Novel Nucleotide Bases for DNA Duplex Recognition by Triple Helix Formation

Jeffrey H. Rothman and W. Graham Richards*

Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford, UK OX1 3QZ

Molecular mechanics modelling demonstrates the geometric and energetic viabilities of proposed novel nucleotide bases in Hoogsteen interactions confined to natural DNA triplet configuration.

The wealth of new genetic sequences becoming available has thrust DNA into the role of a potential drug target. The need is for sequence-selective ligands, with one of the more promising avenues being DNA triple helix-forming oligonucleotides which recognize duplex DNA by virtue of Hoogsteen hydrogen bonding. The ability to target a broad scope of DNA sequences, the high stability of the triplex structure, and the sensitivity to single-base mismatches make this a powerful method for binding exclusive sites within large segments of duplex DNA. However, this approach is severely limited if we restrict attention to natural nucleotides, since triplex formation is limited to pyrimidine oligonucleotides binding duplex AT or GC base-pair DNA sequences at homopurine sites in the major groove parallel to the homopurine strand as T·A-T or C^{+}·G-C triplets.[1–4] Much effort has gone into the design of non-natural nucleotide bases for triplex-forming oligonucleotides,[5–7] especially those that aim to bind specifically to TA or CG base pairs *via* Hoogsteen hydrogen bonding with the same orientational geometry as the more stable known T·A-T and C^{+}·G-C triplexes.[8–12] A successful base should demonstrate favourable stacking and Hoogsteen interaction energy, and a comparable minimized phosphodiester backbone and nucleotide geometry to that of the known natural triplets. Automated ligand design routines are not well suited to binding sites that need not remain close to their stated geometry for optimal binding interactions. Non-natural bases may bind with comparable energetics with respect to known stable triplets, but are allowed conformational latitude without energetic penalty to the host triplex (owing to the relatively flexible backbone geometry). These changes within the target structure may easily eclipse or extend van der Waals contact limits for the ligand design search parameters. Here we report the successful modelling of novel unnatural nucleosides, X1 4-[ethynyl-2-(2-deoxy-β-D-ribofuranosyl)-vinyl]imidazole, and X2 4-[ethynyl-2-(2-deoxy-β-D-ribofuranosyl)vinyl]imidazol-2-one as the X1·T-A and X2·T-A triplets (Figs. 1,2).

These non-natural bases were developed from the C1′ position of the central deoxyribose of the Hoogsteen strand (strand III) using van der Waals boundaries and potential H-bonding sites of the central bases of the Watson–Crick strands (strands I and II) as guidance. Owing to the planar base-stacking constraints the solutions are limited to aromatic structures. Given the same deoxyribose–base dihedral angle as the other bases on the Hoogsteen strand, no five- or six-membered cyclic aromatic structures common to natural bases and their analogues could be placed initially at the C1′ position due to severe steric clashes with the thymine 5-Me of the middle of strand II of the Watson–Crick pair. In fact, these base structures were tipped almost perpendicular to the base-stacking plane upon minimization, thus rendering them ineffective for Hoogsteen H-bonding. Apparent solutions that avoided unfavourable steric contacts were to displace initially the bulk of the non-natural base from the C1′ deoxyribose position *via* a vinyl–ethynyl spacer. This places the Hoogsteen H-bonding C4-substituted imidazole (of X1) and imidazol-2-one (of X2) moieties in favourable positions to interact with the thymine 4-carbonyl and adenine 6-amino functionalities of strands II and I, respectively. The minimized configurations of the triplets X1·T-A and X2·T-A within the triplex are demonstrated in Figs. 1 and 2, respectively. These particular examples demonstrate comparable stacking and Hoogsteen interbase interaction energies and geometries with respect to those of a T·A-T triplet in the centre of a $(T{\cdot}A\text{-}T)_{11}$ triplex.

