LA-UR 93-3013

CONF-9306237--3

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE A NEW APPROACH TO CALCULATE THE HYDRATION OF DNA MOLECULES

AUTHOR(S): GERHARD HUMMER
DIKEOS MARIO SOUMPASIS

SUBMITTED TO: For publication in Proceedings of the Eighth Conversation in the Discipline Biomolecular Stereodynamics, SUNY at Albany, NY, June 22-26, 1993. Adenine Press, NY.

DISCLAIMER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FORM NO 836 R4
ST NO 2629 5/81

A New Approach to Calculate the Hydration of DNA Molecules

Gerhard Hummer[1] and Dikeos Mario Soumpasis[2]

[1]Theoretical Biology and Biophysics Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

[2]Department of Molecular Biology, Max-Planck-Institute for Biophysical Chemistry, P.O. Box 2841, D-3400 Göttingen, F.R.G.

Proceedings of the

Eighth Conversation in the Discipline Biomolecular Stereodynamics

(SUNY at Albany, June 22–26, 1993)

Abstract

A new method to calculate approximate water density distributions around DNA is presented. Formal and computational simplicity are emphasized in order to allow routine hydration studies. The method is based on the application of pair and triplet correlation functions of water-oxygen calculated by computer simulation. These correlation functions are combined with the configurational data of the electronegative atoms on DNA (oxygen and nitrogen) taken from crystal structures. For three B-DNA structures water density distributions are calculated and discussed. The observed characteristic features agree well with the prevalent picture from experiments. The minor groove shows a more structured hydration than the major groove. Also, the minor groove hydration of A·T basepair tracts differs from that found in G·C basepair regions. In A·T tracts single peaks of high water density appear (*spine of hydration*), whereas in G·C regions the minor groove is occupied by two side-by-side ribbons of water.

Introduction

Water plays a crucial role in biomolecular systems. Both structure and reactions of biomolecules such as proteins and nucleic acids are influenced by interactions with the aqueous environment. Therefore, an adequate description of biomolecular hydration is indispensable for a detailed understanding of many important biological processes. Here we will focus on the interaction of DNA with water. A considerable amount of information has been obtained by various experimental techniques (for reviews see refs. (1–5)). Particularly crystal X-rays analysis revealed that the hydration of DNA molecules shows distinct sequence and structure dependent features, such as the characteristic so-called *spine of hydration* first observed by Dickerson's group (6, 7) in the AATT tract of the minor groove of a B-DNA dodecamer.

On the theoretical modeling side, Monte Carlo and molecular dynamics computer simulations provide flexible though costly methods to study the hydration of DNA. Several groups have analysed energetic and structural aspects of DNA-water systems (8–12). However, it must be stressed that in the case of biomolecular hydration analysis the computer simulation

methodology is subject to serious limitations, e.g., the enormous amount of CPU time required owing to the large number of particles involved and the slow relaxation times present in the system. Moreover, the long-range and strong Coulomb interactions cause additional problems; and the low density of bulk water (only about 0.034 particles per $Å^3$) results in considerable statistical uncertainties in water density distributions, even if these are based on very extensive computer simulation studies. Therefore, the development of alternative methods that allow a fast and accurate hydration analysis is essential to study the many biomolecular structures derived experimentally (X-rays, NMR etc.) and through computer modeling.

Here we describe a method to calculate the structural hydration of DNA molecules at comparable accuracy to good quality simulations but several orders of magnitude faster (13). It is based on the concept of potentials of mean force (PMF), which allows an implicit description of the solvent (water, ions) in terms of particle correlation functions. The PMF formalism was already successfully applied to study diffuse ionic effects (14–19) on DNA.

Theoretical Methods

We will here give only a brief outline of the theoretical methodology. For a detailed presentation and discussion of the formalism see ref. (13). The peculiar physical properties of condensed water phases can primarily be ascribed to the ability of water molecules to form complex networks of hydrogen bonds of the type O···H–O. These are of quantum mechanical nature and they basically originate from the angular structure and the charge distribution on the water molecule with the oxygen carrying a net negative charge due to its lone pair electrons. Like oxygen, nitrogen also exhibits a marked electronegativity. On DNA, the two strongly electronegative atoms oxygen and nitrogen are expected to have a predominant influence on the arrangement of water molecules in the surrounding through their capability of forming hydrogen bonds with water molecules. These hydrogen bonds of types N···H–O, N–H···O, and O···H–O are all within quite narrow bands concerning energy, bond angle, as well as bond distance; and they do not differ significantly from those found in bulk water. In a first order attack on the problem at hand, oxygens and nitrogens on DNA are identified with water oxygens, and the water density caused by these is calculated. This is accomplished by expanding the conditional density $\rho_{\mathrm{O}}^{(n,1)}$ of finding a water

