Sequence Dependent Hydration of DNA: Theoretical Results

Adrian H. Elcock* and J. Andrew McCammon

Department of Chemistry and Biochemistry
University of California at San Diego
La Jolla, California 92093-0365

Received July 6, 1995

Hydration effects are crucial to the static and dynamic behavior of DNA and its interactions with other molecules.[1] DNA hydration has been studied by a variety of experimental methods: X-ray crystallography,[2] NMR spectroscopy,[3] volumetric, and densitometric techniques[4] have all been employed. Despite such studies, however, there is little experimental data regarding specifically the thermodynamics of hydration, largely because of the difficulties in separating out effects due solely to hydration from those due to other causes. Drug−DNA binding equilibria, for example, are governed by the energetics of the drug−DNA interaction itself, conformational changes in the drug and/or the DNA, and effects due to changes in the ionic atmosphere, in addition to those resulting solely from changes in hydration. In this paper we use a theoretical method to focus entirely on this latter aspect and report on the sequence dependence of solvation free energies calculated for DNA oligonucleotides.

The Poisson−Boltzmann (PB) equation has been widely used to model electrostatic effects in and around macromolecules,[5] and a number of applications to DNA systems have been reported.[6] It has also been used to estimate solvation effects, and good agreement between small molecule hydration free energies calculated with the PB equation and with the free energy perturbation (FEP) method has been obtained.[7] Owing to the highly-charged nature of DNA, it is to be expected that the PB equation will also provide a good description of solvation effects in DNA systems.

Electrostatic solvation energies were calculated for a number of DNA oligonucleotides using the UHBD (University of Houston Brownian Dynamics) program.[8] Oligonucleotide structures were built in a canonical B-DNA conformation[9] using the molecular modeling program QUANTA.[10] The sequence chosen for study was the 9-mer 5′-d[CGC(XYZ)CGC]·5′-d[GCG(Z′Y′X′)GCG], where X, Y, and Z represent any of the four bases A, C, G, or T and X′, Y′, Z′ represent their respective complementary bases. Structures for all possible sequences of the three central bases ($4^3 = 64$) were investigated. Atomic charges and radii for the DNA atoms were obtained from the OPLS parameter set developed by Jorgensen and co-workers;[11] polar hydrogens were assigned a radius of 1.25 Å. The use of these parameters for the isolated nucleic acid bases has been shown to give good agreement with the results of FEP calculations carried out with the same parameter set.[12] The relative dielectrics of the solute and solvent regions were set to 1 and 78, respectively, while the ionic strength was set to 0.15 M. The nonlinear Poisson−Boltzmann equation was solved in two steps using "focusing",[13] the first grid consisting of 55^3 grid points spaced by 1.6 Å, the second, a 110^3 grid spaced by 0.35 Å. Additional calculations[14] performed to assess the effects of different grid spacings on the results suggest that errors in the sequence dependence of the hydration free energies average ~0.4 kcal·mol^{-1}.

Table 1. Electrostatic Solvation Free Energies of Oligonucleotides of Sequence 5′-d[CGCXYZCGC]·5′-d[GCGZ′Y′X′GCG] Relative to the Value Obtained for the Sequence with XYZ = AAA (−4523.0 kcal·mol^{-1})[a]

sequence	A	C	G	T
A_A	0.0	−2.0	−3.3	−0.9
A_C	−3.1	−7.3	−5.2	−4.3
A_G	−2.4	−1.2	−5.8	−1.9
A_T	−0.6	−3.5	−3.2	−0.9
C_A	−3.9	−8.1	−4.5	−5.6
C_C	−7.0	−13.2	−6.3	−9.2
C_G	−6.0	−7.1	−7.0	−6.5
C_T	−4.6	−9.3	−4.4	−5.4
G_A	−0.6	−1.5	−4.2	−0.5
G_C	−3.9	−6.8	−6.5	−4.0
G_G	−2.9	−0.5	−6.8	−1.5
G_T	−1.4	−2.8	−4.2	−0.2
T_A	−1.4	−4.0	−3.6	−1.5
T_C	−4.6	−9.1	−5.4	−4.9
T_G	−4.0	−3.2	−6.0	−2.5
T_T	−2.1	−5.2	−3.3	−1.3

[a] Columns give the identity of the central base (Y). Rows refer to the flanking bases (X and Z)

