothed boundary conditions were employed. Single strand and duplex are the same structures as used in Table V. All helices were truncated to 7-mers. Numbers in parentheses indicate the total net charge of the helices. [b] Reference 51. [c] Reference 50.

for these two TAT structures reflect this point. The effective accessible surface areas of phosphate oxygen atoms (Table III) are still larger for R base pairing, since both the O1P and O2P oxygen atoms point out from the major groove. This is most important from a solvation standpoint (see below). The minimization result showed TAT **pHt** is more stable with respect to the total intramolecular potential energy, in agreement with the X-ray fiber structure.[36]

After minimizations, the glycosyl dihedrals of W–C bases of CGG **aHs** and **aRt** adopt standard values (as found in X-ray structural averages) ranging from -150° to -170°. For strand III of **aHs**, the value is around 50°, which is syn, as constructed. For **aRt**, however, the dihedrals deviate from standard anti value slightly and tend to be around -130°. This minor adjustment in glycosyl dihedrals gives better **R** base pairing and centering in the major groove.

In addition, CGG **aHs** gives smaller propeller twists than **aRt**. This is a consequence of the closeness of strands II and III which prevents free propeller twisting between the W–C pairs. Furthermore, the base-plane radius of gyration (defined as the square root of the summation of moments of inertia of individual base atoms divided by the total mass) helps to quantify a basic difference between the two geometries, with **aHs** yielding a value of about 1 Å less than that of **aRt**. Figure 5 gives the end-on views down the helices of the backbone structural components. These views reinforce this analysis.

If we compare the roll and tilt parameters for strand III defined as for a W–C duplex, we find them to be comparatively larger for CGG **aHs** versus CGG **aRt**. This is caused by the smaller configurational space available for the third strand (see Chart I). The twist angles and the rise per base values remain rather constant and almost identical for all the CGG triplexes studied. The values are close to the starting parameters and the W–C pairs conform with an *A*-DNA canonical helical structure.

The position and conformation of the phosphates correlate well with all the above-mentioned geometric parameters. For CGG **aRt**, the backbone of strand III split the major groove more or less equally with the two negatively-charged phosphate oxygen atoms directed out from the major groove. For CGG **aHs**, however, the third strand backbone is biased toward and partially buried under strand II, with both phosphate oxygen atoms having

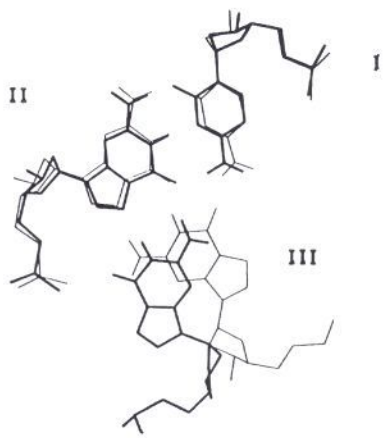

Figure 4. Superimposition of triplets extracted from triplexes 4 (thin) and 5 (thick) in order to illustrate the relative placement of guanosine of strand III in the major groove. The W–C pairs are at the top. CGG **pHt** has both the bases and backbone of strand III closer to strand I.

