

Theoretical considerations on the "spine of hydration" in the minor groove of d(CGCGAATTCGCG)-d(GCGCTTAAGCGC): Monte Carlo computer simulation

P. S. SUBRAMANIAN, G. RAVISHANKER, AND D. L. BEVERIDGE

Chemistry Department, Hall-Atwater Laboratories, Wesleyan University, Middletown, CT 06457

Communicated by Max Tishler, October 8, 1987

ABSTRACT A theoretical description of aqueous hydration in the minor groove of a B-form DNA is presented on the basis of a liquid-state Monte Carlo computer simulation on a system consisting of the oligonucleotide duplex d(CGCGAATTCGCG)-d(GCGCTTAAGCGC) in a canonical B-form together with 1777 water molecules contained in a hexagonal prism cell and treated under periodic boundary conditions. The results are analyzed in terms of solvent density distributions. The calculated minor-groove solvent density shows considerable localization, indicative of discrete solvation sites and providing theoretical evidence for a well-defined ordered water structure. In the AATT sequence, this corresponds to the "spine of hydration" described by H. R. Drew and R. E. Dickerson [(1981) *J. Mol. Biol.* 151, 535-556] based on the x-ray crystal structure of the dodecamer hydrate. We find, however, that the calculated ordered water structure also extends into the CGCG flanking sequences, supported by the N² hydrogen bond donors of the guanine residues and indicating that the spine of hydration could thus extend throughout the minor groove of a B-form DNA. This provides a possible explanation of the positive binding entropies observed by L. A. Marky and K. J. Breslauer [(1984) *Proc. Natl. Acad. Sci. USA* 84, 4359-4363] for both A·T and C·G sequences on the complexation of netropsin to the minor groove of DNAs. Implications of these results with regard to the thermodynamic stability of DNA in water and the sequence specificity of the minor groove hydration are discussed.

The DNA molecule is now well known to exist in a variety of conformational families, both right-handed (A, B, C, D, etc.) and left-handed (Z_I, Z_{II}) helical duplexes (1). The relative stability of the various conformational forms of DNA is observed to be highly sensitive to environmental effects—i.e., hydration and ionic strength. Hydration, dehydration, and reorganization of the ion atmosphere are also important in the thermodynamics of protein binding to DNA and in drug-DNA interactions (2). Whereas a considerable literature on the environmental effects on DNA exists (1), our knowledge of detail at the molecular level is fairly sparse. Our principal source of information to date comes from the positions of ordered water molecules observed in x-ray crystallography of DNA oligonucleotides (3). Particularly, the dodecamer duplex d(CGCGAATTCGCG)-d(GCGCTTAAGCGC) has been found by Dickerson and coworkers (4-6) to crystallize as a hydrate in the B-form of DNA, which corresponds closely to the Watson-Crick double helix and features a distinct parallel groove motif, alternating wide (major) with narrow (minor). Although the amount of water that turned out to be crystallographically ordered is small (≈25%), interesting and provocative features of the hydration emerged, as discussed in detail by Drew and Dickerson

(ref. 7, but see also ref. 8). In particular, an ordered water structure was discovered in the minor groove of the AATT region. It has been indicated that this so-called "spine of hydration" was specific to A·T-rich tracts, where the N-3 atom of adenine and the O² atom of thymine are readily available as hydrogen bond acceptors in interactions with water molecules. The N² donor group on guanine has been thought to disrupt the spine, but hydration of the C·G region of the dodecamer was blocked in the crystal by a spermine ion and by helix-helix packing and thus was not observed.