Table 1 shows the electrostatic free energies of hydration for each of the sequences studied, relative to the value obtained for a central sequence of AAA (this sequence having the least favorable solvation energy). Table 2 presents the same data differently so that the effects of mutating the central base of a given triplet can be more easily seen. For the first six rows in Table 2 there are two entries: some of the sequences studied are essentially identical (a central ATC sequence is the same as the complementary strand of GAT, for example) and differ only in the regions flanking the central trinucleotide (i.e., 5′-dCGC**ATC**CGC vs 5′-dGCG**GATC**GCG). Similarly, for the last four rows, the first and fourth columns should be compared, as should the second and third. These comparisons provide an indication both of the effects of bases outside the central triplet and of the effects of the use of a finite grid representation on the hydration energies obtained. Most differences are only minor (average difference 0.3 kcal·mol^{-1}, maximum difference

(1) (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York. (b) Berman, H. M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 345.
(2) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535.
(3) (a) Kubinec, M. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 8739. (b) Liepinsh, E.; Otting, G.; Wüthrich, K. *Nucleic Acids Res.* **1992**, *20*, 6549.
(4) (a) Rentzeperis, D.; Kupke, D. W.; Marky, L. A. *Biopolymers* **1993**, *33*, 117. (b) Rentzeperis, D.; Kupke, D. W.; Marky, L. A. *Biopolymers* **1992**, *32*, 1065. (c) Chalikian, T. V.; Sarvazyan, A. P.; Plum, E.; Breslauer, K. J.; *Biochemistry* **1994**, *33*, 2394.
(5) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(6) (a) Jayaram, B.; Sharp, K. A.; Honig, B. *Biopolymers* **1989**, *28*, 975. (b) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245. (c) Misra, V. K.; Hecht, J. L.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 264. (d) Zacharias, M.; Luty, B. A.; Davis, M. E.; McCammon, J. A. *Biophys. J.* **1992**, *63*, 1280.
(7) (a) Arald, J. C.; Nicholls, A.; Sharp, K. A.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454. (b) Ewing, T. J. A.; Lybrand, T. P. *J. Phys. Chem.* **1994**, *98*, 1748.
(8) (a) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. *Rev. Comput. Chem.* **1994**, *5*, 229. (b) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.*, submitted.
(9) Arnott, S.; Chandrasekeharan, R.; Birdsall, D. L.; Leslie, A. G. W.; Ratliffe, R. L. *Nature* **1980**, *283*, 743.
(10) QUANTA version 4.0, Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA (1994).
(11) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810.
(12) (a) Elcock, A. H.; Richards, W. G. *J. Am. Chem. Soc.* **1993**, *115*, 7930. (b) Mohan, V.; Davis, M. E.; McCammon, J. A.; Pettitt, B. M. *J. Phys. Chem.* **1992**, *96*, 6428.
(13) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1987**, *12*, 435.
(14) The expression derived for the electrostatic free energy of a system described by the nonlinear Poisson−Boltzmann equation (Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684) contains terms evaluated as volume integrals. The total calculated energy may therefore depend not only on the grid spacing but also on the overall volume of the grid.

0002-7863/95/1517-10161$09.00/0

Table 2. Electrostatic Solvation Free Energies of Oligonucleotides with Central Sequence XYZ Relative to the Value Obtained for the Sequence XAZ[a]

sequence	A	C	G	T
A_A	0.0	−2.0	−4.2	−1.4
		−2.0	−4.2	−1.0
A_C	0.0	−4.2	−2.5	−0.9
		−4.0	−2.8	−1.1
A_G	0.0	+0.8	−4.5	0.0
		+1.0	−4.4	+0.4
C_A	0.0	−4.2	−1.3	−1.7
		−3.5	−0.9	−1.0
C_C	0.0	−6.2	+0.4	−1.3
		−5.3	+0.6	−1.1
G_A	0.0	−0.9	−4.8	−0.5
		−0.5	−4.9	0.0
A_T	0.0	−2.9	−2.6	−0.3
C_G	0.0	−1.1	−1.0	−0.5
G_C	0.0	−2.9	−2.6	+0.1
T_A	0.0	−2.6	−2.2	−0.1

[a] Columns give the identity of the central base (Y). Rows refer to the flanking bases (X and Z). Where two entries are given for the same central sequence, the top value refers to the sequence 5′-d[CGCXYZCGC] and the bottom value refers to the sequence 5′-d[GCGXYZGCG.

0.9 kcal·mol^{-1}), a result which suggests that, with regard to electrostatic hydration free energies, it is probably sufficient to consider only a trinucleotide sequence to estimate sequence-dependent effects; that is, the hydration of a central base pair is only strongly affected by the immediately adjacent base pairs.

