Evidence for a New Spine of Hydration: Solvation of DNA Triple Helices

V. Mohan,‡ Paul E. Smith,† and B. Montgomery Pettitt*

Chemistry Department
University of Houston
Houston, Texas 77204-5641

Received March 26, 1993

Triple helices of DNA have recently come under considerable scrutiny, both experimentally[1-3] and theoretically.[4-7] Early studies showed that single-stranded oligonucleotides could be used to inhibit gene transcription in vitro.[1] This is a different approach compared to inhibition of translation or antisense reagents.[7] Expanding on this observation to target any random base sequence in a promoter region requires a thorough understanding of the microscopic reasons for stability in these systems. To this end we have performed computer simulation studies of a triple helix in explicit aqueous salt solution. We report patterns of water associations specific to triple-helical DNA. These structures are not present in duplex DNA. We have found a new spine of hydration unrelated to the one in the minor groove of TA-rich double helices in the B-form. This unique hydration structure is located in the groove formed between the first and third stands of a CGG triplex. These waters of hydration are thought to play an important role in stabilizing G-rich antiparallel triple helices at neutral to physiological pH.

Early fiber diffraction studies showed that a poly(U) stand binds to a UA duplex with the triplex-forming U strand oriented parallel (3′–5′ sense) to the purine A strand.[8] In other experimental studies, oligomers predominantly with protonated cytosines (C^+) in acidic pH solutions have also been shown to bind parallel to the G-rich strand of CG duplexes.[9-11] Promoter regions often have a highly CG-biased composition. It has recently been shown that antiparallel is the preferred orientation when G-rich synthetic oligonucleotides bind to the purine-rich strand of several duplex gene promoter sequences.[12-14] This suggests that the preference in the orientation is due to the specific base composition of both the target duplex and the triplex-forming oligonucleotide (TFO). Because of the multiplicity of hydrogen-bonding groups, Hoogsteen and reverse-Hoogsteen patterns (base orientations) are possible. Without complete empirical structural data, we previously used a modeling study[4] to help interpret some of the existing experimental observations.[1-3,7] Here, that structure[4] is used as a starting structure for a pair of computer simulations.

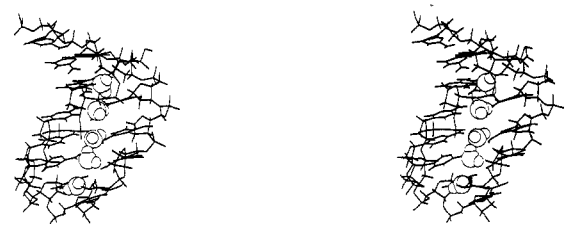

Figure 1. Spine of hydration in the M1 groove of a rigid triplex. Picture shown is a single configuration from a 100-ps simulation where the water and ions moved freely about a rigid triple helix. The second strand has been removed for clarity.

We have analyzed the results of a 100-ps and a 500-ps molecular dynamics simulation study of triple-helical DNA with explicit water, counterions, and salt using periodic boundary conditions at 300 K for a rigid and a flexible triplex, respectively. The simulations employed an antiparallel reverse-Hoogsteen homopolymeric CGG 7-mer triplex with 837 water molecules, 21 Na^+ ions, and excess salt (NaCl) at a concentration of 1 M. Triplexes are known to form in millimolar divalent cations or high concentrations of monovalent cations.[7] The Ewald method was used to compute the long-ranged electrostatic interactions. An all-atom force field[15] was employed in the present simulations. The charge on the DNA solute was $-|e|$ per phosphate group. We used the three-site simple point charge (SPC/E) water model of Berendsen *et al.*[16] The Lennard-Jones parameters for Na^+ and Cl^- were adopted from Chandrasekhar *et al.*[17]

Current knowledge on triple helix formation in a CGG system[7] suggests that the third G strand binds in the major grove of a CG duplex in a reverse-Hoogsteen fashion. This forms two new grooves; we denote M1 as the groove between the first or C strand and the third G strand and M2 as the groove between the second and third strands of G.

In the M1 groove, we found a series of water molecules bound between the NH_2 of the third-strand G and the NH_2 of C at each base plane. These water molecules were not placed in these positions initially. The initial water placement, starting from the idealized rigid structure obtained from modeling studies, followed the traditional route of overlaying a sample of preequilibrated water, removing any waters that overlap the DNA within 1.8 Å, and then randomly placing the salt ions by replacing a water with each ion. Formation of the M1 spine of water was complete after the first 10 ps of equilibration, and this water structure remained stable for the rest of the rigid simulation. A representative structure is shown in Figure 1. Subsequently, starting from the end of the first simulation, an additional 500 ps of dynamics was performed, allowing the DNA (and all components) to move freely. Again, the spine of hydration remained intact. The spine waters of the short helix exchanged infrequently during the simulations, maintaining the spine's overall probability density. The structure can be seen in Figure 2. Radial distribution functions between the N2 of each guanine and water oxygen and between the N4 of cytosine and water oxygen were computed and showed a first peak at 3.1 and 2.9 Å, respectively. Both peaks integrated to nearly one water molecule per base plane. On further analysis of the contributions to those peaks, it was found that different individual water molecules were responsible. In fact, 10 waters were found for the central five base planes that were within 3.5 Å of either N4 of C or N2 of G.

