Leigh Ann Lipscomb
Fang Xiao Zhou
Loren Dean Williams*
School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, GA 30332

Clathrate Hydrates Are Poor Models of Biomolecule Hydration

Clathrate hydrates form the basis of a general model of biomolecule hydration. In clathrate hydrate crystal structures, the size of hydrogen-bonded water rings is highly constrained to five members. The clathrate hydrate model predicts that the size of water rings near biomolecule surfaces is similarly constrained to five members. This report describes a test of this model of biomolecule hydration. We have demonstrated that five-membered water rings are not a general feature of protein or nucleic acid hydration. The clathrate hydrate model appears to be inappropriate for soluble biomolecules. © *1996 John Wiley & Sons, Inc.*

INTRODUCTION

The hydrophobic effect is a dominant force in biomolecular conformation and stability.[1,2] Nonpolar residues cannot participate in hydrogen-bonding interactions with bulk water. Consequently the thermodynamics of transfer of a nonpolar molecule from liquid cyclohexane to water corresponds to −6 kcal/mol at room temperature.[3] This unfavorable entropy of hydration is thought to arise from formation of an ordered network of hydrogen-bonded water molecules around nonpolar solutes. The hydrogen-bonded network of water molecules surrounding nonpolar molecules in solution may be similar to that of clathrate hydrates.[4–7] In this report we compare the hydrogen-bonded network of water molecules in clathrate hydrates to that surrounding biomolecules.

Clathrate hydrates are crystalline substances in which nonpolar molecules are encapsulated by networks of hydrogen-bonded water molecules. The "ordering" of these water molecules refers to their localization and to regular repeats of certain structural motifs. One outstanding motif of clathrate hydrates is a five-membered ring of fully hydrogen-bonded water molecules. Crystallographically, one can observe such rings originating at one water molecule, passing via hydrogen bonds through four others before returning to the original water molecule. The sizes of water rings in clathrate hydrates are highly constrained, with five, or rarely, four or six water molecules.

Water molecules of clathrate hydrates are firmly bound with well-defined positions and orientations. Preferential formation of five-membered rings implies long-range through-bond and/or through-space interactions. These interactions would be attenuated by rotational and translational motions of the water molecules. Dunitz[8] has noted that the limiting entropy of transfer of a water from bulk to a well-defined position and orientation corresponds to around −2 kcal/mol at room temperature. Thus the limiting entropic cost of formation of a single five-membered ring corresponds to −10 kcal/mol.

The clathrate hydrates, with five-membered

Received December 20, 1994; accepted June 7, 1995.
* To whom correspondence should be addressed.
Biopolymers, Vol. 38, 177–181 (1996)

CCC 0006-3525/96/020177-05

rings of highly localized, fully hydrogen-bonded water molecules, have provided a model for hydration of hydrophobic regions of biomolecules. Support for the clathrate hydrate model of bimolecular hydration has been provided by observation of five-membered water rings in hydration shells of proteins and nucleic acids. Five-membered water rings have been observed in the hydration shells of crambin,[9] B-DNA d(CG)·proflavine,[10] A-DNA [d(GG^{Br}UA^{Br}UACC)]$_2$,[11] and insulin.[12]

The clathrate hydrate model predicts that five-membered hydrogen-bonded water rings are statistically overrepresented around hydrophobic regions of biomolecules, both nucleic acid and protein. We have developed assays for the validity of the clathrate hydrate model and have tested this prediction. Our results suggest that clathrate hydrates may not provide a useful model for biomolecule hydration.

The arrangement of water molecules of the *n*-propylamine clathrate hydrate is expected to closely resemble biomolecule hydration (G. A. Jeffrey, personal communication). Hydrogen bonds polarize water molecules. Rings maximize the induced polarities.

We have examined several proteins and nucleic acids representing each major secondary structural class, including A-, B-, and Z-DNA, and DNA–drug complexes. Crambin, d(GG^{Br}UA^{Br}UACC), and d(CpG)-proflavine were chosen because five-membered water rings have been described near their surfaces. We have used proteins and representatives of each DNA structural class that were determined to high resolution (see Table I).

Table I Refinement Statistics for Biomolecules Studied

Biomolecule	Resolution (Å)[a]	*R*-Factor (%)[b]
Erabutoxin[15]	1.4	14.4
Crambin[9]	0.95	12.9
Achromobacter protease I	1.2	14.9
d(CGCGCG)[17]	1.0	18.0
d(CGATATATCG)[16]	1.70	17.8
d(CG)·proflavine[10]	0.9	15.1
d(CGATCG)·adriamycin[13]	1.4	20.0
d(GG^{Br}UA^{Br}UACC)[11]	1.7	14.0

[a] The accuracy of a crystal structure depends strongly on the resolution of the diffraction pattern. The resolution is calculated as $\lambda/2 \sin \Theta$, where Θ is the angle at which the diffraction pattern fades, and λ is the radiation wavelength (Å). The structures of all biomolecules studied in this paper were determined to relatively high resolution.