The spine of hydration has been subsequently considered (9) to be a central stabilizing feature of the B-DNA structure, and that disruption of the spine as by dehydration would effect a conformational change to the A- or Z-form, depending on sequence and other environmental conditions. Sequences rich in A·T base pairs are known to preferentially stabilize B-form DNA (10), and the spine of hydration has provided a possible explanation (9). Further evidence for the stabilizing nature of the spine comes from an experiment based on CD spectra, whereby an A-form calf thymus DNA in complex with the minor groove binding molecule netropsin is pulled into a B-like form (11). Thermodynamic binding studies by Breslauer and coworkers (12, 13) on netropsin binding to the poly[d(A-T)]-poly[d(A-T)] duplex revealed a positive binding entropy that has been associated (14) with the Dickerson water spine. However, Marky and Breslauer (15) have found a similar binding entropy for formation between netropsin and the poly[d(G-C)]-poly[d(G-C)] duplex and suggest alternative interpretations of the entropy data involving non-spine-hydration shells less sensitive to sequence or binding-induced release of counterions.

CALCULATIONS

To gain additional perspective on the nature of the proposed spine of hydration in DNA, we have carried out (T, V, N) ensemble Monte Carlo-Metropolis computer simulation by using a modified version of the program MMC (16) on a system consisting of d(CGCGAATTCGCG)-d(GCGCTTAAGCGC) in the canonical B-form (17) conformation and 1777 water molecules, a number chosen to provide an excess of two solvation shells for the solute. The calculations were carried out for a hexagonal prism central cell (Fig. 1) under periodic boundary conditions, and thus comprise a representation of a dilute aqueous solution or the aqueous hydration of the dodecamer. The volume of the system was taken to be consistent with an environmental density of 1 gm/ml and the temperature was 298 K. The configurational energies of the system were evaluated under the assumption of pairwise additivity by using atom site potentials, the TIP4P function was developed by Jorgensen *et al.* (18) for water-water interactions, and TIP4P was spliced with the coulomb and van der Waals terms in the AMBER FF2 force field of Weiner *et al.* (19) for water-solute interactions. A spherical cutoff at 7.5 Å was applied in the evaluation of water-water interaction energies, and solute-water interactions were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

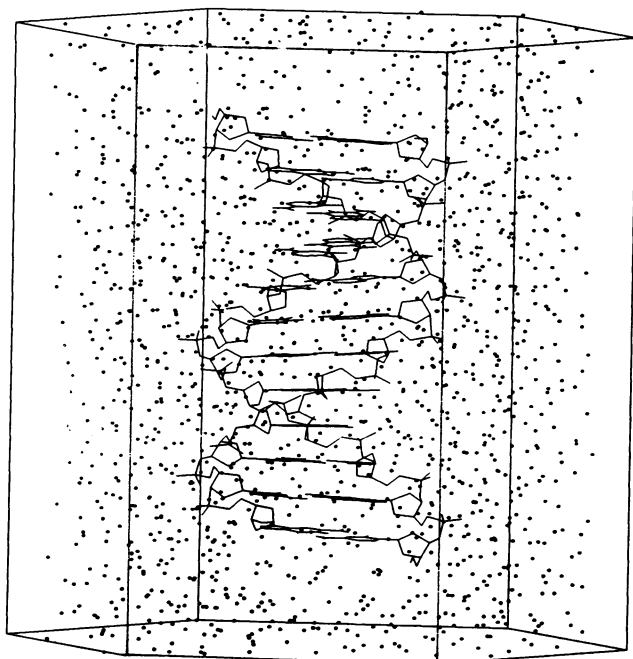


FIG. 1. The hexagonal prism elementary cell used in Monte Carlo simulations on hydrated $d(\text{CGCGAATTCGCG})\text{-}d(\text{GCGCTTAAGCGC})$. The dots surrounding the dodecamer indicate the center-of-mass positions of solvent water molecule in the initial, randomly chosen, configuration.

