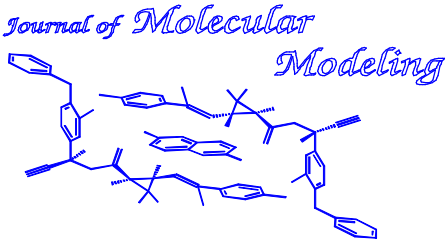
J. Mol. Model. **1996**, *2*, 456 - 466



Novel Hoogsteen-like Bases for Recognition of the C-G Base Pair by DNA Triplex Formation

Jeffrey H. Rothman†, W. Graham Richards*

Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford, UK OX1 3QZ
(jhr@vax.ox.ac.uk)

† Present address: Department of Chemistry, Havemeyer Hall, Box 3154, New York, NY 10027-6948, USA

Received: 21 June 1996 / Accepted: 28 October 1996 / Published: 27 November 1996

Abstract

Effective sequence-specific recognition of duplex DNA is possible by triplex formation with natural oligonucleotides via Hoogsteen H-bonding. However, triplex formation is in practice limited to pyrimidine oligonucleotides that bind duplex A-T or G-C base pair DNA sequences specifically at homopurine sites in the major groove as T·A-T and C^+·G-C triplets. Here we report the successful modelling of novel unnatural nucleosides that recognize the C-G DNA base pair by Hoogsteen-like major groove interaction. These novel Hoogsteen nucleotides are examined within model A-type and B-type conformation triplex structures since the DNA triplex can be considered to incorporate A-type and/or B-type configurational properties. Using the same deoxyribose-phosphodiester and base-deoxyribose dihedral angle configuration, a triplet comprised of a C-G base pair and the novel Hoogsteen nucleotide, Y2, replaces the central T·A-T triplet in the triplex. The presence of any structural or energetic perturbations due to the central triplet in the energy-minimized triplex is assessed with respect to the unmodified energy minimized $(T{\cdot}A\text{-}T)_{11}$ starting structures. Incorporation of this novel triplet into both A-type and B-type natural triplex structures provokes minimal change in the configuration of the central and adjacent triplets.

Keywords: DNA Triplex, Hoogsteen, molecular recognition

Introduction

The ever increasing knowledge of gene sequences has made DNA a suitable drug target. It has become worthwhile to consider sequence selective ligands such as DNA triple helix forming oligonucleotides (TFOs), as one of the more promising routes. Sequence-specific recognition of duplex DNA by triplex formation is induced by major groove Hoogsteen H-bonding to the duplex Watson-Crick base pairs. Oligodeoxynucleotide-implemented triple helix formation also furnishes one of the most versatile methods for sequence-specific recognition of double helical DNA [1,2]. The ability to target a broad scope of DNA sequences, its high stabilities, and single-base mismatch sensitivity make this a powerful method for binding exclusive sites within large segments of duplex DNA. Since the base sequence of a 17-mer oligonucleotide is statistically unique in the sequence of the

* *To whom correspondence should be addressed*

human genome, extremely selective intervention ought to be possible [3]. However, that approach is severely limited if we restrict attention to natural nucleotides since triplex formation is limited to pyrimidine TFOs binding duplex A-T or G-C base pair DNA sequences specifically at homopurine sites in the major groove parallel to the homopurine strand as T·A-T or C^+·G-C triplets. Helix-coil transition melting temperature, UV mixing curve, and ^{1}H NMR experiments all give credence to these homopolymeric structures [4-7]. However, the construction of homopolymeric triplex structures structurally and configurationally analogous to T·A-T and C^+·G-C via TFOs binding in the major groove of a T-A or C-G duplex parallel to the homopyrimidine strand have yet to be experimentally confirmed.

As a preliminary stage, triplets composed of novel Hoogsteen nucleosides designed to bind in the major groove of T-A or C-G base pairs should show structural stability enclosed within a known stable triplex structure. Its appropriateness is then determined by scrutiny of any possible configurational perturbations imposed upon adjacent structure by the triplet and the configuration of the test triplet itself. Previous molecular modelling studies [8-10] have involved a proposed series of novel unnatural nucleosides which demonstrate selective binding to the major groove of a T-A base pair in the center of a T·A-T triplex. In this study base design is targeted for recognition of the major groove of a C-G base pair.

Application in DNA duplex recognition

Depending entirely upon the recognition inherent in the T·A-T and C^+·G-C triplets, the usage of natural TFOs has been successful in mimicing repressors and the construction of artificial restriction enzymes [11-13]. For example, TFOs have been successful in accomplishing single or double site specific cleavage of yeast and human chromosomal DNA [14]. Even at micromolar TFO concentrations sequence-specific inhibition of DNA binding proteins such as prokaryotic modifying enzymes and eukaryotic transcription factor have been successful [15]. TFOs have also been shown to be useful as competitors for DNA-binding proteins and as site-specific DNA damage or cleavage reagents [12,13,16]. As illustrated by the suppression of human *c-myc* gene transcription with nanomolar TFO concentrations, suppression of gene expression via triplex formation also has potential [17]. Additionally, successful suppression of transcription has also been implemented by blocking the promoter region thereby inhibiting the binding of the eukaryotic transcription factor [18]. In light of the adaptability portrayed in these examples, this method demonstrates potential to be a universal solution for DNA recognition.

Much effort has gone into the design of nonnatural nucleotide bases for TFOs, especially those that aim to bind specifically to T-A or C-G base pairs via Hoogsteen H-bonding with the same parallel orientational geometry as the known T·A-T and C^+·G-C natural triplexes [19-24]. Previous energy minimization modelling studies [8,10] of Hoogsteen-like bases, X3 (Figure 1a), and X5, (Figure 1b) have shown successful specific binding of the T-A major groove within an 11mer T·A-T triplex in the A-type and B-type configuration, respectively. These energy minimization studies demonstrate minimal or no structural perturbation to the adjacent triplets with respect to the energy minimized control $(T{\cdot}A\text{-}T)_{11}$ model. Further molecular dynamics studies of this $(T{\cdot}A\text{-}T)_5(A{\cdot}T\text{-}X)(\ T{\cdot}A\text{-}T)_5$ explicitly solvated system with counter ions showed similar structural root mean square deviation behavior in both configurations to that of its related $(T{\cdot}A\text{-}T)_{11}$ triplex configuration [9].