[b] The crystallographic *R* factor provides a measure of agreement between calculated (F_c) and observed (F_o) structure factors. The *R* factor is calculated as $\Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$. For a macromolecular crystallographic experiment, *R* factors are typically about 20%. The structures of all biomolecules studied in this paper were refined to relatively low *R* factors.

MATERIALS AND METHODS

The initial step in our test of the clathrate hydrate model was to define the normal distribution of water ring sizes in the absence of preferential formation of any particular ring size. A definition of normal distribution allowed quantitative assessment of the degree of overrepresentation of any given ring size in observed hydration shells. The definition of normal distribution was given by a *statistical model*. The statistical model contrasts with the clathrate hydrate model by predicting relative populations of water rings solely by the number of possible combinations that can produce each ring. The statistical model predicts normal distributions in which no ring size is energetically preferred. In the statistical model the relative population [$M(n, w)$] of rings with n members is given by

$$M(n, w) = (w)!/2(w - n)!(n)$$

M is a function of the number of ring members (n), and the total number of hydrating water molecules (w). Predictions of the statistical and clathrate hydrate models were compared to each other and to observed distributions via the program CYCLONE.[13] CYCLONE uses three-dimensional coordinates to determine populations of rings in the hydration shell of a macromolecule.

With the exception of crambin and *n*-propylamine clathrate hydrate, coordinates of water molecules were obtained from the Protein Data Bank (PDB).[14] The coordinates of the water molecules surrounding crambin were provided by Dr. Martha Teeter. Water molecule coordinates of *n*-propylamine clathrate hydrate were obtained from the original citation.[7]

CYCLONE sets up a matrix of all interatomic contacts less than 3.4 Å and sorts the matrix to count rings. The hydrogen-bond cutoff distance, 3.4 Å, was determined as described previously.[13] Distributions of rings are not sensitive to changes (3.0–3.6 Å) in the cutoff distance (data not shown). CYCLONE operates within a defined three-dimensional box. Lattices were generated with appropriate symmetry to ensure that rings involving water molecules of different asymmetric units were properly counted. There is a critical box volume below which an accurate distribution is not obtained. This edge effect arises when the ratio of volume to surface area is too low. With a sufficiently large volume, increasing the volume increases the numbers of rings observed, but does not alter the distribution.

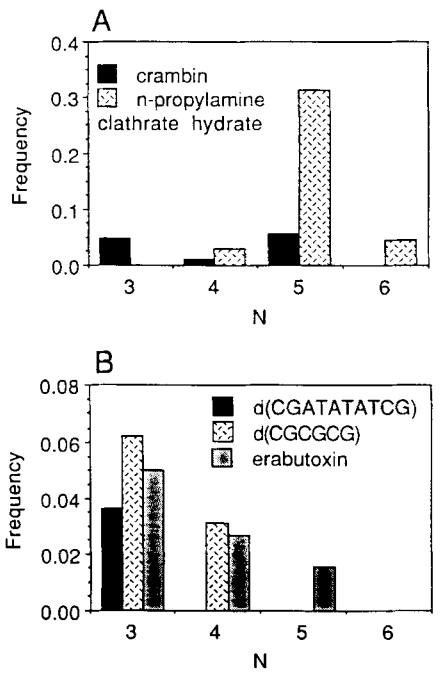

FIGURE 1 Observed distribution of rings in the crystal structures of (A) crambin and *n*-propylamine clathrate hydrate, and (B) erabutoxin, [d(CGATATATCG)]$_2$ and [d(CGCGCG)]$_2$. In each case the frequency has been normalized to the number of water molecules. N is ring size.

RESULTS

We have determined the distribution of water ring sizes in a series of previously reported crystal structures. The series includes *n*-propylamine clathrate hydrate (structure described in Ref. 7), crambin,[9] erabutoxin,[15] achromobacter protease I (PDB entry 1 ARB), B-DNA [d(CGATATATCG)]$_2$,[16] and Z-DNA [d(CGCGCG)]$_2$.[17] The x-ray intensity data of each structure is of sufficiently high quality so that water molecules can be confidently located. These biomolecules represent both proteins and nucleic acids of widely varying sequences and in different crystallographic packing environments. Observed relationships between ring size and frequency are shown for crambin and *n*-propylamine clathrate hydrate in Figure 1A, and for erabutoxin, B-DNA [d(CGATATATCG)]$_2$, and Z-DNA [d(CGCGCG)]$_2$ in Figure 1B. The results described here suggest that the clathrate hydrate model is not of general utility for describing biomolecule hydration.