treated under the minimum image convention. A somewhat similar approach to potentials for this system was used earlier in a molecular dynamics simulation on the dodecamer by Seibel *et al.* (20). Electroneutrality was established by uniformly scaling the nucleic acid charges. This was found to alter any single atomic charge only in the second decimal place, so that the phosphate group remains substantially electronegative, *ca.* -0.8 atomic units of charge. The simulation was first allowed to proceed for 1,500,000 configurations of equilibration, and the ensemble averages were formed over the next 1,500,000 configurations of the realization. Force bias (21) and preferential sampling (22) were applied to accelerate the Monte Carlo convergence (23). All calculations were carried out on the Cray X-MP/48 supercomputer at the Pittsburgh Supercomputer Center, at a sampling rate of $\approx 100,000$ configurations per hr. Files were returned locally for display and analysis on an Evans and Sutherland PS-350 color graphics unit by using the comprehensive molecular graphics program "DOCK" (24).

A full description of the aqueous hydration of our model dodecamer is possible from this simulation. Detailed analysis on first-shell solvent coordination, on solute-water pair interaction, and on solute binding energies as well as solvent density distributions, all partitioned by using the proximity criterion (25, 26) into contributions from the major groove, minor groove, and sugar-phosphate backbone of the DNA will be presented elsewhere (P.S.S. and D.L.B., unpublished results). The solvent density plots for the minor groove of the dodecamer have revealed a considerable amount of particular detail relevant to the nature of the spine of hydration and are thus particularly reported and discussed herein. All results are subject to the approximations in our calculations, particularly the choice of canonical B-form for the dodecamer, the assumed intermolecular potential functions, and implicit treatment of counterions.

RESULTS AND DISCUSSION

The solvent density in the minor groove is depicted for presentation here as the linear superposition of 16 individual

configurations (snapshots) collected at equally spaced intervals along the production segment of the Monte Carlo realization. Water molecules found within the first-shell radii of the N-9, C-4, N-3, C-2, and N-1 atoms of adenine, of the N-3, O², and N-1 of thymine, of the N-9, C-4, N-3, C-2, and N² of guanine, and of the O² and N-1 of cytosine were assigned as the first coordination shell of the minor groove. A second shell comprised of all water molecules forming a primary coordination to first-shell waters was defined to further elaborate the minor-groove hydration. Proximity analysis turned up ≈ 50 waters of hydration in the minor groove of the dodecamer, 12 primary and 38 secondary with respect to the bases.

The calculated solvent density in the minor groove of the duplex of our model $d(\text{CGCGAATTCGCG})\text{-}d(\text{GCGCTTAAGCGC})$ dodecamer is shown in Fig. 2. The small circles outside the nucleic acid are the positions of the oxygen atoms of the solvent water molecules, with filled points indicating the first-shell coordination running along the floor of the minor groove and the open points indicating the position of second-shell groove waters (some of these second-shell groove waters may reside in a first-shell coordination with atoms of the sugar-phosphate backbone). An individual point simply indicates a water molecule is present in one or another of the solvent configurations contributing to the statistical state of the