oxygen at a position $\mathbf{r}$, given n water oxygens at the positions $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of the electronegative atoms oxygen and nitrogen on DNA in terms of water oxygen pair and triplet correlations $g^{(2)}_{OO}$ and $g^{(3)}_{OOO}$,

$$\rho_O^{(n,1)}(\mathbf{r}|\mathbf{r}_1, \ldots, \mathbf{r}_n) \approx \rho \times \prod_{i=1}^{n} g^{(2)}_{OO}(\mathbf{r}, \mathbf{r}_i) \times \prod_{j=1}^{n-1} \prod_{k=j+1}^{n} \frac{g^{(3)}_{OOO}(\mathbf{r}, \mathbf{r}_j, \mathbf{r}_k)}{g^{(2)}_{OO}(\mathbf{r}, \mathbf{r}_j) g^{(2)}_{OO}(\mathbf{r}_j, \mathbf{r}_k) g^{(2)}_{OO}(\mathbf{r}_k, \mathbf{r})}, \quad [1]$$

where ρ is the bulk water density. This formula for the conditional density is based on the expansion of n-particle PMF's in terms of lower order PMF's (20). It is important to retain the three particle correction in equation 1 in the case of water. This guarantees that the preferentially tetrahedral arrangement of water molecules is taken into account properly, which stems from the highly anisotropic intermolecular interactions in water. The water pair and triplet correlations $g^{(2)}$ and $g^{(3)}$ were calculated previously (21) for the highly elaborate *ab initio* Niesar-Clementi-Corongiu (NCC) water model (22) for pair distances between 2.3 and 7.9 Å with 0.2 Å increment and are used here along with linear interpolation.

Results and Discussion

We will show results for the sequence and structure dependent hydration of three B-DNA crystal structures. The positional data were taken from the Brookhaven Protein Databank. The water density was calculated using equation 1 for a volume of $30 \times 30 \times 42$ Å^3 around the molecules on a cartesian grid with a grid width of 0.3 Å.

Figures 1–3 show stereo plots of the main features of the structural hydration obtained for the Dickerson B-dodecamer d(CGCGAATTCGCG) (6) re-refined by Westhof (23), the dodecamer d(CGCATATATGCG) (24), and the decamer d(CCAACGTTGG) (25). The positions marked by points correspond to positions of highest probability for water oxygens ($\geq$10 times the bulk value). In the case of the dodecamers, the narrow central A·T base-pair region of the minor groove is covered with a distinct *spine of hydration.* In the wider G·C region, the spine splits into pronounced double ribbons of high water density. The major grooves do not show as much localized water as the minor grooves. Also along the sugar-phosphate backbone the water molecules are not strongly localized, except for clusters of high probability points at the phosphates.

The dodecamer d(CGCATATATGCG) exhibits a quite continuous high water density region in the central part of the minor groove that is only interrupted at the two 5'-TpA steps. A disruption of the *spine of hydration* at 5'-TpA steps was recently observed by X-rays crystallography (26); and simple energy calculations showed that the 5'-TpA step is unfavorable for a hydration *spine* formation (27).

The decamer d(CCAACGTTGG) shows a minor groove hydration pattern at the central 5'-CpG step that differs from the side-by-side ribbons observed at the terminal G·C regions of the dodecamers. Though less pronounced than in the A·T tracts of the dodecamers, the high density regions of water are well structured. This is in agreement with the X-rays crystallography observations of a short *spinelike* hydration in the central part of the minor groove of this molecule (25). Interestingly, the minor groove of the A·T tract is covered with a double ribbon of high water density, unlike the hydration pattern found in A·T basepair regions of the dodecamers. This finding again agrees with the X-rays results (25).

Concluding Remarks

We have presented a method that allows an accurate and computationally extremely efficient calculation of the structure and sequence dependent water density distributions around DNA molecules. Calculations like those described above (1.4×10^6 grid points) require only approximately 25 CPU minutes on a workstation. The results we obtained for the hydration of three B-DNA crystal structures are in very good agreement with the experimentally established picture of B-DNA hydration. The formalism allows routine hydration investigations of the many nucleic acids structures found experimentally or modeled on the computer. Results for other DNA structures (A- and Z-DNA) will be presented elsewhere. Moreover, the generality of the approach does not limit its application to nucleic acids. The extension to protein hydration analysis is under way.

Acknowledgements

The authors wish to thank P. Procacci, G. Corongiu, E. Clementi, A. E. García, R. Klement, T. Jovin, and M. Neumann for their collaborations and discussions. The work has been funded by the Bundesministerium für

Forschung und Technologie, the Max-Planck-Society (both F.R.G), and the Department of Energy (U.S.A.).