We have also investigated the correlation between the distribution of ring sizes and the extent of hydration for DNA and DNA–drug complexes (see Figure 2). Extent of hydration is given by the number of crystallographically ordered water molecules normalized to the size of the DNA fragment, including other ligands (i.e., the number of water molecules per base pair or drug molecule). The distribution of ring sizes is given by the slope of the log $Q(n, w)$ vs log $M(n, w)$ plot, where $Q(n, w)$ is the number of rings found by CYCLONE and $M(n, w)$ is calculated as described above. The slope of the log $Q(n, w)$ vs log $M(n, w)$ plot increases with the ratio of large rings to small rings. The results in Figure 2 suggest that the distribution of ring sizes is related to the extent of hydration.

DISCUSSION

Proteins

The hydration shell of crambin has similarities but also significant differences from predictions of the clathrate hydrate model. Crambin is of particular interest because five-membered water rings were described near portions of this molecule.[9] Our results confirm that five-membered rings are statistically overrepresented in the hydration shell of crambin. The observed relationship between ring size and population in the hydration shell of crambin is illustrated in Figure 1A. There are many five- and three-membered rings and many fewer four- or six-membered rings. Thus five-membered water rings are abundant in crambin and the *n*-propyl-

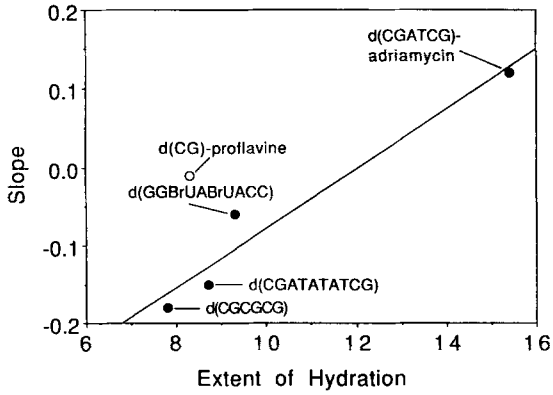

FIGURE 2 Slopes of log $M(n, w)$/log $Q(n, w)$ plot vs extent of hydration. $M(n, w)$ is the number of rings with n members predicted by the statistical mode. $Q(n, w)$ is the number of rings with n members as determined by the program CYCLONE. The extent of hydration is given by the number of crystallographically ordered water molecules per base pair normalized to the size of the DNA fragment, including ligand.

amine clathrate hydrate. Three-membered water rings are abundant near the surface of crambin, but absent for *n*-propylamine clathrate hydrate.

Our results show that overrepresentation of five-membered rings is not a general characteristic of protein hydration. The distribution of water ring sizes near crambin appears to be anomalous. A search for protein hydration shells in which five-membered rings are statistically overrepresented yielded only crambin. For example, the relationship of ring size and frequency in the hydration shell of the protein erabutoxin is illustrated in Figure 1B. In this case five-membered rings occur at low frequency. We have also determined that only three-membered water rings are present near the surface of Achromobacter protease I (data not shown; structure obtained from the PDB, entry number 1 ARB). In summary, three of four protein structures studied with CYCLONE do not display a preference for five-membered water rings. Similarly, five-membered rings have been reported to occur at low frequency in the hydration shell of lysozyme.[18]

Three-membered rings are the most abundant in the hydration shells of erabutoxin, achromobacter protease I, and lysozyme.[18] In addition three-membered rings occur with high frequency in the hydration shell of crambin. This result is surprising because small rings are thought to be intrinsically unstable due to ring strain and thus are expected at low frequency.[19]

DNA

We previously provided evidence that the clathrate hydrate model may not be generally applicable to nucleic acid hydration.[13] A quantitative determination of water ring size distribution did not reveal overrepresentation of five-membered rings even in structures where five-membered rings had been described. Previous observation of five-membered rings does not indicate statistical overrepresentation. Specifically, we reported that five-membered hydrogen-bonded water rings are not statistically overrepresented near the A conformation of $[d(GG^{Br}UA^{Br}UACC)]_2$, or B-DNA–drug complexes d(CGATCG)·adriamycin and d(CG)·proflavine. Although these results are inconsistent with the general applicability of the clathrate hydrate model of DNA hydration, the possibility remains that the clathrate model correctly describes hydration near B-DNA or specific DNA residues or sequences. It has been proposed that five-membered water rings occur at high frequency near 5-methyl and 5-bromo groups of thymine and uracil[11] and near the sequence d(TATA) in B-DNA.[20]

We have determined the distribution of water ring sizes in the hydration shell of B-form $[d(CGATATATCG)]_2$ (structure reported in Ref. 16). The observed relationship between ring size and population is illustrated in Figure 1B. Five-membered rings are not overrepresented. Three-membered rings are the most abundant.