We also calculated the NMR rotational correlation times for both bulk waters and the 10 closest waters bound to the C and

‡ Present address: Isis Pharmaceuticals, Carlsbad, CA.
† Present address: Physical Chemistry, ETH-Zentrum, Zürich, Switzerland.

(1) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456.
(2) Hélène, C.; Toulmé, J. J. *Biochem. Biophys. Acta* **1990**, *1049*, 99.
(3) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081.
(4) Cheng, Y. K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465. van Vlijmen, H. W. Th.; Ramé, G. L.; Pettitt, B. M. *Biopolymers* **1990**, *30*, 517.
(5) Laughton, C. A.; Neidle, S. *J. Mol. Biol.* **1992**, *223*, 519; *J. Chem. Phys.* **1991**, *88*, 2597; *Nucleic Acids Res.* **1992**, *95*, 8430.
(6) Hausheer, F. H.; Singh, U. C.; Palmer, T. C.; Saxe, J. D. *J. Am. Chem. Soc.* **1990**, *112*, 9468. Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Colvin, O. M.; T'so, P. O. P. *Anti-Cancer Drug Design* **1990**, *5*, 159. Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Flory, J. P.; Tufto, K. B. *J. Am. Chem. Soc.* **1992**, *114*, 5356.
(7) Cheng, Y. K.; Pettitt, B. M. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225.
(8) Arnott, S.; Selsing, B. E. *J. Mol. Biol.* **1974**, *88*, 509. Arnott, S.; Bond, P. J.; Selsing, B. E.; Smith, P. J. C. *Nucleic Acids Res.* **1976**, *3*, 2459.
(9) Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73.
(10) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967.
(11) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.
(12) 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; pp 565–578.
(13) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246.
(14) Moser, H. E.; Dervan, P. B. *Science* **1991**, *251*, 1360.
(15) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187; Charmm 22 set.
(16) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
(17) Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903.

0002-7863/93/1515-9297$04.00/0

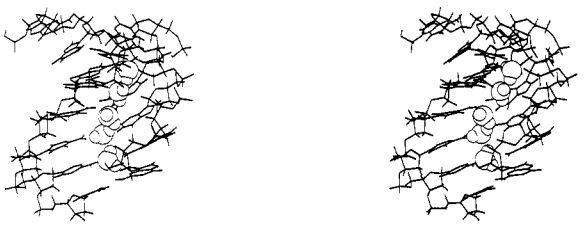

Figure 2. Spine of hydration in the M1 groove for a flexible triplex. This figure is a single configuration taken from a 500-ps simulation where all atoms moved dynamically. The second strand has been removed for clarity.

G exocyclic amino groups. The value of the bulk water in the simulation is 12 ps, somewhat higher than the saltwater bulk value of around 6 ps.[22] The corresponding value for the spine waters is 30 ps, significantly higher than the bulk, indicating a substantial restriction in reorientational motions. As such, those water molecules are able to help enthalpically to stabilize the triplex structure. An orientational preference is found for the water close to the helix axis. For waters in the M1 and M2 groove, this corresponds to an average angle of the water dipole vector with respect to the helix axis of about 130°. In molecular terms, there is a preference for having one of the OH bond vectors parallel to the helix axis. This arrangement is most prominent in the M2 groove.

In duplex DNA, the spine of waters in B-form TA-rich samples lines the minor groove, as seen by X-ray diffraction[19] and for more general cases, including GC tracts, by Monte Carlo simulations,[18] and is thought to help enthalpically to stabilize the system. The agreement between simulation and experiment in that case was essentially quantitative. Similarly, in GC-rich duplexes, filaments of observed water have been proposed as stabilizing B-form structures.[19] Those water molecules also have been implicated in the structural transitions of the B-form of DNA.[20] Unfortunately, far less structural data exist on triple helices, and no 3D crystal structure has been reported at atomic resolution. Based on our computer simulations, the spine of hydration in the M1 groove and possibly that in the M2 groove of triple-helical DNA play a role in stabilizing the CGG system. This feature in the M1 groove is unique to the CGG triple helix and would not be present in parallel or antiparallel TAT systems. It is due to the unique positioning of the amino groups of the first-strand C and the third-strand G. We notice that while the minor groove is electrostatically negative, the M1 groove is filled with hydrogen-bond donors.

The presence of organized water molecules may play a significant role in determining the different stabilities between pure CGG systems and samples where TAT and other combinations are interdispersed. It is known that a single mismatch in a CGG triple helix destabilizes the system by a factor of 10.[21] Given that kind of sensitivity, understanding the details of solvation effects is crucial for determining the full set of triple helix formation principles.

Acknowledgment. The authors acknowledge the National Institutes of Health, the Texas Coordinating Board through the Advanced Technology Program, Triplex Pharmaceutical Corporation, and the Robert A. Welch Foundation for partial support of this work. Profs. M. Hogan and Lian Gao and Mr. Jason Cheng are thanked for many useful conversations. The authors thank SDSC and the NSF chemical instrumentation program for computational support.

(18) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535.
(19) Beveridge, D. L.; Subramanian, P.; Jayaram, B.; Swaminathan, S. In *Theoretical Biochemistry and Molecular Biophysics*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: New York, 1990; p 17.
(20) Saenger, W. *Principles of Nucleic Acids Structure*; Springer Verlag: New York, 1983.
(21) Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab, F. *FASEB J.* **1988**, *2*, 2939.
(22) Smith, P. E.; Pettitt, B. M. *J. Chem. Phys.* **1991**, *95*, 8430.