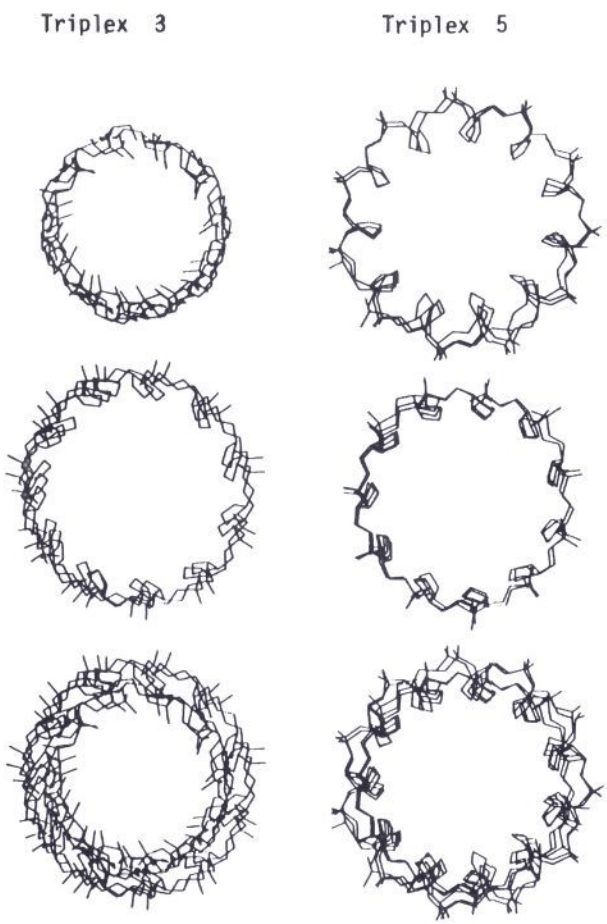

Figure 5. The end-on views of the backbones of strands II and III of CGG **aHs** (triplex 3, see Table I) and CGG **aRt** (triplex 5, see Table I). The views are along the average helical axes. Top and middle rows are for strands III and II, respectively, and they are shown superimposed on the bottom row. Strand I is similar in its distribution of masses around the center of ring for both triplexes and so is omitted. There is an inverse relationship of relative distribution of masses of strands II and III for the two triplexes as seen. Strand III of **aHs** has the phosphate oxygen atoms pointing inside its ring so that it is almost completely within the inner ring of strand II.

little solvent exposure (Chart I). The average effective accessible surface areas of phosphate oxygens are listed in Table III. As expected, the surface areas for both the oxygen atoms on strand III of CGG **aHs** almost vanish. For CGG **aRt**, however, the

Chart I. Space-Filling Models of CGG **aRt** (left) and **aHs** (right)[a]

[a] Phosphate oxygen atoms are in red; backbones of strands I, II, and III are in blue, purple, and yellow, respectively; and all base atoms were darkened. All atoms were given the same radius. It is clear that strand III of **aHs** (yellow) is in close contact with strand II (purple), and its phosphate oxygen atoms (red) are buried inside the backbone of strand II. Overall, **aHs** has a smaller radius of gyration, as is easily observed. Also, the phosphate oxygen atoms of **aRt** are pointing out of the major groove as a distinguishing feature to illustrate the spatial difference in the major grooves.

corresponding values are larger. Energetically, the comparable II–III and I–III backbone interactions (Table II) in CGG **aRt** also reflect the balanced position of its strand III in the major groove.

(b) Solvation. The approximate free energies of solvation computed from a linearized Poisson–Boltzmann equation[49] for truncated versions of all triplexes are given in Table VII. Two sets of partial charges[50,51] were employed for comparison and to check the sensitivity of calculations with respect to these two charge models. Besides minor variations, the major difference between the models is the use of reduced charges[51] on the phosphate oxygen atoms to mimic counterion condensation in one set and full charges on the other.[50] On employing the reduced charge model for the transition from single helix to duplex, and then to triplex, the solvation free energy decreases but less than additively. For full charges,[50] the solvation energy is considerably larger and more nearly additive. The order of solvation energies for CGG **aHs** and **aRt** was interchanged by the two model computations. It is reasonable to expect the phosphate oxygen atoms are coordinated by countercations. Therefore without using explicit free ions, the reduced charge[51] approach should be reasonably realistic and relevant in the present study. The simplistic Debye–Hückel screening, useful for very low salt concentrations with Poisson–Boltzmann calculations, did not significantly alter our results and was not pursued further.