References and Footnotes

1. J. Texter, *Prog. Biophys. Mol. Biol. 33*, 83-97 (1978).

2. W. Saenger, "*Principles of Nucleic Acid Structure*", Springer, New York, pp. 368-384 (1984).

3. W. Saenger, *Ann. Rev. Biophys. Biophys. Chem. 16*, 93-114 (1987).

4. E. Westhof and D. L. Beveridge, in "*Water Science Reviews*", Vol. 5, ed. F. Franks, Cambridge University Press, Cambridge, pp. 24-136 (1990).

5. H. Berman, *Current Opinion Struct. Biol. 1*, 423-427 (1991).

6. H. R. Drew and R. E. Dickerson, *J. Mol. Biol. 151*, 535-556 (1981).

7. M. L. Kopka, A. V. Fratini, H. R. Drew, and R. E. Dickerson, *J. Mol. Biol. 163*, 129-146 (1983).

8. E. Clementi and G. Corongiu, *Biopolymers 20*, 551-571 (1981).

9. P. S. Subramanian and D. L. Beveridge, *J. Biomol. Struct. Dyn. 6*, 1093-1122 (1989).

10. F. Eisenhaber, V. G. Turmanyan, F. Eisenmenger, and W. Gunia, *Biopolymers 28*, 741-761 (1989).

11. V. P. Chuprina, U. Heinemann, A. A. Nurislamov, P. Zielenkiewicz, R. E. Dickerson, and W. Saenger, *Proc. Natl. Acad. Sci. USA 88*, 593-597 (1991).

12. A. V. Teplukhin, V. I. Poltev, and V. P. Chuprina, *Biopolymers 32*, 1445-1453 (1992).

13. C. Hummer and D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA* (submitted).

14. D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA 81*, 5116-5120 (1984).

15. D. M. Soumpasis, J. Wiechen, and T. M. Jovin, *J. Biomol. Struct. Dyn. 4*, 535-552 (1987).

16. A. E. García and D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA 86*, 3160-3164 (1989).

17. R. Klement, D. M. Soumpasis, and T. M. Jovin, *Proc. Natl. Acad. Sci. USA 88*, 4631-4635 (1991).

18. D. M. Soumpasis, A. E. García, R. Klement, and T. M. Jovin, in "*Theoretical Biochemistry and Molecular Biophysics*", eds. D. L. Beveridge and R. Lavery, Adenine Press, New York, pp. 343-360 (1990).

19. D. M. Soumpasis, in "*Computation of Biomolecular Structures. Achievements, Problems and Perspectives*", eds. D. M. Soumpasis and T. M. Jovin, Springer, Berlin, pp. 223-239 (1993).

20. A. Münster, "*Statistical Thermodynamics*", Vol. 1, Springer, Berlin, p. 338 (1969).

21. D. M. Soumpasis, P. Procacci, and G. Corongiu, *IBM DSD Report*, Oct. 1991; and *Mol. Phys.* (to be published).

22. U. Niesar, G. Corongiu, E. Clementi, G. R. Kneller, and D. K. Bhattacharya, *J. Phys. Chem. 94*, 7949-7956 (1990).

23. E. Westhof, *J. Biomol. Struct. Dyn. 5*, 581-599 (1987).

24. C. Yoon, G. G. Privé, D. S. Goodsell, and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA 85*, 6332-6336 (1988).

25. G. G. Privé, K. Yanagi, and R. E. Dickerson, *J. Mol. Biol. 217*, 177-199 (1991).

26. J. R. Quintana, K. Grzeskowiak, K. Yanagi, and R. E. Dickerson, *J. Mol. Biol. 225*, 379-395 (1992).

27. V. P. Chuprina, *Nucl. Acids Res. 15*, 293-311 (1987).

Figure Legends

Figure 1: Hydration of the B-DNA crystal structure of the dodecamer d(CGCGAATTCGCG). The stereo plot shows grid points with high water density ($\geq$10 times the bulk water density).

Figure 2: Hydration of the B-DNA crystal structure of the dodecamer d(CGCATATATGCG). Details as in figure 1.

Figure 3: Hydration of the B-DNA crystal structure of the decamer d(CCAACGTTGG). Details as in figure 1.