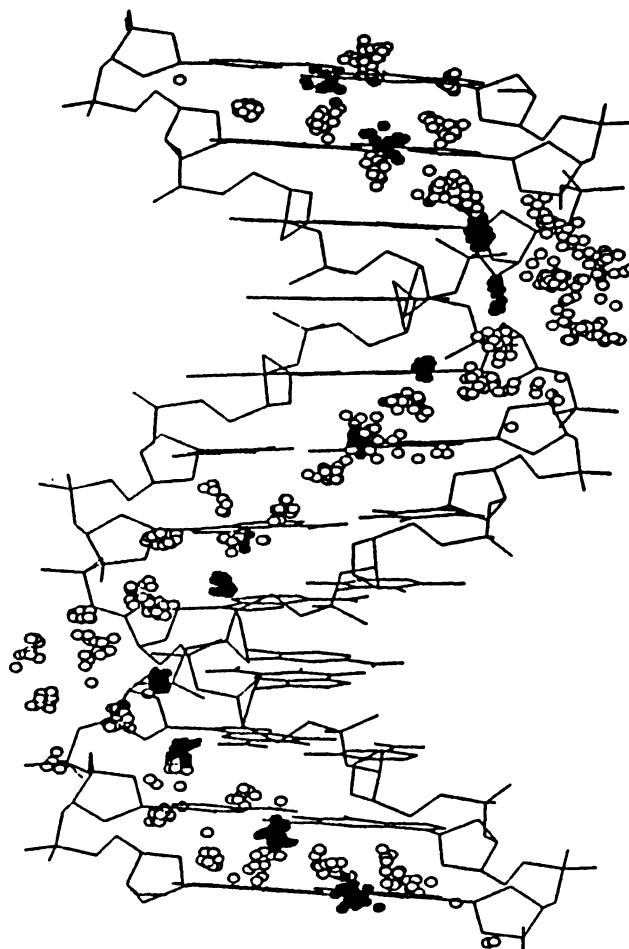


FIG. 2. Calculated hydration density in the minor groove of $d(\text{CGCGAATTCGCG})\text{-}d(\text{GCGCTTAAGCGC})$ presented as a superposition of water oxygen positions obtained from a sequence of 16 configurations extracted from equally spaced intervals along the simulation. Solid circles indicate first-shell waters of hydration with respect to nucleotide bases, and open circles indicate the remaining minor-groove waters.

system. The important information conveyed by this figure is the clustering of points, indicating the concentration of solvent density in that region. The calculated minor-groove solvent density for the model dodecamer (Fig. 2) shows considerable localization and provides theoretical evidence for the existence of discrete hydration sites in the minor groove and for the existence of a well-defined ordered water structure in this region corresponding to Dickerson's observed spine of hydration.

The nature of the calculated spine is further revealed by a display of waters in the first shell of the minor groove and the nucleotide bases of the dodecamer (Fig. 3). The first-shell coordination consists of water molecules hydrogen bonded to the polar atoms along the floor of the minor groove, and more or less situated in the plane of the nucleotide base pairs. By contrast, the first-shell waters in the crystallographic spine of hydration were found to bridge the O and N atoms of successive base pairs. Some discrepancies in the calculated and observed results are expected due to the difference between the canonical B-form and the crystallographic dodecamer structure as well as the approximations inherent in the intermolecular force field. The preferred calculated hydration site for an A·T base pair turns out to be the adenine N-3 acceptor site. In the C·G region the guanine N² donor site is clearly favored over the cytosine O² acceptor site. The localization of solvent density is a bit more pronounced in the C·G than in the A·T region, reflect-

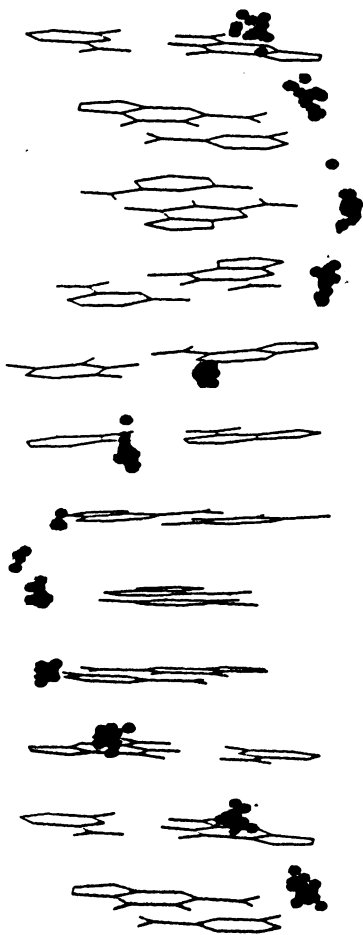


FIG. 3. Calculated hydration density along the floor of the minor groove of d(CGCGAATTCGCG)-d(GCGCTTAAGCGC)—i.e., the first hydration shell of groove waters for the nucleotide bases. Only the nucleotide bases of the model dodecamer are shown to simplify the illustration.

ing the more demanding geometrical restrictions of a donor compared with an acceptor site on the solute (28).