(c) Quenched Dynamics. To further refine our estimates for the two antiparallel CGG cases, room temperature molecular dynamics simulations were performed to explore configuration space near the model-built structures. These short simulations are not meant to represent either the actual dynamics or the equilibrium distributions; they are only meant as a conservative

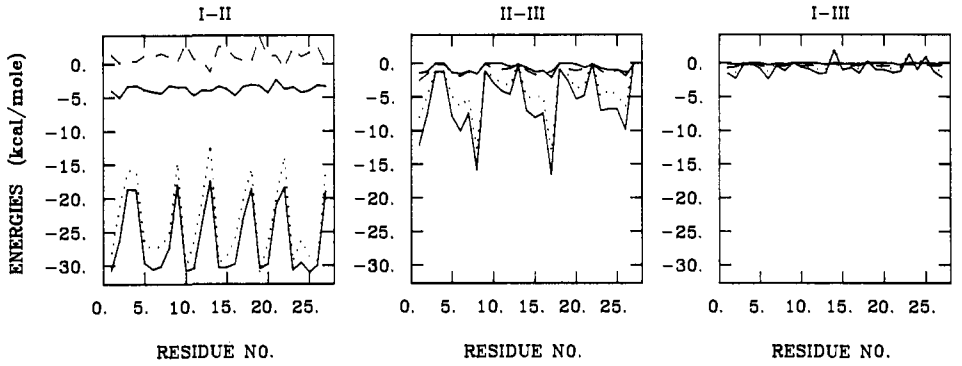

Figure 6. Base-plane interaction energies versus relative residue numbers for hetero **aHs**. The total interaction potential energy (solid) was decomposed into electrostatic (dotted), van der Waals (dashed), and hydrogen-bonding (solid) components. (a) W–C pair strands I and II. Reduced interactions for the TA (residue numbers 3, 4, 13, and 22) and AT duplets (residue numbers 9, 18, and 27) are revealed as local spikes with similar average magnitude of Watson–Crick pair interactions as in TAT **pHt** and **aRt**; (b) Hoogsteen pair strands II and III (large local distorted interactions appear (see ref 32 for detailed discussion)); and (c) interaction between strands I and III.

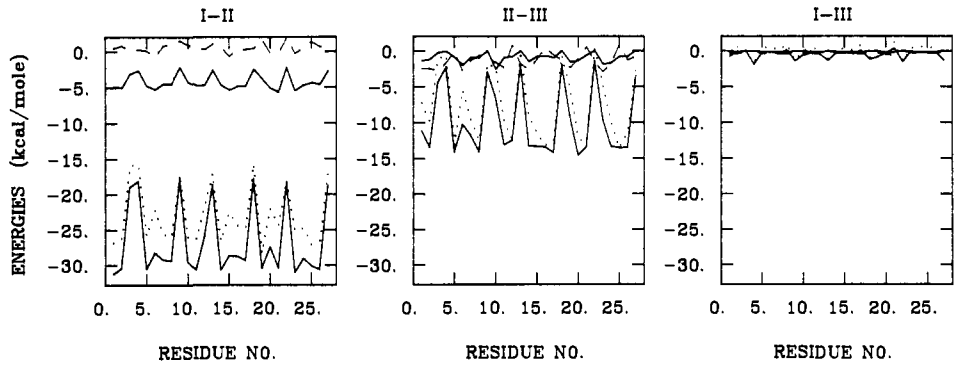

Figure 7. Base-plane interaction energies for hetero **aRt**. The axes and energy components are the same as in Figure 6, except panel (b) represents **R** base pairing between strands II and III. The extent of local distortion is smaller when compared with hetero **aHs** (cf. Figures 7b and 6b).

conformational search. CGG **aHs** was relatively stable and its potential energy stayed almost constant around −4100 kcal/mol for the duration. In contrast, CGG **aRt** started at a somewhat higher potential energy which gradually decreased until about −4000 kcal/mol was reached. The model thermal motions appeared to move CGG **aRt** from a relatively shallow local potential energy minimum to another minimum nearby.