For these testing purposes, a host undecamer TAT DNA triplex, $(T{\cdot}A\text{-}T)_{11}$, was constructed from fibre diffraction data.[13] The centre triplet was then modified to create the binding region for testing non-natural base candidates by removing the nucleotide base of the Hoogsteen strand and switching the A-T nucleotide bases of the Watson–Crick strand to T-A, retaining the strand-specific deoxyribose–base dihedral angles. Analyses of the proposed bases within the host triplex were performed with the AMBER[14] suite of programs. For purposes of

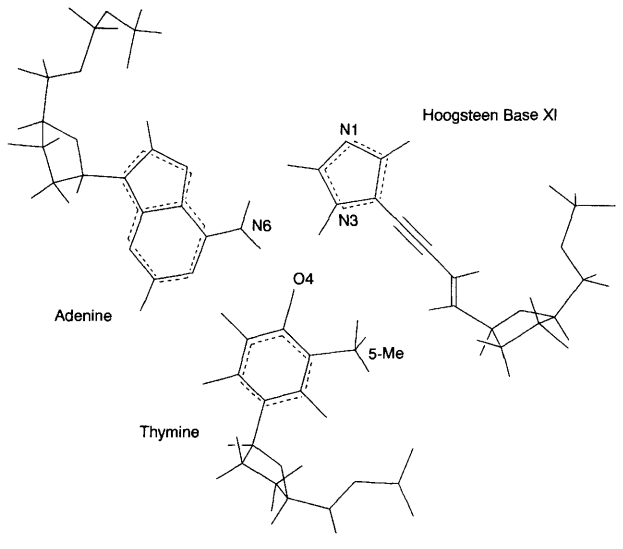

Fig. 1 Energy-minimized X1·T-A triplet configuration within the $(T{\cdot}A\text{-}T)_5$–(X1·T-A)–$(T{\cdot}A\text{-}T)_5$ triplex

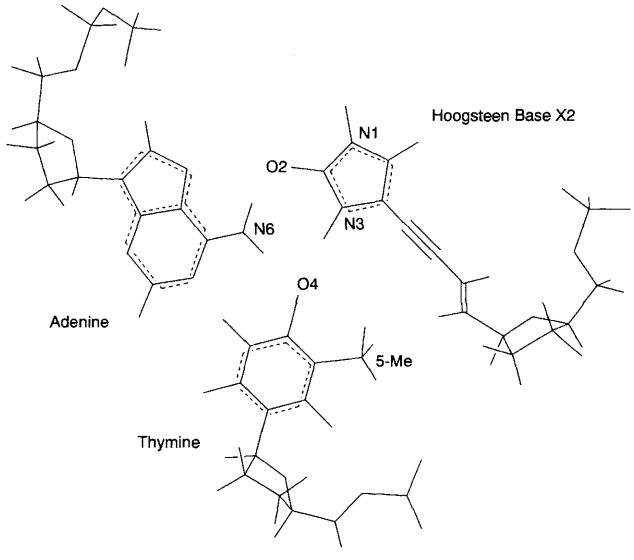

Fig. 2 Energy-minimized X2·T-A triplet configuration within the $(T{\cdot}A\text{-}T)_5$–(X2·T-A)–$(T{\cdot}A\text{-}T)_5$ triplex

calculating charges, the geometries of the proposed bases were determined with MOPAC[15] using the AM1 Hamiltonian. The charges of the MOPAC-determined structure were then obtained from GAUSSIAN90[16] with an RHF/STO-3G basis set, using the CHELPG[17] method, and scaled to the AMBER force field for nucleotides. Energy minimizations, comprising 100 steps deepest descent and 1000 steps of conjugate gradient method, were performed with a nonbonded cutoff distance of 8 Å, and a linear distance-dependent dielectric of 4 modelling the implicit solvation.

Before analysing the energetic perturbation effects on the modified triplexes, $(T{\cdot}A\text{-}T)_5$–(X1·T-A)–$(T{\cdot}A\text{-}T)_5$ and $(T{\cdot}A\text{-}T)_5$–(X2·T-A)–$(T{\cdot}A\text{-}T)_5$ due to replacement of the central triplet within the $(T{\cdot}A\text{-}T)_{11}$ triplex, the inherent perturbations in the $(T{\cdot}A\text{-}T)_{11}$ triplex model itself must be studied. Total interaction energies between bases of the Watson–Crick strands and their base-stacking interaction energies appear consistent throughout the neat triplex and insensitive to end effects. However, end effects are duly noted as deviations in Hoogsteen (strand III) base total interaction energies in only the last two triplets at each end of the neat $(T{\cdot}A\text{-}T)_{11}$ triplex. Consequently, the central nine triplets remain energetically and structurally consistent.