In the present study a Hoogsteen base is designed to target the C-G major groove. Favorable stacking and Hoogsteen interaction energies, and a comparable minimized phosphodiester backbone and nucleotide geometry to that of the known natural triplets are necessary requirements for a successful Hoogsteen base. Ligand construction within con-

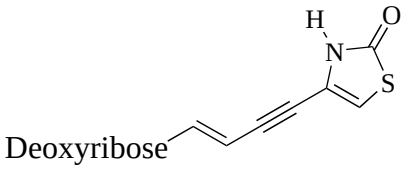


Figure 1a. *Hoogsteen base X3*

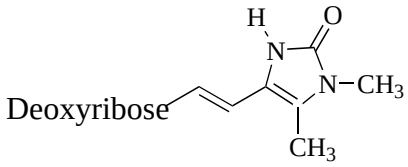


Figure 1b. *Hoogsteen base X5*

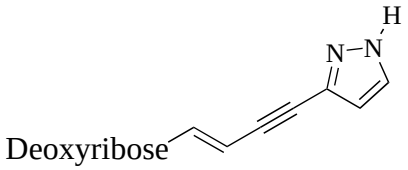


Figure 2a. *Hoogsteen base Y1*

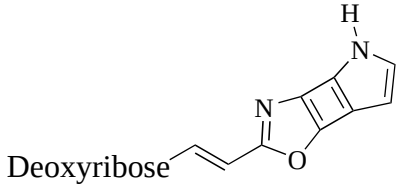


Figure 2b. *Hoogsteen base Y2*

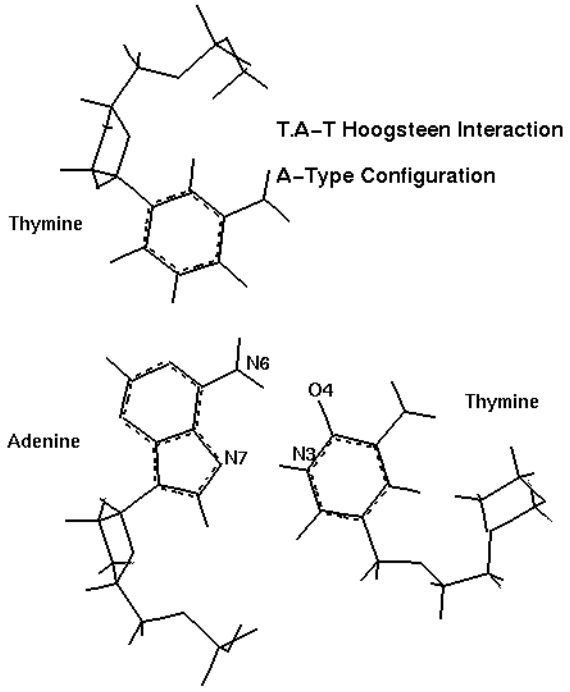


Figure 3a. *Energy minimized T.A-T triplet configuration within the (T.A-T)$_{11}$ triplex from the A-type configuration*

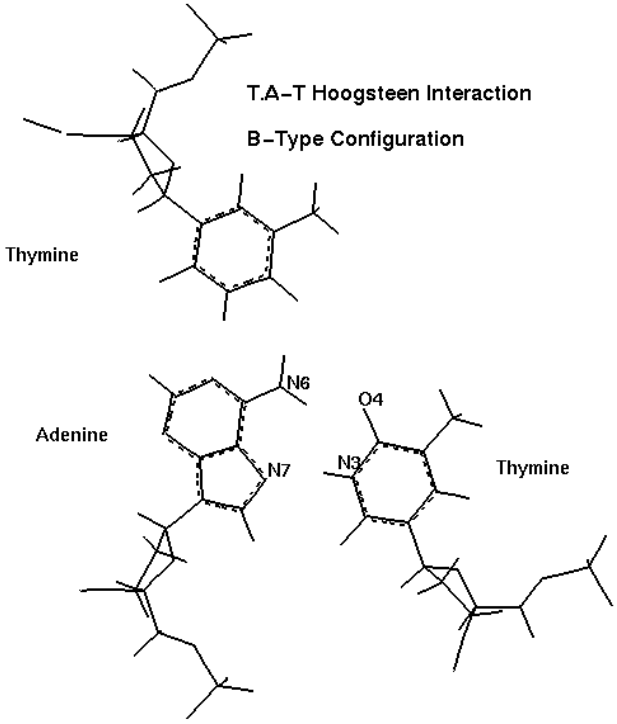


Figure 3b. *Energy minimized T.A-T triplet configuration within the (T.A-T)$_{11}$ triplex from the B-type configuration.*

formationally accommodating targets which need not remain near to their given geometry for optimal binding interactions is not easily accommodated by current algorithm-based ligand design routines. Nonnatural bases may bind with comparable energetics with respect to known stable triplets, but are allowed conformational latitude without significant energetic penalty to the host triplex. This is best portrayed by the detectable binding of guanine in the G·T-A triplet by sequence specific binding-cleavage methods [21], while NMR evidence indicates significantly distorted triplex geometry [25]. Due to this Hoogsteen triplet energy degeneracy any subsequent distortions imposed upon the adjacent triplets must be monitored for interaction energy decreases and maintenance of comparable structural geometry. Here we report the successful modelling of the novel nucleoside, Y2, (Figure 2b) cyclobuta [1,2-d] (*E*) 2-(1-(2-deoxy-b-D-ribofuranosyl)-1-ethene) oxazole [4,3-b] pyrrole, as the Y2·C-G triplet within a T·A-T triplex.

Design Premise

Development of the nonnatural bases from the C1' position of the central deoxyribose of the Hoogsteen strand (strand III) were directed by van der Waals boundaries and potential H-bonding sites of the central bases of the Watson-Crick strands (strands I and II). Due to the planar base stacking constraints the constructed nonnatural bases are limited to aromatic structures. In order to achieve maximal interaction energy and specificity the Hoogsteen base design plan includes H-bonding with both the cytosine of strand II and guanine of strand I rather than only its neighboring strand II base, which is the circumstance for the natural T·A-T triplet (Figures 3a,b). Positioning the aryl portion of Y2 by an ethenyl linker accomplishes this H-bonding that spans these Watson-Crick nucleotide bases (Figure 4). By spacing appropriately the oxazole N1 and pyrrole 1NH H-bonding components of Y2, repulsive cross interference is minimized [26] between N1 and Gua-O6 and between 1N-H Cyt-4NH_2. Initial modelling studies demonstrated that close proximity between H-bond donor and acceptor moieties such as in a pyrazole type Hoogsteen base, Y1, (Figure 2a) causes repulsive interactions between the same H-bond moieties of the Watson-Crick nucleoside bases (Figure 5a,b). For this reason Y2 was developed to allow maximum separation of the adjacent H-bonding components, N: and N-H, and place them in favorable positions to interact with the Cyt-4NH_2 (strand II) and Gua-O6 (strand I), respectively. The Y2·C-G triplet demonstrates comparable stacking and Hoogsteen interbase interaction energies and geometries with respect to those of a T·A-T triplet in the center of a (T·A-T)$_{11}$ triplex.