Following the dynamics, the triplexes were further minimized (quenched) and energetic and geometric analyses were done. No specific change in energetic properties for CGG **aHs** (cf. Tables II and VIII) was observed except a slight gain in electrostatic energy. For CGG **aRt**, however, a relatively large increase in total energy was obtained (cf. Tables II and VIII), from both van der Waals and electrostatic contributions. Strand III moved closer to the W–C duplex by removing minor construction artifacts. A slight decrease in radius of gyration observed supports this argument.

(d) Heteropolymers. Next, we considered the consequences of base pairing and orientation on the G-rich 27 base-pair sequence of the c-*myc* gene promoter site. Because of the preponderance of CG pairs versus TA or AT pairs, given the results of the energetic and solvation comparisons in the homopolymeric CGG series, the antiparallel orientation of strand III was considered consistent with experimental data. The internal potential energy components of hetero **aHs** and **aRt** are listed in Table II. When they are compared with corresponding values from the homopolymeric sequences, all the intermolecular interactions diminish. The decrease can be explained by considering the base-plane interactions (Figures 6 and 7). As expected, around the TAT triplets or ATT mismatches (see Table I for the sequences), spikes arise, indicating either reduced or unfavorable interactions.[32] **aHs**, with strand III close to strand II, seems to be perturbed to a greater extent near the mismatches. Other energetic and geometric properties of the two triplexes have a similar pattern. However, hetero **aHs**, which has the syn dihedrals, shows greater fluctuations in the computed properties, especially for the angles, dihedrals,

Table VIII. Results of Quenched Molecular Dynamics Simulations at Room Temperature for CGG **aHs** and **aRt**[a]

	I–II	II–III	I–III	III–(I–II)	total
		CGG **aHs**			
Base Interactions					
ener	−991	−551	−404	−955	−6118
vdw	−119	−91	−56	−147	−2737
elec	−771	−415	−319	−734	−5206
hbon	−102	−46	−28	−74	−264
intramolecular					1838
Backbone Interactions					
ener	0	−242	−3	−245	
vdw	0	−218	−1	−219	
elec	0	−19	−2	−20	
		CGG **aRt**			
Base Interactions					
ener	−994	−636	−219	−855	−5807
vdw	−118	−139	−66	−206	−2445
elec	−769	−464	−119	−583	−5267
hbon	−107	−33	−34	−67	−207
intramolecular					2112
Backbone Interactions					
ener	0	−209	−141	−350	
vdw	0	−69	−25	−94	
elec	0	−140	−116	−256	

[a] Refer to Table II for the identification of various columns. All energies are given in kcal/mol.

roll and tilt openings, and base-stacking patterns. This is in accord with the observation from the study of homopolymers that **H** CGG triplexes are more compact and vulnerable to perturbations. Tilt and roll angles of strand III are large and uniform in CGG **aHs**, but are less uniform in the corresponding heteropolymeric **aHs**. The base mismatched sites caused distortion in the uniform intrastrand base-stacking pattern of strand III, and the effect was

Table IX. The Theoretical Proton–Proton NOE Intensity Differences between CGG **aHs** and **aRt** (minimized structures after equilibration)[a]

H–H type	indicator	intensity diff (arbitrary units)
COPLANAR		
H41---H21	R > H	1661
H41---H22	R > H	1207
H42---H21	R > H	8586
H42---H22	R > H	1229
H42---H1	H > R	2098
H41---H1	H > R	707
ADJACENT		
H41---H21	R > H	2468
H41---H22	R > H	2905[b]
H41---H1	H > R	2398
H42---H21	R > H	1289
H42---H22	R > H	3268
H42---H1	H > R	1148

[a] See Figure 1 for hydrogen atom labels. For each proton pair, the first atom belongs to the cytosine of strand I and the other belongs to the guanine(s) of strand III. "COPLANAR" denotes a proton pair from the corresponding residue level of strands I and III, and "ADJACENT" indicates a proton pair from the individual residue of strand I and the two residues just above and below the corresponding residue level of strand III. Relative intensities of CGG **aHs** and **aRt** are listed in the "indicator" column. "R > H" indicates the intensity of **aRt** is greater than that of **aHs**. [b] Zero NOE intensity was obtained for **aHs** according to calculation.

transmitted along the segment to nearby residues.