The calculated spine of hydration can be traced out by following the hydrogen-bonded water network up the minor groove. The calculated spine, as in the crystal structure, is not just a first-shell entity but involves for the most part an alternation of first-shell and second-shell water molecules of the minor groove. This can be more clearly seen in Figs. 4 and 5, where snapshots of the hydration of the minor groove are shown in molecular detail with the nucleic acid "trellis" removed. Note that once the second shell is involved the spine is not necessarily a single path, but an incipient water network in which the motif of tetrahedral coordination characteristic of liquid water coordination is quite discernible. This point was noted in the discussion of the crystal structure as well (ref. 7, but see also ref. 8), even though some of the participating waters were crystallographically disordered and thus not observed. The water network in the C·G region (Fig. 5) shows some interesting water polygons of order 4 and 5.

An important result from our calculation is the extent to which the C·G as well as the A·T region can support an ordered water structure in the minor groove. We find the C·G region, with the floor of the minor groove lined with

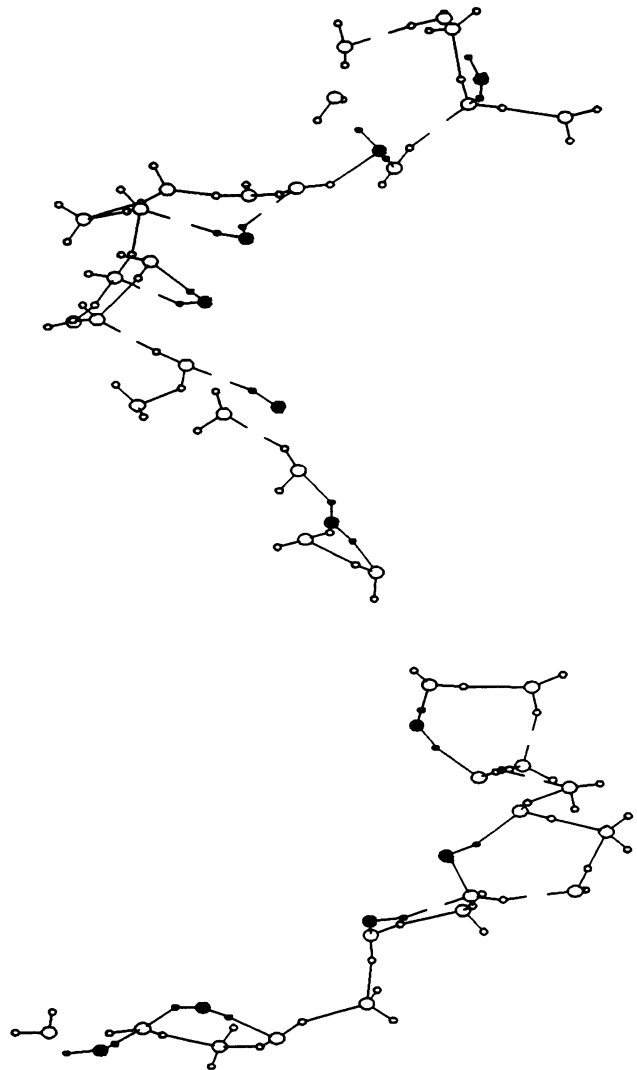


FIG. 4. A single snapshot view of the calculated ordered water structure in the minor groove of d(CGCGAATTCGCG)-d(GCGCTTAAGCGC), with only the water depicted.