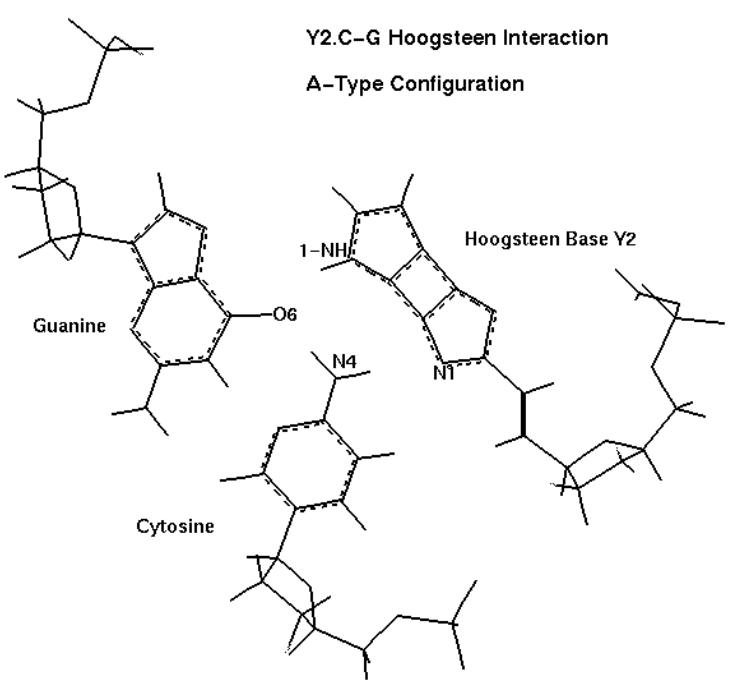


Figure 4a. *Energy minimized Y2.C-G triplet configuration within the (T.A-T)$_5$-(Y2.C-G)-(T.A-T)$_5$ triplex from the A-type configuration.*

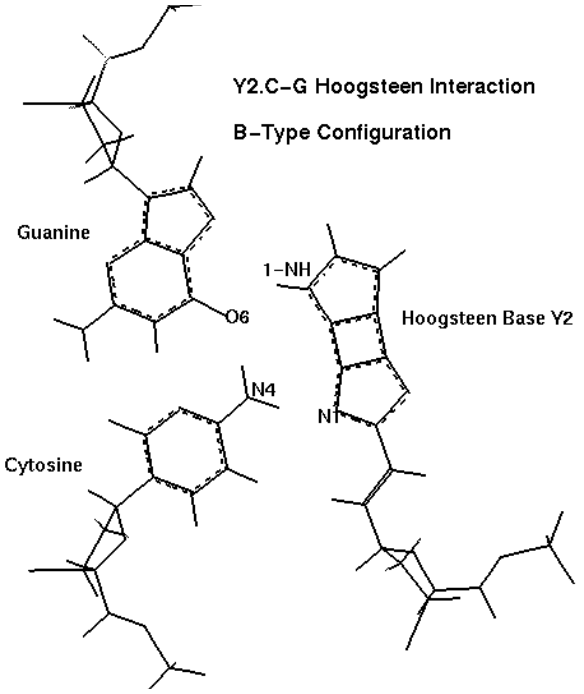


Figure 4b. *Energy minimized Y2.C-G triplet configuration within the (T.A-T)$_5$-(Y2.C-G)-(T.A-T)$_5$ triplex from the B-type configuration.*

Triplex conformations

The Arnott fibre diffraction data [27] provided the first DNA triplex models. Configurations of T·AT triplex derived from these results are considered to be similar to A-DNA with C3'-endo ribose puckers. Results from NMR NOE experiments are consistent with the fibre diffraction data indicating C3'-endo configuration on all three strands of T·AT triplex [28]. At present date, there are still no reports of crystal diffraction studies of DNA triplex. However, recent solution IR spectroscopy studies [29] of T·AT triplex and solution NMR studies [30] of various triplex oligonucleotide systems suggest that many of the nucleotide residues have ribose puckers nearer to a C2'-endo configuration which would be in better agreement with a B-DNA type conformation. Consequently, in light of these new findings a more appropriate model T·AT triplex has been devised [31,32] in which all three strands have the same phosphodiester geometry and C2'-endo ribose pucker that characterizes B-type DNA geometry. More insight into this dilemna may be gained from recent molecular dynamics simulations [33] of the T·AT DNA triplexes from both the A-type and B-type starting configurations. These simulations show similar trajectories which converge to a structure that is structurally equidistant from, but not very similar to either of the initial A-type or B-type structures. Although it may be difficult to produce an accurate representation of a time-averaged helical unit from these dynamics simulations, the resultant convergence to similar structure suggests available configurational pathways between A-type and B-type DNA triplex conformations. Unfortunately, overwhelming proof or rejection of the preference of one type of conformation over the other is not evident. In light of this situation it would be prudent to design Hoogsteen bases that are viable for each general triplex conformation.

Methods

For test purposes, the A-type conformation host undecamer T·A-T DNA triplex, $(T{\cdot}A\text{-}T)_{11}$, was constructed from the Arnott fibre diffraction model [27] where the triplet step height of 3.26Å and a turn angle of 30.0° were used. This is related to placement of a Hoogsteen binding pyrimidine strand into the major groove of an A-DNA duplex parallel to the purine strand. The B-type conformation $(T{\cdot}A\text{-}T)_{11}$ DNA triplex was constructed from the T·AT triplex structure proposed by Sasisekharan [31] in which all three strands have the same ribophosphodiester geometry, nevertheless the step height and turn angle are equivalent to those of the A-type

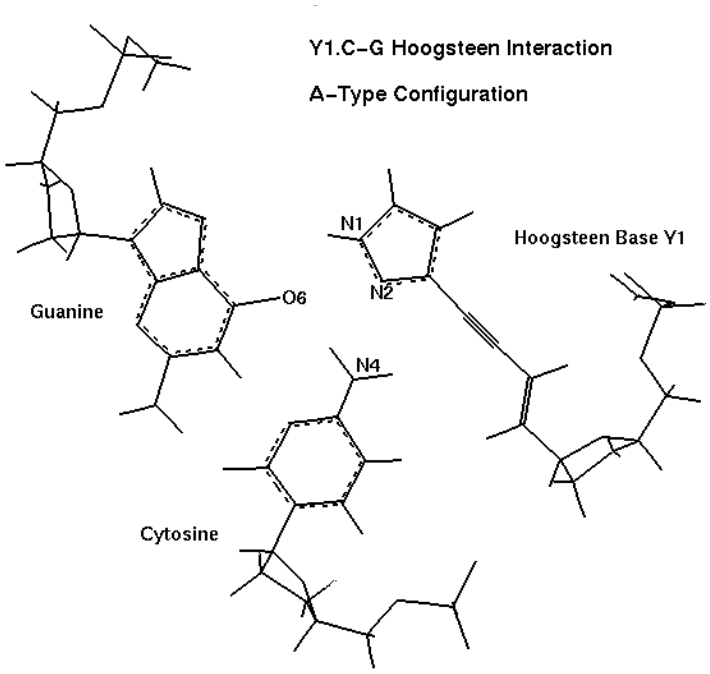


Figure 5a. *Energy minimized Y1.C-G triplet configuration within the (T.A-T)$_5$-(Y1.C-G)-(T.A-T)$_5$ triplex from the A-type configuration.*