The effective accessible surface areas of backbone phosphate oxygen atoms were computed (Table III). The result is analogous to that found for homopolymers. The spatial features of the major groove are apparently determined by the base composition.

(e) H–H Distances. There is a fundamental difference in the base pairings between CGG **aHs** and **aRt**. This geometric difference should be resolvable experimentally and NOE measurements may provide a practical method of analysis. Thus, we have computed distances and theoretical NOE intensities for all possible diagnostic hydrogen pairs in these two structures. Table IX gives the dominant theoretical H–H NOE intensity differences between CGG **aHs** and **aRt** (minimized structures after dynamics). Overall, the **R** base pairing is more susceptible to giving diagnostic NOE signals. This is easy to understand by considering the CGG planar triplets (Figure 1). For the **R** scheme, both hydrogen atoms H21 and H22 are situated in the middle of the major groove, so they are necessarily close to atoms H41 and H42 attached to nitrogen N4 of the cytosine of strand I. However, the guanine base on strand III of CGG **aHs** is closer to strand I than in CGG **aRt**, so the distances between imino proton H1 of guanine and amino protons H41 or H42 of cytosine are shorter, as expected. Other H–H distances between strands I and II are not pertinent given the criterion we have adopted. Thus, based on these model studies, the existence of the coplanar H42–H21 or the adjacent H41–H22 signal would indicate a structure similar to that of CGG **aRt**.

V. Discussion

To assess the relative stabilities of the triplexes, it is natural to consider the free energy of formation of a triplex from its precursors—a single strand and a corresponding duplex. For a specific related set (TAT **pHt** and **aRt** or CGG **aHs**, **pHt**, and **aRt** or hetero **aHs** and **aRt**), the assumption that triplexes with various strand orientations and base pairings are formed from precursors with essentially the same conformations is reasonable. Therefore, for comparison, the total internal potential energies ΔE (Tables II and VIII) calculated can serve as an approximate enthalpic component in vacuum. Similarly, the free energy of solvation data ΔA_{sol} (Table VII) include the entropic as well as enthalpic parts of the solvation free energy of formation of triplexes for comparison of identical sequences. Entropic contributions from vibrational and conformational motions are assumed to approximately cancel out by subtraction in the comparisons. In fact, this assumption is justified in the duplex case from thermodynamic data.[53] In the discussion that follows, only the solvation data from the reduced charge approach[51] implicitly including some counterion screening are used.

For TAT **pHt** and **aRt**, both energetic and solvation terms favored the former (ΔE = −4669 versus −4533 kcal/mol and ΔA_{sol} = −794 versus −770 kcal/mol). Room temperature quenched molecular dynamics simulations have been performed on both triplexes. The results (not shown) indicated that, at room temperature, **pHt** obtained a lower potential energy and became much more stable than **aRt**. It is gratifying to see these results agree with the experimental structural data for this system.[36,45]