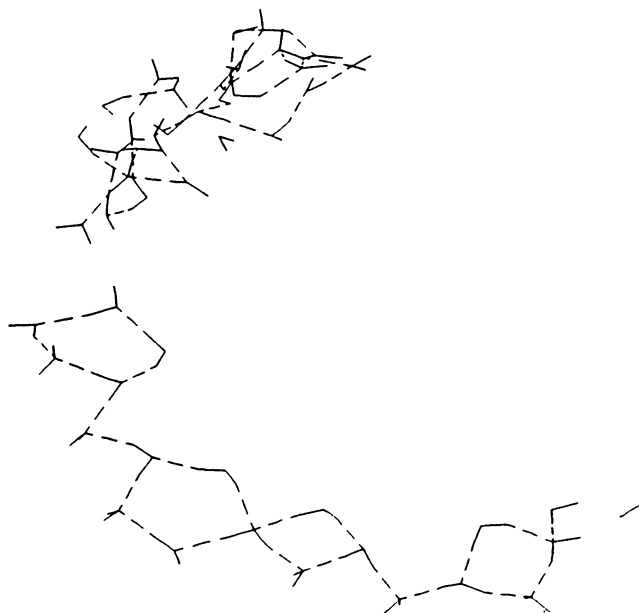


FIG. 5. Another view of the structure presented in Fig. 3, presented as a framework structure and rotated to show the water network in CG flanking sequence.

guanine N² donor groups and cytosine O² acceptor sites, to be just as capable of supporting a spine of hydration as is an A·T region. The calculated spine of hydration in the minor groove of our model dodecamer extends completely from one end of the structure to the other, with the interface between the C·G and A·T regions dealt with easily by the geometric flexibility of water–water hydrogen bonding interactions. The penetration of water into the DNA is of course greater for A·T tracts. Thus our calculations support the general idea of the spine of hydration, but indicate that the spine is not necessarily specific to A·T-rich regions as originally suspected. An ordered water network in the minor groove of DNA supported by C·G base pairs provides an attractive alternative explanation of the positive binding entropies for netropsin complexed with poly[d(G-C)]·poly-[C-G] observed by Marky and Breslauer (15).^{*} The role of the Dickerson spine of hydration in the preferential stabilization of B-DNA specific for A·T-rich sequences also needs to be reconsidered in light of this result.

A cautionary note on the relationship of the spine of hydration to thermodynamic stability of the DNA duplex has been aired by our group (29). We pointed out that ordered water in the minor groove may be in an energetically favorable state but is entropically unfavorable [enthalpy–entropy compensation (30)] and that more disordered water in the major groove, quite capable of favorable energetics, could well be as important to the stability of B-DNA as the minor-groove spine of hydration. At this point we must emphasize that the spine of hydration as observed crystallographically and as calculated as described herein, is a structural entity for which the corresponding energetics are not unequivocally established.

In conclusion, we note with considerable interest the extent of localization of the calculated hydration sites in the A·T and G·C regions found in the simulation and also the difference in penetration of the groove waters into DNA for A·T and C·G base pairs. This raises another potentially interesting thermodynamic issue, the extent to which or-

dered water structure in DNA is sequence dependent. A particular sequence that supports an unusually ordered water structure could be particularly favorable for binding substrate due to the extra entropy increase on desolvation, providing an additional entropic drive to the binding free energy (27). We believe that the present calculations and results indicate that further theoretical and experimental investigation of this idea would be worthwhile.

Very helpful discussions with Dr. Helen Berman of Fox Chase Cancer Research Institute and Prof. Kenneth J. Breslauer of Rutgers University are gratefully acknowledged. D.L.B. particularly acknowledges the hospitality and scientific discussions and interactions of the 1986 CECAM Workshop on Nucleic Acid Hydration (D. E. Westhof, organizer) at the University of Paris, Orsay and the “morte subite” group. This research was supported by grants from the National Institutes of Health (GM-37909), National Science Foundation (CHE-8696117), the Office of Naval Research, as well as the generosity of Merck Sharpe and Dohme Research Laboratories and the National Bureau of Standards. Computer facilities were provided by the Pittsburgh Supercomputer Center and Wesleyan University.