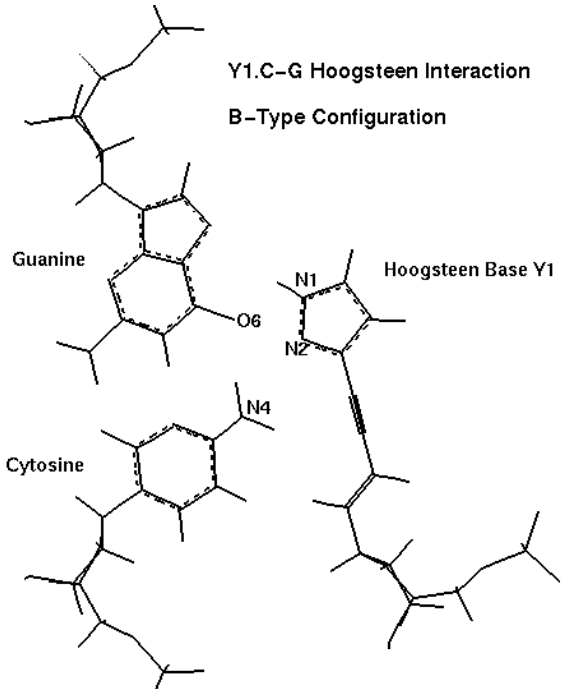


Figure 5b. *Energy minimized Y1.C-G triplet configuration within the (T.A-T)$_5$-(Y1.C-G)-(T.A-T)$_5$ triplex from the B-type configuration.*

conformation. In order to create the test triplexes, the center triplet of each host undecamer was then modified to create the binding region for testing nonnatural base candidates by replacing the A-T nucleotide bases of the Watson-Crick strand with C-G and replacing the nucleotide base of the Hoogsteen strand with the nonnatural bases to be tested, retaining the strand specific deoxyribose-base dihedral angles.

The AMBER force field does not contain all of the parameters required for the proposed nucleoside bases, Y1 and Y2. For purposes of calculating the atomic partial charges of these proposed bases their geometries were determined with MOPAC [34] using the AM1 hamiltonian. The charges of the MOPAC determined structure were then obtained from GAUSSIAN90 [35] with an RHF/STO-3G basis set, using the CHELPG [36] method, and scaled to fit the AMBER 4.0 force field for natural nucleotides. Analyses of the proposed bases within the host triplex were performed with the AMBER [37] suite of programs. Energy minimizations of these triplexes were performed via 100 steps steepest descent and 1000 steps of conjugate gradient method with a nonbonded cutoff distance of 8Å, and a linear distance dependent dielectric to model the implicit water solvation [38].

Results

Charge compatibility of proposed unnatural nucleoside bases with AMBER

Since there are no pyrazole constructs in the AMBER force field, comparison of the Y1 fitted charges is not possible. However, comparison of the fitted charges on the significantly electrostatic interacting portions of Y2 show similar values to those of analogous 5-membered heterocyclic substructures in AMBER. The fitted charges of the pyrrole N-H (N: -0.131, H: 0.246) in Y2 are similar to that of the imidazole (N: -0.142, H: 0.228) portion of histidine in AMBER. The fitted charge of the oxazole N: (N: -0.536) portion of Y2 is also similar to that of the imidazole portion of histidine (N: -0.502), and the other 5-membered aromatic imine substructures of guanine and adenine (both N: -0.543) in AMBER. In fact these charge values of Y2 are also comparable to the similarily calculated charge values of imidazole N-H (N: -0.245, H: 0.232) and N: (N: -0.550). The apparently comparable charges calculated by the two procedures for imidazole and related substructures indicates a reasonable compatibility between the fitted charges of the unnatural Y2 base and charges on similar structures in the AMBER force field.

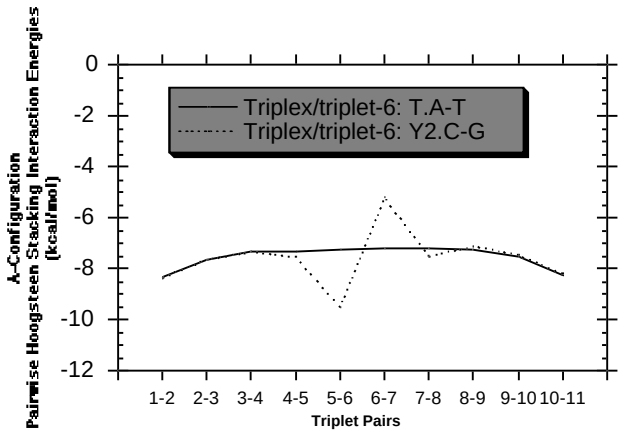


Figure 6a. *Pairwise Hoogsteen Stacking interaction energies for triplexes with the central triplets: T·A-T, Y2·C-G in the A-type conformation.*

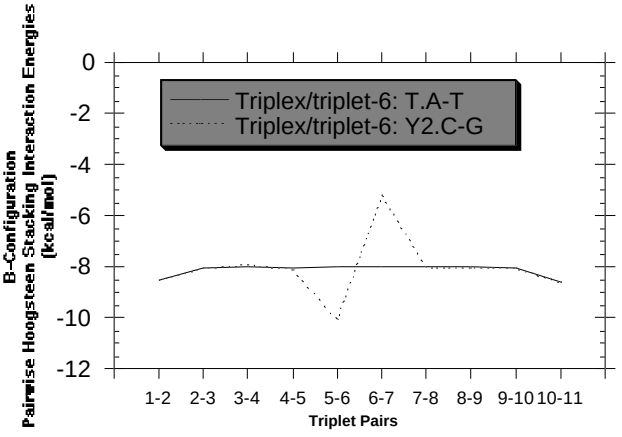


Figure 6b. *Pairwise Hoogsteen Stacking interaction energies for triplexes with the central triplets: T·A-T, Y2·C-G in the B-type conformation.*