The situation for CGG triplexes is more complicated, since we have three triplex schemes to consider (**aHs**, **pHt**, and **aRt**). In order to make a comparison with experimental footprinting results, we first concentrate on **aHs** and **aRt**, which have the same strand orientations as revealed by experiment.[3,7] **H** pairs between strands II and III give **aHs** much lower total internal potential energy than **aRt**. This is caused by the relative distances between strands II and III (see Chart I). Upon quenching the structures from molecular dynamics simulations, the total internal potential energy difference between **aHs** and **aRt** was significantly reduced, but the relative ordering of stability was unchanged. **aHs** still has a more stable ΔE than that of **aRt** (−6118 versus −5807 kcal/mol). From the point of view of solvation data, **aRt** is more favorable than **aHs**. **aRt** had a change in free energy of solvation upon complexation (from precursors single strand and duplex) of −176 kcal/mol (calculation not shown) in contrast to −158 kcal/mol for **aHs** (truncated 7-mers were used). This is in accord with the calculations of the effective accessible surface area of the phosphate oxygen atoms. Essentially, the two backbone phosphate oxygen atoms of strand III in **aHs** are more screened from the solvent environment by the backbone of strand II. Combining the estimated internal potential energy terms from the simulations and free energies of solvation (Table VII), we find that **aHs** is more stable by ~2 kcal/mol per triplet than **aRt** in terms of free energy. It is not clear, however, that our method is reliable for such small differences. Although CGG **pHt** has an internal potential energy similar in magnitude to that of **aRt** (in fact slightly better), its much less negative free energy of solvation upon complexation (−126 kcal/mol for a truncated 7-mer) in comparison with **aHs** and **aRt** has eliminated it as a realistic possible model.

Thus, the strand orientation is strongly coupled to the solvation characteristics as demonstrated by a comparison of CGG **pHt** versus **aHs** or **aRt**. When both types of homogeneous triplexes, TAT and CGG, are compared, the extent of perturbation of the W–C base pairs also contributes to the relative direction of strand II with respect to strand III. In the TAT case, **H** base pairing on **pHt** between strands II and III induces less distortion on the already existing W–C pairs (Table II) than the R in **aRt**. In contrast, the **R** pairing in CGG **aRt** affects the W–C interaction less than that of the **H** in CGG **pHt**.

In this study, hetero **aHs** and **aRt** were treated alike and behaved as homopolymers with point perturbations. The four dT·dA·dT triplets or three dA·dT·dT triplets are almost like isolated sites of irregularity. The effect of dT·dA·dT and dA·dT·dT triplets is to superimpose their local interaction patterns on that of a homogeneous CGG triplex. In this way, the detailed studies on homopolymers gives a valid model of the biologically significant triplex formation in the inhibition of the c-*myc* gene promoter site.

However, it should be pointed out there is a delicate balance among the energy, solvation, strand orientation, and compositional aspects. **aHs** has a slightly more negative solvation free energy than **aRt** (Table VII), as opposed to the trend observed in homopolymeric CGG triplexes, where **aRt** has a somewhat more negative free energy value. There are several possible reasons for this observation. Hetero triplexes are different in composition with respect to CGG triplexes. The former pair contains TAT

(53) Klump, H. H. In *Studies in Modern Thermodynamics 8: Biochemical Thermodynamics*; Jones, M. N., Ed.; Elsevier: New York, 1988; Chapter 3.

and ATT triplets which have different solvation patterns with respect to CGG triplets. As is shown in Table VII, TAT **pHt** has a more negative solvation free energy than its **aRt** counterpart. Therefore, the solvation aspects of hetero triplexes should deviate from that of the homopolymeric CGG triplexes. Second, the solvation calculations are sensitive to the parameter sets (charges and radii) employed. We have varied the charge and radius sets in order to test the sensitivity and have observed that the relative ordering of solvation data for corresponding species is rarely changed. Specific screening effects which are induced by countercations near the molecular surface still remain largely unexplored. Finally, owing to the numerically small difference obtained in the relative stabilities of CGG **aHs** versus CGG **aRt** or hetero **aHs** versus hetero **aRt**, the assumption of counterbalanced entropic contributions from vibrational and thermal motions for a specific set of triplexes may be questionable although it was observed to be valid[53] in duplexes. For more complicated heterogeneous modifications in the sequences, insertions of alternate pyrimidine–purine, or a long random region, further work is required which is beyond the scope of the present paper. This point is reflected by the preferences in orientation.[3,6] which depend on the composition.