1. Saenger, W. (1984) *Principles of Nucleic Acid Structure* (Springer, New York), pp. 368–384.
2. Marky, L. A., Snyder, J. G., Remata, D. P. & Breslauer, K. J. (1983) *J. Biomol. Struct. Dyn.* **1**, 487–507.
3. Berman, H. M. (1986) *Ann. N.Y. Acad. Sci.* **482**, 166–178.
4. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **287**, 755–758.
5. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. (1983) *J. Mol. Biol.* **163**, 129–146.
6. Drew, H. R., Samson, D. & Dickerson, R. E. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4040–4044.
7. Drew, H. R. & Dickerson, R. E. (1981) *J. Mol. Biol.* **151**, 535–556.
8. Westhof, E. (1985) *Biochimie* **67**, 811–817.
9. Dickerson, R. E., Drew, H. R. & Conner, B. (1981) in *Biomolecular Stereodynamics I*, ed. Sarma, R. H. (Adenine, New York), pp. 1–34.
10. Minyat, E. E., Minchenkova, L. E. & Ivanov, V. I. (1980) *Studia Biophysica* **79**, 37.
11. Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D. & Schyolkina, A. K. (1974) *J. Mol. Biol.* **87**, 817–833.
12. Marky, L. A., Blumenfeld, K. S. & Breslauer, K. J. (1983) *Nucleic Acids Res.* **11**, 2857–2870.
13. Marky, L. A., Curry, J. & Breslauer, K. J. (1984) in *Molecular Basis of Cancer*, ed. Rein, R. (Liss, New York), pp. 155–173.
14. Kopka, M. L., Pjura, P., Yoon, C., Goodsell, D. & Dickerson, R. E. (1985) in *Structure and Motion: Membranes, Nucleic Acids and Proteins*, eds. Clementi, E., Corongiu, G., Sarma, M. H. & Sarma, R. H. (Adenine, New York), pp. 461–483.
15. Marky, L. A. & Breslauer, K. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4359–4363.
16. Mezei, M. & Beveridge, D. L. (1986) *MMC: A Digital Computer Program for Liquid State Monte-Carlo Metropolis Calculations* (Hunter College, New York).
17. Arnott, S., Campbell-Smith, P. J. & Chandrasekaran, R. (1976) in *CRC Handbook of Biochemistry and Molecular Biology*, ed. Fasman, G. (CRC Press, Cleveland, OH), 3rd Ed., Vol. 2, pp. 411–432.
18. Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. L. (1983) *J. Chem. Phys.* **79**, 926–935.
19. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
20. Seibel, G. L., Singh, U. C. & Kollman, P. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6537–6540.
21. Pangali, C., Rao, M. & Berne, B. J. (1979) *Chem. Phys. Lett.* **55**, 413–417.
22. Owicki, J. C. & Scheraga, H. A. (1977) *Chem. Phys. Lett.* **47**, 600–602.
23. Mehrotra, P. K., Mezei, M. & Beveridge, D. L. (1983) *J. Chem. Phys.* **78**, 3156–3166.
24. Manion, F., Norman, E., Stodola, B. & Wood, B. (1986)

^{*}A referee cautions that positive binding entropies are also observed for intercalators that presumably do not seriously influence the minor groove bound water.

- DOCK Version 5C* (Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia).
25. Mehrotra, P. K. & Beveridge, D. L. (1980) *J. Am. Chem. Soc.* **102**, 4287-4294.
 26. Mezei, M. & Beveridge, D. L. (1986) *Methods Enzymol.* **127**, 21-47.
 27. Berman, H.M. (1986) *Trans. Am. Cryst. Assoc.* **22**, 107-113.
 28. Mezei, M. & Beveridge, D.L. (1981) *J. Chem. Phys.* **74**, 622-632.
 29. Beveridge, D. L., Maye, P. V., Jayaram, B., Ravishanker, G. & Mezei, M. (1984) *J. Biomol. Struct. Dyn.* **2**, 261-270.
 30. Lumry, R. & Rajender, S. (1970) *Biopolymers* **9**, 1125-1227.