Analysis of Control Structure

The structural and energetic behavior of the Y2·C-G triplet is compared to that of a T·A-T triplet as their structures are energy minimized within the center of a $(T\cdot A\text{-}T)_{11}$ triplex. The C^{+}·G-C triplet was not chosen as a basis for comparison over that of the T·A-T in triplet-6 due to the inconsistencies in comparisons as the protonated form. Before analyzing the energetic perturbation effects on the modified triplex $(T\cdot A\text{-}T)_5(Y2\cdot C\text{-}G)(T\cdot A\text{-}T)_5$ due to replacement of the central triplet within the $(T\cdot A\text{-}T)_{11}$ triplex in both the A-type and B-type configurations, the naturally occuring inherent perturbations in the energy minimized $(T\cdot A\text{-}T)_{11}$ triplex model itself must be studied. End effects concerning the end triplets are evident mainly due to the lack of base stacking interactions on their outer surface. Not only is this manifested in the weaker total interaction energies of the end bases in the Hoogsteen strand, but also the configuration of the bases of the end triplets demonstrate increased buckle angle, thus placing the center of the triplet closer to their adjacent triplet. This distortion appears to allow a more favorable stacking energy between Hoogsteen bases, and is propagated in a decreasing manner towards the center of the triplex for A-type and B-type conformations as demonstrated by the pairwise stacking energies in Figures 6a,b. This distortion also allows a closer and more favorable H-bonding distance between the Hoogsteen base and the Watson-Crick bases as demonstrated by slightly increased Hoogsteen interaction energies with the nucleosides of its triplet nearer to the ends of the triplex as demonstrated in Figures 8a,b. Aside from end effects due to the lack of outlying base stacking interactions, energies between bases of the Watson-Crick strands (I-II) and their base-stacking interaction energies (Figures 6a,b) appear consistent throughout the $(T\cdot A\text{-}T)_{11}$ triplex in

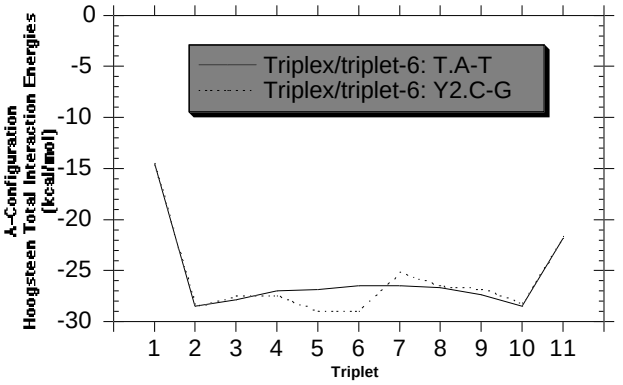


Figure 7a. *Interaction energies for each Hoogsteen base with the rest of the triplex with central triplets: T·A-T, Y2·C-G for the A-type conformation.*

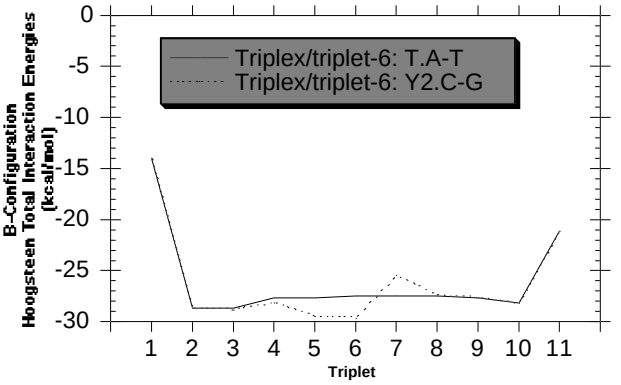


Figure 7b. *Interaction energies for each Hoogsteen base with the rest of the triplex with central triplets: T·A-T, Y2·C-G for the B-type conformation.*

Table 1. *Deoxyribose-phosphodiester dihedral angles (degrees) for A-configuration triplex structures: Triplex-0: (T.A-T)$_{11}$ and Triplex-1: (T·A-T)$_5$-(Y2·C-G)-(T·A-T)$_5$ (see page 464)*

Strand-I Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-64.7	-64.0
O3'-C3'	-169.1	-168.1
C3'-C4'	79.2	78.1
C4'-C5'	61.6	63.4
C5'-O5'	173.7	173.1
O5'-P	-73.0	-71.3
C1'-Base	29.9	23.3

Strand-II Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-72.9	-62.05
O3'-C3'	172.2	-167.4
C3'-C4'	63.3	80.6
C4'-C5'	78.2	64.8
C5'-O5'	-169.2	171.1
O5'-P	-63.8	68.0
C1'-Base	25.2	28.5

Strand-III Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-73.4	-63.3
O3'-C3'	174.4	-163.8
C3'-C4'	61.2	81.4
C4'-C5'	79.8	60.9
C5'-O5'	-167.9	173.7
O5'-P	-65.3	-72.42
C1'-Base	31.5	39.93

both configurations and relatively insensitive to terminal configurational distortion for both types of triplex starting conformations. Similarily, the Hoogsteen interaction energies between Hoogsteen bases (strand III) and the Watson-Crick bases (strands I,II) are also consistent throughout the neat triplex in both configurations (Figures 8a,b). Consequently, the central nine triplets remain energetically and structurally consistent.

A-configuration results

Upon examination of the Y2·C-G triplet within the energy minimized (T·AT)$_5$(Y2·C-G)(T·AT)$_5$ triplex starting in the A-configuration, the total interaction energy of the unnatural Hoogsteen nucleoside Y2 (-28.94 kcals) is 109% that of the analogous value in the control (T·AT)$_{11}$ triplex (-26.57 kcals). This is primarily due to the Y2 nucleoside having favorable interactions in recognition of both of the Watson-Crick nucleosides in its triplet whereas the Hoogsteen base in a natural T·AT and C$^+$·GC triplet only recognize the strand II Watson-Crick base. Accordingly, the Hoogsteen interaction energy of the Y2 nucleoside (-5.59 kcals) with the C-G Watson-Crick nucleosides is 112% that of the analogous value in the control (T·AT)$_{11}$ triplex (-5.00 kcals). Although the Y2 base accomplishes two H-bonds just like a natural Hoogsteen base, it orients one to each Watson-Crick base whereas a natural Hoogsteen base orients both to one. The improvement in the Y2 Hoogsteen interaction in comparison to a natural Hoogsteen interaction lies in the added favorable van der Waals contact which is available due to its proximal orientation to both Watson-Crick bases. The Hoogsteen H-bonding distances of 2.80Å and 3.24Å are within range of viable H-bonding interactions between Y2-1NH — Gua-O6, and Y2-N1 — Cyt-4NH$_2$, respectively. Analogous H-bonding distances in the A-configuration for a T·AT central triplet are 2.90Å and 2.88Å for the Thy-O4 — Ade-6NH$_2$ and Thy-3NH — Ade-N7 interactions, respectively. The Y2 H-bonding distance to Cyt-4NH$_2$ is somewhat attenuated with respect to the equilibrium distance, and this subsequent difference in H-bonding strength is reflected in the Y2 -2.92 kcal interaction energy with the strand II cytosine and the -2.67 kcal interaction energy with the strand I guanine.