In hydration shells of nucleic acids, the distribution of water ring sizes is determined largely by the extent of hydration. Extent of hydration is given by the number of crystallographically ordered water molecules normalized to the size of the DNA fragment, including other ligands. The distribution of ring sizes is given by the slope of the log $Q(n, w)$ vs log $M(n, w)$ plot. This slope increases with the ratio of large rings to small rings. As shown in Figure 2 the distribution of ring sizes correlates remarkably with the extent of hydration. When five DNA fragments and DNA·drug complexes are included, the relationship between extent of hydration and distribution of ring sizes is nearly linear. The correlation coefficient is 0.73. The d(CG)·proflavine complex is an outlier as might be expected from the small size of the DNA fragment and the high ratio of drug to DNA (two molecules of proflavine per dinucleotide duplex). In the absence of d(CG)·proflavine, the correlation coefficient for the relationship plotted in Figure 2 is 0.95.

The model of biomolecule hydration emerging from our studies is one in which the distribution of water ring sizes is random. In dehydrated systems such as Z-form $[d(CGCGCG)]_2$, the larger ring sizes (i.e., $n = 4$–7) are underrepresented. The water density near such biomolecules is sufficiently low that formation of larger ring sizes occurs infrequently. The water molecules are widely separated from one another and interact predominantly with the biomolecule. There is not a continuous network of hydrogen-bonded water molecules. This model can explain the high frequency with which three-membered water rings are observed near the surfaces of proteins and nucleic acids in crystals.

Conclusion

The clathrate hydrate model of hydration implies many water molecules are firmly bound to a macromolecular surface and/or to other water molecules. Such firmly bound water molecules are energetically expensive and may in fact signal insolubility in water. This proposal is consistent with the cluster of 5-membered rings near crambin, a water-

insoluble protein.[21] We have been unable to discern overrepresentation of five-membered rings in hydration shells of soluble macromolecules. Our statistical model, with randomly associating water molecules, is consistent with loosely bound, disordered water molecules in the hydration shells of soluble macromolecules.

This work was supported by American Cancer Society Grant NP-831. The coordinates of the water molecules surrounding crambin were provided by Dr. Martha Teeter, Boston College.

REFERENCES

1. Kauzmann, W. (1959) *Adv. Protein Chem.* **14,** 1–63.
2. Tanford, C. (1980) *The Hydrophobic Effect* John Wiley & Sons, New York.
3. Creighton, T. E. (1993) *Proteins,* W. H. Freeman, New York.
4. Klotz, I. M. (1958) *Science* **128,** 815–822.
5. Jeffrey, G. A. & McMullan, R. K. (1967) *Progr. Inorg. Chem.* **8,** 43–108.
6. Jeffrey, G. A. (1969) *Acc. Chem. Res.* **2,** 344–352.
7. Brickenkamp, C. S. & Panke, D. (1973) *J. Chem. Phys.* **53,** 5284–5295.
8. Dunitz, J. (1994) *Science* **264,** 860.
9. Teeter, M. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6014–6018.
10. Schneider, B., Ginell, S. L. & Berman, H. M. (1992) *Biophys. J.* **63,** 1572–1578.
11. Kennard, O., Cruse, W. B. T., Nachman, J., Prange, T., Shakked, Z. & Rabinovich, D. (1986) *J. Biomol. Struct. Dynam.* **3,** 623–647.
12. Baker, T., Dodson, E., Dodson, G., Hodgkin, D. & Hubbard, R. (1985) in *Crystallography in Molecular Biology,* Moras, D., Drenth, J., Strandberg, B., Suc, D. & Wilson, K., Eds., Plenum Press, New York, pp. 179–192.
13. Lipscomb, L. A., Peek, M. E., Zhou, F. X., Bertrand, J. A., VanDerveer, D. & Williams, L. D. (1994) *Biochemistry* **33,** 3649–3659.
14. Bernstein, F. C., Koetzle, T. F., Williams, J. G. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
15. Smith, J. L., Corfield, P. W. R., Hendrickson, W. A. & Low, B. W. (1988) *Acta Cryst.* **A44,** 357–368.
16. Yuan, H. S., Quintana, J. R. & Dickerson, R. E. (1992) *Biochemistry* **31,** 8009–8021.
17. Bancroft, D., Williams, L. D., Rich, A. & Egli, M. (1994) *Biochemistry* **33,** 1073–1086.
18. Madhusudan, Kodandapani, R. & Vijayan, M. (1993) *Acta Cryst.* **D49,** 234–245.
19. Del Bene, J. & Pople, J. A. (1970) *J. Chem. Phys.* **52,** 4858–4866.
20. Saenger, W. (1987) *Ann. Rev. Biophys. Chem.* **16,** 93–114.
21. Teeter, M. M. & Hendrickson, W. A. (1979) *J. Mol. Biol.* **127,** 219–223.