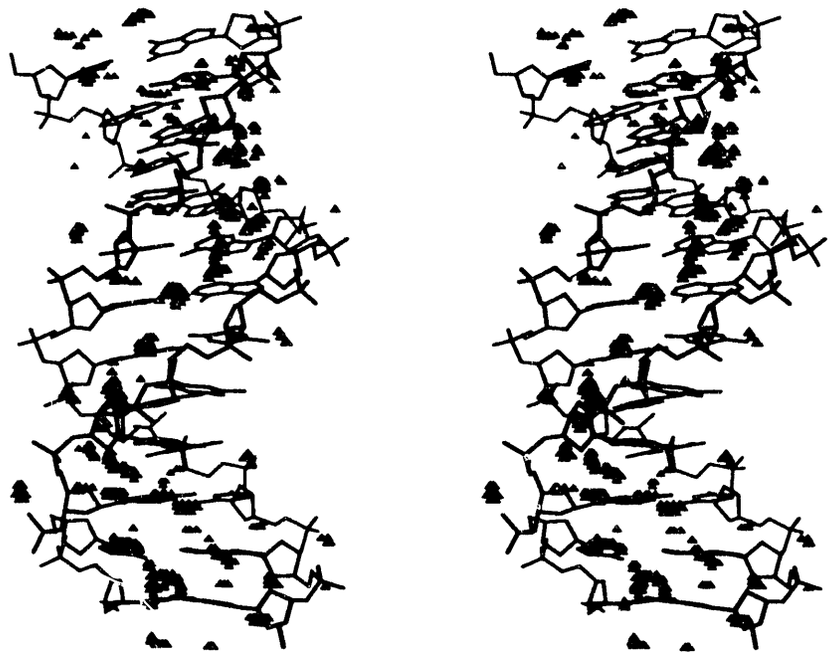

Figure 1

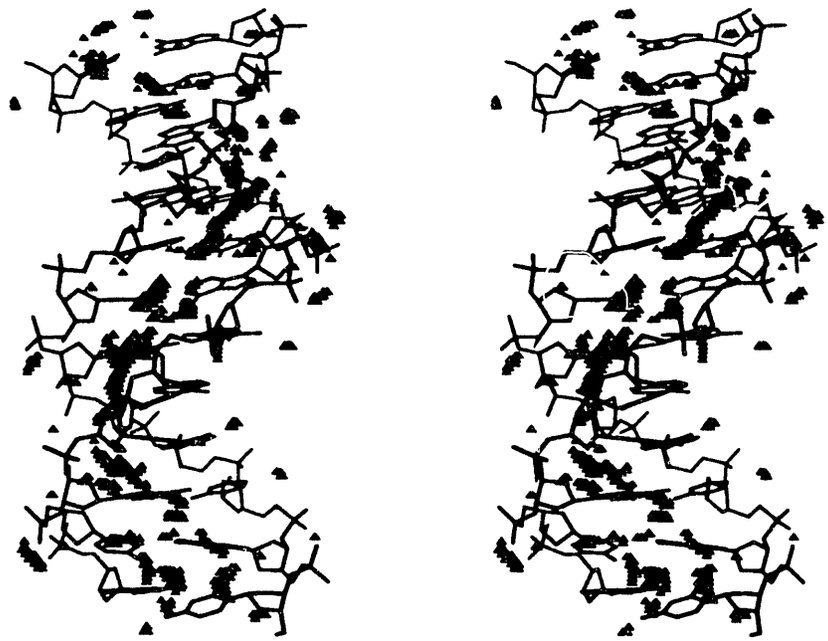

Figure 2

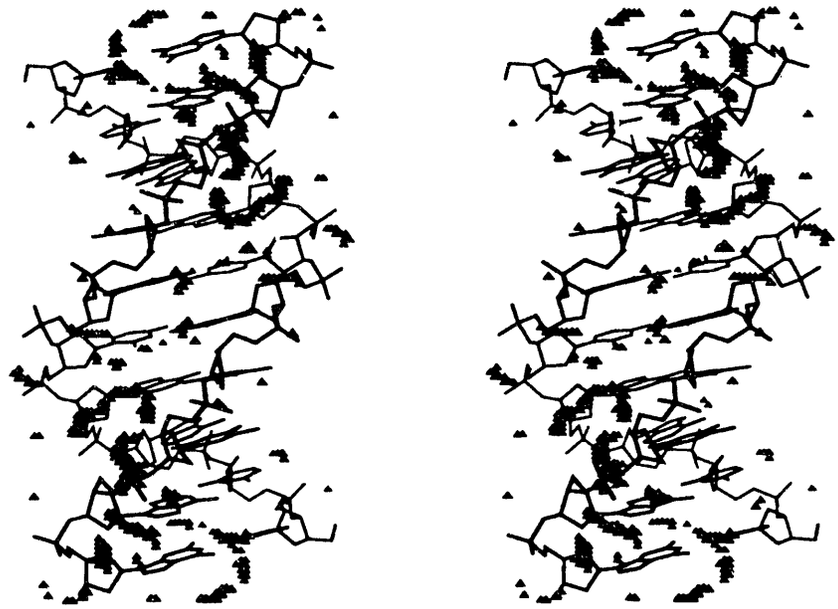

Figure 3

END

DATE FILMED

11/30/93