Additionally the p-stacking interaction energy of the Y2 Hoogsteen base (-14.75 kcals) with its adjacent Hoogsteen bases are 102% that of the analogous value in the control (T·AT)$_{11}$ triplex (-14.45 kcals). However, due to the orientation of the Y2 base within its triplet plane with respect to the analogous Hoogsteen base in a central T·AT triplet, the 3' adjacent thymine base (triplet 7) receives less p-overlap and the 5' adjacent thymine base (triplet 5) receives more p-overlap than their analogous adjacent bases in the control (T·AT)$_{11}$ triplex. Their p-interaction energies, -5.23 kcals and -9.53 kcals, 5' and 3' respectively, reflect the disparity in p-overlap above and below the Y2 base in comparison to the -7.26 kcals and -7.19 kcals p-interaction energies of the analogous central triplet Hoogsteen base in the control (T·AT)$_{11}$ triplex.

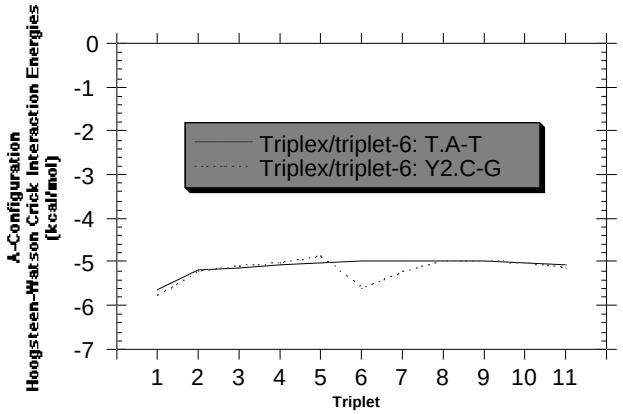


Figure 8a. *Interaction energies between the Hoogsteen base and its accompanying Watson-Crick base pairs for triplexes with central triplets: T·A-T, Y2·C-G for the A-type conformation.*

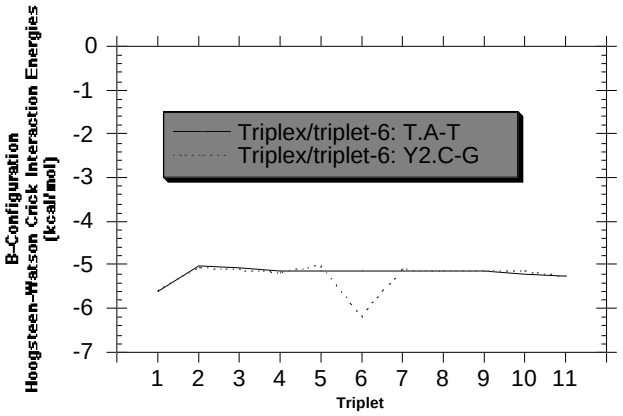


Figure 8b. *Interaction energies between the Hoogsteen base and its accompanying Watson-Crick base pairs for triplexes with central triplets: T·A-T, Y2·C-G for the B-type conformation.*

However, this does not cause any detrimental structural asymmetries as shown by the phosphodiester and c-dihedrals of the energy minimized $(T\cdot AT)_5(Y2\cdot C\text{-}G)(\ T\cdot AT)_5$ triplex structure as compared to those of the energy minimized $(T\cdot AT)_{11}$ control triplex (Table 1, Figure 9a). Consequently, the Hoogsteen nucleoside stacking interaction energy, Hoogsteen nucleoside total interaction energy profiles, and Hoogsteen-Watson Crick energy profiles also closely follow that of the $(T\cdot AT)_{11}$ control triplex (Figures 6a,7a,8a). The Watson-Crick interaction energy profile closely follows that of the control $(T\cdot AT)_{11}$ triplex except for the stronger interaction energy of C-G in triplet-6 due to its three H-bond interaction in contrast to two for the A-T pair. This is in accord with the dihedral results (Table 1) which show almost no structural perturbation to the Watson-Crick strands (I-II).

Table 2. *Deoxyribose-phosphodiester dihedral angles (degrees) for B-configuration triplex structures: Triplex-0: $(T.A\text{-}T)_{11}$ and Triplex-1: $(T\cdot A\text{-}T)_5$-(Y2·C-G)-$(T\cdot A\text{-}T)_5$* *(see page 464)*

Strand-I Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-86.0	-86.2
O3'-C3'	-175.8	-175.8
C3'-C4'	103.4	106.2
C4'-C5'	58.6	57.5
C5'-O5'	171.5	172.2
O5'-P	-66.3	-65.7
C1'-Base	42.1	40.3

Strand-II Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-87.2	-86.3
O3'-C3'	-176.3	-176.2
C3'-C4'	105.7	104.2
C4'-C5'	57.6	57.8
C5'-O5'	172.9	172.0
O5'-P	-66.2	-66.7
C1'-Base	41.6	40.7

Strand-III Triplet-6 Backbone Dihedrals

Dihedrals	Triplex-0	Triplex-1
P-O3'	-86.6	-89.1
O3'-C3'	-176.2	-176.9
C3'-C4'	104.7	113.0
C4'-C5'	58.4	57.3
C5'-O5'	172.6	174.1
O5'-P	-66.9	-66.3
C1'-Base	41.1	50.4

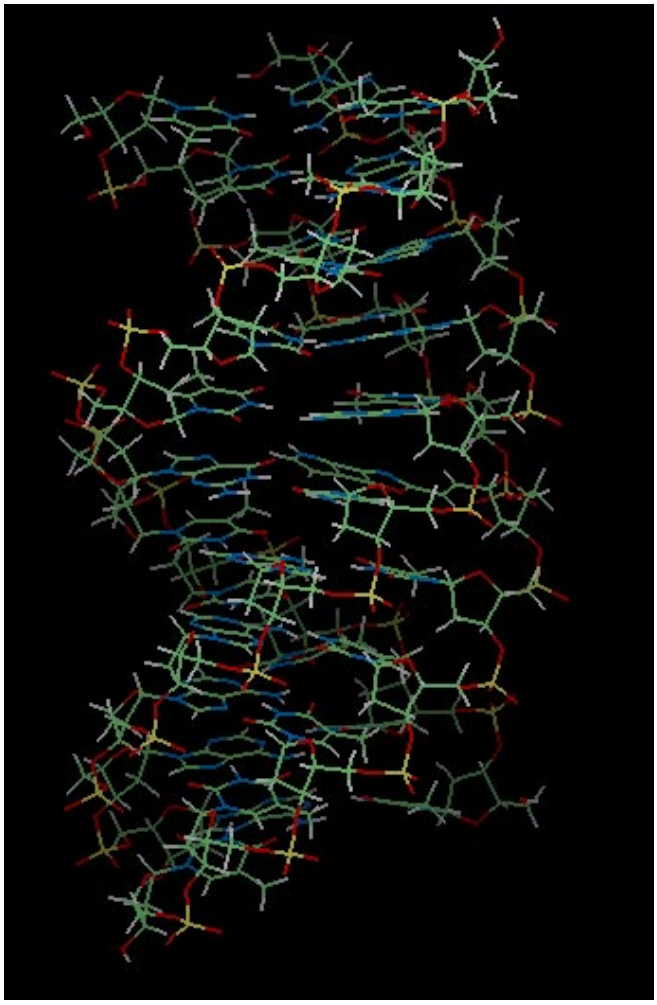


Figure 9a. *Energy minimized (T·A-T)$_5$-(Y2·C-G)-(T·A-T)$_5$ A-type triplex configuration. 3D-structure as PDB-file included.*