Upon examination of the X1·T-A triplet within the T·A-T triplex, the total interaction energy of the unnatural Hoogsteen base X1 with the T-A Watson–Crick base pair is 80%, whereas the total interaction energy between the Watson–Crick bases is 97% of that of the analogous interactions in the central T·A-T triplet of the $(T{\cdot}A\text{-}T)_{11}$ triplex (Table 1). The total Hoogsteen interaction energy profile indicates that the inserted X1·T-A triplet will additionally cause a minor perturbation of a 10% decrease in the Hoogsten total interaction energy of the adjacent triplet in the 3′ direction. Otherwise, the total Hoogsteen interaction and Watson–Crick interaction energy profiles of the bases closely follow those of the $(T{\cdot}A\text{-}T)_{11}$ triplex.

The Hoogsteen interaction of X2 with the T-A Watson–Crick base pair performs significantly better than X1 by recognizing Watson–Crick bases on strands I and II with a total interaction energy of 108% that of the analogous interactions of the natural T·A-T triplet in the corresponding position of the $(T{\cdot}A\text{-}T)_{11}$ triplex (Table 1). In fact, Hoogsteen recognition of Watson–Crick bases in natural triplets is only limited to strand II. The interaction energy between the Watson–Crick bases on this triplet is only perturbed to 95% of the value modelled in the analogous central T·A-T triplet of the $(T{\cdot}A\text{-}T)_{11}$ triplex (Table 1). Although a perturbation again exists in the Hoogsteen total interaction energy of the adjacent triplet in the 3′ direction, as it did in the X1·T-A inserted triplex, it is still only a 10% energetic decrease of that particular Hoogsteen interaction. Otherwise, the total Hoogsteen interaction and Watson–Crick interaction energy profiles closely follow those of the $(T{\cdot}A\text{-}T)_{11}$ triplex.

Table 1 Interaction energies between bases of the central triplet within the host T·A-T undecamer triplex (kcal mol^{-1}; 1 cal = 4.184 J)

Central triplet III·II-I	Interaction energy Hoogsteen(III)–WC(II)[a]	Interaction energy Hoogsteen(III)–WC(I)[a]	Interaction energy WC(II)–WC(I)[a]
T·A-T	−5.032	0.032	−5.352
X1·T-A	−3.273	−0.745	−5.189
X2·T-A	−3.151	−2.238	−5.056

[a] WC = Watson–Crick.

With respect to these preliminary results, the forementioned bases and triplet constructs appear to be geometrically and energetically viable within naturally occurring triplex structures. Additional modelling studies incorporating molecular dynamics simulations of these systems will be used to further investigate stability.

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

Received, 21st December 1994; Com. 4/07795D

References

1 H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645.
2 S. A. Strobel, H. E. Moser and P. B. Dervan, *J. Am. Chem. Soc.*, 1988, **110**, 7927.
3 P. Rajagopal and J. Feigon, *Nature*, 1989, **339**, 637.
4 R. H. Durland, D. J. Kessler and M. E. Hogan, *Biochemistry*, 1991, **30**, 9246.
5 A. Ono, P. O. Ts'o and L. Kan, *J. Am. Chem. Soc.*, 1991, **113**, 4032.
6 T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.*, 1989, **111**, 3059.
7 J. S. Koh and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 1470.
8 L. C. Griffin and P. B. Dervan, *Science*, 1989, **245**, 967.
9 V. Mohan, Y. K. Cheng, G. E. Marlow and B. M. Pettitt, *Biopolymers*, 1993, **33**, 1317.
10 H. U. Stilz and P. B. Dervan, *Biochemistry*, 1993, **32**, 2177.
11 L. C. Griffin, L. L. Kiessling, P. A. Beal, P. Gillespie and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 7976.
12 K. M. Koshlap, P. Gillespie, P. B. Dervan and J. Feigon, *J. Am. Chem. Soc.*, 1993, **115**, 7908.
13 S. Arnott, P. J. Bond, E. Selsing and P. J. C. Smith, *Nucl. Acids Res.*, 1976, **3**, 2459.
14 D. A. Pearlman, D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, P. A. Weiner and P. A. Kollman, AMBER 4.0, University of California, San Francisco, 1991.
15 J. P. Stewart, *J. Computer aided Mol. Design*, 1990, **4**, 1.
16 M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Ragavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, L. R. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol and J. A. Pople, GAUSSIAN 90, Revision J. Gaussian Inc., Pittsburgh, PA, 1990.
17 L. E. Chirlian and M. M. Francl, *J. Comput. Chem.*, 1987, **8**, 894.