A number of interesting points are apparent from the results. In general, for example, the replacement of A by C or G results in a more favorable hydration free energy, a result which is in line with the hydration free energies of the isolated bases.[12] This preferential hydration associated with G and C is seen to be particularly pronounced when substitution of the central A results in formation of a pair of adjacent G's (or C's) and is still more pronounced when it results in a triplet run of G's. There are, however, cases where such an effect is not obtained. The substitution CAC → CGC results in a hydration energy *less* favorable by about 0.8 kcal·mol^{-1}, while the change AAG → ACG results in a similar unfavorable change of ~1.1 kcal·mol^{-1}. These effects can be understood by examining the spatial arrangement of functional groups in the structures. The more favorably hydrated sequences are characterized by having functional groups of similar polarity grouped close together (carbonyl with carbonyl, for example). The closer grouping of like-charged atoms is expected to be unfavorable from the point of view of intrinsic DNA stability, but more favorable from the perspective of solvation. Effects such as these mirror the variation in gas-phase electrostatic potentials calculated previously[15] and are similar in spirit to the secondary electrostatic effects noted as being important in determining base-pair binding affinity.[16]

The general finding that GC base pairs are associated with more favorable hydration free energies is in line with recent experimental results which show that these base pairs cause an increase in water density relative to AT base pairs.[4c] An apparent discrepancy is, however, obtained in the case of central sequences containing *only* AT base pairs. In the present results, substitution of any of these bases with C or G results in a more favorable hydration free energy. Experimental work[4c] suggests, however, that sequences with 100% AT content are more hydrated than sequences with 55−60% AT content (though less hydrated than 100% GC sequences). The authors of that work emphasize the difficulties of interpreting macroscopic experimental data in microscopic terms, and it does not necessarily follow that apparent molar volume will correlate with hydration free energy. Nevertheless, this result may point to inadequacies in the approach adopted here. On the other hand, the neighboring base-pair dependencies found here are in accord with the experimental finding that DNA hydration is not simply a function of the percentage content of AT and GC base pairs.

The results reported here provide a comprehensive list of the effects of base-pair substitutions on the hydration free energy of a DNA oligonucleotide. The finding that substitution of A by C or G is not always favorable shows that in order to understand such sequence-dependent hydration effects it is not sufficient to consider only a single base pair; instead, attention must be paid to the identity of the adjacent base pairs in both directions. It is, however, important to stress the weaknesses of the current approach. Firstly, no account has been made of nonelectrostatic factors such as the hydrophobic effects believed to attend the thymine methyl group; assuming that these are proportional to the solvent-accessible surface area of the molecule,[17] calculations suggest that such effects should be slight.[18] Secondly, while the present results were obtained with the OPLS parameter set, there is as yet no consensus as to which of the many molecular mechanics parameter sets is most suited to treating DNA; similar trends were, however, obtained with the CHARMM parameter set.[19] Finally, the assumption that all the oligonucleotides conform to a standard B-DNA structure may not be appropriate; alternating AT sequences, for example, are believed to adopt an "alternating B-DNA" structure in solution.[20] The results reported here are, however, expected to be relatively insensitive to minor structural changes. Despite these notes of caution, the results presented here should provide a good estimate of sequence-dependent hydration effects and as such are expected to be of significance to studies of processes which involve changes in DNA solvation, such as ligand−DNA and protein−DNA interactions.

Acknowledgment. This work was supported in part by grants from the National Institutes of Health and the National Science Foundation MetaCenter Program. A.H.E. is very grateful to the Wellcome Trust for financial support under Grant 041157/Z/94/Z.

JA9522140

(15) Pullman, A.; Pullman, B. *Q. Rev. Biophys.* **1981**, *14*, 289.
(16) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.
(17) Honig, B.; Sharp, K. A.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101.
(18) Of the sequences studied here, most have total surface areas within 20 Å^2 of each other, with the maximum difference being 50 Å^2. Assuming a constant of proportionality of 5 cal·mol^{-1}·A^{-2} (Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978), this represents a maximum difference of only 0.3 kcal·mol^{-1}.
(19) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Nguyen, D. T.; Ngo, T.; Prodhom, B.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Wiorkiewicz-Kuczera, J.; Karplus, M. *FASEB J.* **1992**, *6*, A143.
(20) Klug, A.; Jack, A.; Viswamitra, M. A.; Kennard, O.; Shakked, Z.; Steitz, T. A. *J. Mol. Biol.* **1979**, *131*, 669.