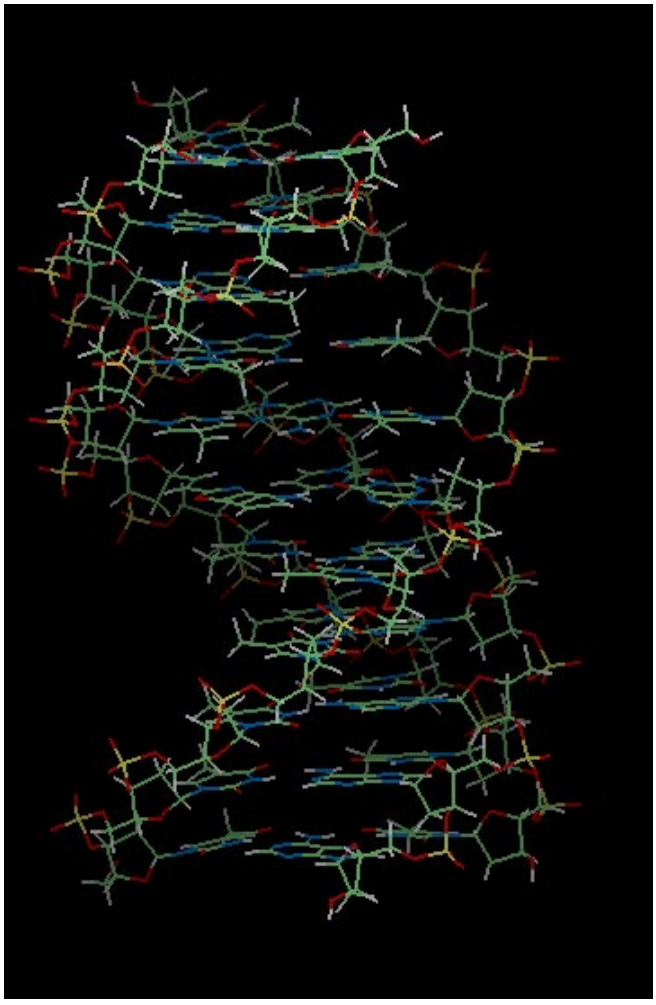


Figure 9b. *Energy minimized (T·A-T)$_5$-(Y2·C-G)-(T·A-T)$_5$ B-type triplex configuration. 3D-structure as PDB-file included.*

B-configuration results

With respect to energetic trends the Y2·C-G triplet within the energy minimized (T·AT)$_5$(Y2·C-G)(T·AT)$_5$ triplex performs similarily in a starting B-configuration in comparison to the results for the starting A-configuration. While the total interaction energy of the unnatural Hoogsteen nucleoside Y2 (-29.58 kcals) is 108% that of the analogous value in the control (T·AT)$_{11}$ triplex (-27.52 kcals), its Hoogsteen interaction energy with the Watson-Crick bases (-6.17 kcals) is 120% of the analogous interaction energy in the central triplet of the control (T·AT)$_{11}$ triplex (-5.14 kcals) in the starting B-configuration. The Hoogsteen H-bonding distances of 2.81Å and 2.92Å are within range of viable H-bonding interactions between Y2-1NH — Gua-O6, and Y2-N1 — Cyt-4NH$_2$, respectively. In comparison, the analogous H-bonding distances in the B-configuration for a T·AT central triplet are a very near 2.87Å and 2.92Å for the Thy-O4 — Ade-6NH$_2$ and Thy-3NH — Ade-N7 interactions, respectively. The similar Y2 interaction energies to the strand II cytosine (-3.12 kcals) and the strand I guanine (-3.05 kcals) reflect the likeness in the H-bonding distance to each Watson-Crick base.

The planar orientation of the Y2 base also causes a similar disparity in the Hoogsteen stacking interaction energies with its adjacent Hoogsteen bases in the B-configuration. A biased increased p-overlap with the 5'-adjacent Hoogsteen base (-10.08 kcals) and decreased overlap with the 3'-adjacent Hoogsteen base (-5.22 kcals) compared to the p-overlap energies (-8.02 kcals) and (-7.99 kcals) of the analogous 5' and 3' adjacent Hoogsteen bases in the control (T·AT)$_{11}$ triplex is as evident in the B-configuration as in the A-configuration. However, the total p-stacking interaction energy for the Y2 base (-15.30 kcals) is 95.6 % of the total stacking interaction for the analogous Hoogsteen base (-16.01 kcals) in the control (T·AT)$_{11}$ triplex whereas the total p-stacking energy for its 5'-adjacent Hoogsteen base (triplet-5) (-17.09 kcals) is 107% and the 3'-adjacent Hoogsteen base (triplet-7) (-12.76 kcals) is 80% of the total p-stacking interaction for the analogous Hoogsteen base in the control (T·AT)$_{11}$

triplex. This disparity in pstacking interaction energies is noticeable in the Hoogsteen p-stacking interaction profile (Figure 6b) as well as the Hoogsteen nucleoside total interaction energy profile (Figure 7b). However, this does not cause any detrimental structural asymmetries as shown by the phosphodiester and c-dihedrals of the energy minimized $(T\cdot AT)_5$(Y2·C-G)($T\cdot AT)_5$ triplex structure as compared to those of the energy minimized $(T\cdot AT)_{11}$ control triplex (Table 2, Figure 9b). Just like the A-configuration triplex results, the Hoogsteen nucleoside stacking interaction energy, Hoogsteen nucleoside total interaction energy profiles, and Hoogsteen-Watson Crick energy profiles in the B-configuration closely follow that of its $(T\cdot AT)_{11}$ control triplex (Figures 6b,7b,8b). The Watson-Crick interaction energy profile closely follows that of the control $(T\cdot AT)_{11}$ triplex except for the stronger interaction energy of C-G in triplet-6 due to its three H-bond interaction in contrast to two for the A-T pair in accord with the dihedral results (Table 2).