We have not included explicit solvent molecules and countercations in our calculations, although a dielectric constant and reduced phosphate charges consistent with counterion condensation theory have been adopted.[32,51,54,55] This limits the accuracy of our results. However, a recent study[56] has implied the charges on phosphate oxygen atoms were not critical to the geometry of single or double helices when modeled in water without salt during a short simulation.

VI. Conclusion

By combining minimization, solubility calculations, and dynamics, the base pairings of triplexes have been explored. For homogeneous TAT triplexes, the present result that strands II and III preferred Hoogsteen base pairing and parallel orientation is in agreement with the known X-ray fiber structures.[36] For a pyrimidine–purine–purine CGG 27-mer, the Hoogsteen base-pairing scheme between the two guanine-rich strands which orient in an antiparallel manner gives the most favorable interactions.

However, the reversed-Hoogsteen hydrogen-bonding scheme of CGG is geometrically simple and introduces the least perturbation to the W–C pair. In addition, the position of the TFO in the major groove is observed to be centered in this scheme, with the negatively-charged phosphate oxygen atoms pointing outward from the major groove and opposed to the solvent and countercation environments. On the contrary, the Hoogsteen base-paired stereoisomer, with the glycosyl dihedral in the syn conformation, resulted in the purine-rich guanine strand of the W–C pair shielding the TFO strand. This causes a distortion from ideal Hoogsteen hydrogen bonding and renders a triplex with a smaller radius of gyration. The major groove is shared unequally and the phosphate oxygen atoms of strand III are partially buried in the backbone of strand II. CGG **pHt** is unlikely to be a possible model because of its much less negative solvation free energy. Comparison of CGG- and TAT-type triplexes shows that perturbations of the W–C pairs in duplexes by the TFO in triplex formations may be a factor in determining the relative orientations of strands II and III.

Triplexes formed by TFOs with c-*myc* gene promoter site (hetero triplexes) are essentially like homogeneous CGG triplexes. Disturbances introduced by the other triplets lead to isolated perturbations. Overall, Hoogsteen base pairing with strands II and III oriented **a** led to somewhat more stable triplexes although solvation effects favor reversed Hoogsteen.

Dickerson's convention[40] for geometric interpretation of a base or base pair was employed in the present study. Currently, several groups[57,58] have been questioning the validity of Dickerson's convention and have proposed more rigorous methods to interpret and manipulate nucleic acid helices. This will not change any conclusions we have drawn. If more complex sequences or DNA–drug interactions were considered, more rigorous methods,[57,58] as proposed very recently, should be very useful for local structural analysis.

Until recently, drug design in this area was centered on small molecule intercalations, minor groove binders, protein–DNA interactions, or mRNA inhibitions. However, recent observations have demonstrated the uptake of intact or modified oligonucleotide by cells.[6,59] Oligonucleotides have been shown to inhibit transcription in several genes by triplex formation in cellular environment.[6] This is a distinctly different approach to genetic medication than the antisense technique.[59] A better understanding of the forces that stabilize triplex structures will be important in developing applications of this method of genetic medication to arbitrary sequences.

Acknowledgment. The authors thank M. Hogan for many interesting conversations. We also thank R. H. Durland, J. M. Hudson, V. Mohan, and J. Perkyns for useful suggestions. The Robert A. Welch Foundation and the Texas Advanced Technology Program are thanked for partial support of this work. The authors thank an anonymous referee for notational suggestions.

(54) Record, M. T., Jr.; Lohman, T. M.; de Haseth, P. *J. Mol. Biol.* **1976**, *107*, 145–158.
(55) Manning, G. S. *Q. Rev. Biophys.* **1978**, *11*, 179–246.
(56) Hausheer, F. H.; Singh, U. C.; Palmer, T. C.; Saxe, J. D. *J. Am. Chem. Soc.* **1990**, *112*, 9468–9474.
(57) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655–667.
(58) Soumpasis, D. M.; Tung, C.-S. *J. Biol. Struct. Dyn.* **1988**, *6*, 397–420.
(59) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 544–584.