Conclusion

The emulation of a known deoxyribose-phosphodiester triplex configuration with novel base arrangements shows promise. Previous modelling work has demonstrated the viability of unnatural Hoogsteen bases that specifically recognize the T-A major groove [8,10]. Molecular dynamics studies further confirmed the stability of these triplets with respect to T·A-T triplet [9]. With respect to the presented results, the Y2 base and its triplet constructs in both the A-type and B-type configurations appear to be structurally and energetically viable within naturally occuring triplex structures. Although the B-configuration is presently believed to be most representative of the triplex solution structure, elements of A-type geometry are not necessarily obviated. The use of the T·A-T triplet as a standard for comparison is merely a guide and not a goal. The Y2·C-G triplet studies performed here are merely a starting point for dynamic evaluation which allows a more rigorous assessment of the stability of the anticipated configuration. Having a similar stable triplet configuration to that found in nature is not a guarantee for configurational stability during dynamics simulation. However, it is likely to be a prerequisite.

With respect to molecular modelling exercises, much caution is required to assess the correctness and precision of the results. For the purposes of this study the calculated interaction energies are only utilized for comparative purposes and obviously do not represent free energies of binding. In this manner systematic inaccuracies are reduced. However, more formal modelling of solvation with respect to explicit waters and counter-ions and the application of molecular dynamics to these systems are the next obvious steps to follow this preliminary study. The final test must be synthesis and binding energy studies, but the simulations should provide encouragement for experimental work.

Acknowledgement: J. H. R. is supported by a Hitchings-Elion postdoctoral fellowship from The Burroughs-Wellcome Fund.

References

1. Strobel, S.A.; Moser, H.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1988,** *110*, 7929-9.
2. Povsic, T.J.; Dervan, P.B. *J. Am. Chem. Soc.* **1989,** *111*, 3059-61.
3. Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 544-579.
4. Pilch, D.S.; Brousseau, R.; Shafer, R.H. *Nucl. Acids Res.* **1990,** *18*, 5743-50.
5. Pilch, D.S.; Levenson, C.; Shafer, R.H. *Proc. Natl. Acad. Sci. U.S.A.* **1990,** *87*, 1942-6.
6. Pilch, D.S.; Levenson, C.; Shafer, R.H. *Biochemistry* **1991,** *30*, 6081-7.
7. Plum, E.G.; Park, Y.W.; Singleton, S.F.; Dervan, P.B.; Breslauer, K.J. *Proc. Natl. Acad. Sci. U.S.A.* **1990,** *87*, 9436-40.
8. Rothman, J.H.; Richards, W.G. *J. Chem. Soc. Chem. Comm.* **1995,** 1589-90.
9. Rothman, J.H.; Richards, W.G. *Molecular Simulation* **1995,** *18*, 13-42.
10. Rothman, J.H.; Richards, W.G. *Biopolymers* in press.
11. Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645-650.
12. Strobel, S.A.; Moser, H.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1988,** *110*, 7927-29.
13. Le Doan, T.; Perroualt, L.;Praseuth, D.;Helene, C. *Nucl.Acid Res.* **1987,** *15*, 1749-60.
14. Strobel , S.A.; Dervan, P.B. *Science* **1990,** *249*, 73-5.
15. Maher, L.J.; Wold, B.; Dervan, P.B. *Science* **1989,** *245*, 725-30.
16. Povsic, T.J.; Dervan, P.B. *J. Am. Chem. Soc.* **1990,** *112*, 9428-30.
17. Cooney, M.; Czernuszewicz, G.; Postal, E.H.; Flint, S.J.; Hogan, M.E. *Science* **1988,** *241*, 456-9.
18. Durland, R.H.; Kessler, D.J.; Gunnell, S.; Duvic, M.; Pettitt, B.M.; Hogan, M.E. *Biochemistry* **1991,** *30*, 9246-55.
19. Griffin, L.C.; Kiessling, L.L.; Beal, P.A.;Gillespie, P.; Dervan, P.B. *J. Am. Chem. Soc*. **1992,** *114*, 7976-82.
20. Griffin, L.C.; Dervan, P.B. *Science* **1989,** *245*, 967-71.
21. Koshlap, K.M.; Gillespie, P.; Dervan, P.B., Feigon, J. *J. Am. Chem. Soc.* **1993,** *115*, 7908-9.
22. Ono, A.; Ts'o, P.O.; Kan, L. *J. Am. Chem. Soc*. **1991,** *113*, 4032-3.
23. Koh, J.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1992,** *114*, 1470-8.
24. Mohan, V.; Cheng, Y.K.; Marlow, G.E.; Pettitt, B.M. *Biopolymers* **1993,** *33*, 1317-25.
25. Radhakrishnan, I.; Patel, D.J. *Structure* **1994**, *2*, 17-32.
26. Burrows, A.D., Chan, C-W, Chowdhry, M.M., McGrady, J.E., Mingos, D.M.P. *Chem. Soc. Rev.* **1995,** *24*, 329-339.

27. Arnott, S.; Bond, P.J.; Selsing, E.; Smith, P.J.C. *Nucl. Acids Res.* **1976,** *10*, 2459.
28. Umemoto, K.; Sarma, M.H.; Gupta, G.; Luo, J.; Sarma, R.H. *J.Am. Chem. Soc.* **1990**, *112*, 4539-4545.
29. Howard, F.B.; Miles, H.T.;Liu, K.; Frazier, J.; Raghunathan, G.; Sasisekharan, V. *Biochemistry*, **1992**, *31*, 10671-7.
30. Macaya, R.F.; Wang, E.;Schultze, P.; Sklenar, V.; Feigon, J. *J. Mol. Biol.*, **1992**, *225*, 755-73.
31. Raghunathan, G.; Miles, H.T.; Sasisekharan, V. *Biochemistry*, **1993**, *32*, 455-62.
32. Liu, K.; Miles, H.T.; Parris, K.D.; Sasisekharan, V. *Struct. Biol*, **1994**, *1*, 11-12.
33. Laughton, C.A. *Molecular Simulation*, **1995**, *14*, 275-89.
34. Stewart, J.J.P. *J. Comp. Aided Mol. Design* **1990,** *4*, 1.
35. Frisch, M.J.; Head-Gordon, M.; Trucks, G.W.; Foresman, J.B.; Schlegel, H.B.; Ragavachari, K.; Robb, M.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seeger, R.; Melius, C.F.; Baker, J.; Martin, L.R.; Kahn, L.R.; Stewart, J.J.P.; Topiol, S.; Pople, J.A. GAUSSIAN 90, **1990,** Revision J. Gaussian Inc., Pittsburgh, PA.
36. Chirlian, L.E.; Francl, M.M. *J. Comp. Chem.* **1987,** *8*, 894.
37. Pearlman, D.A.; Case, D.A.; Caldwell, J.C.; Siebel, G.L.; Singh, U.C.; Wiener, P.A.; Kollman, P.A., 1991, AMBER 4.0, University of California, San Francisco.
38. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, **1983,** *4*, 187.