Journal of Biomolecular Structure and Dynamics, ISSN 0739-1102
Volume 1 (1983), Ądenine Press (1983).

Monte Carlo Studies on Water in the dCpG/Proflavin Crystal Hydrate

Mihaly Mezei, David L. Beveridge
Chemistry Department
Hunter College of the City University of New York
695 Park Avenue
New York, N.Y. 10021

Helen M. Berman
Institute for Cancer Research,
Fox Chase Cancer Center,
Philadelphia, PA 19111

Julia M. Goodfellow, John L. Finney
Department of Crystallography,
Birkbeck College,
University of London,
London, WC1E 7HX, UK.

and

Stephen Neidle
Department of Biophysics,
King's College,
University of London,
London WC2B 5RL, UK.

Abstract

The extensive water network identified in the crystallographic studies of the dCpG/Proflavin hydrate by Neidle, Berman and Shieh (*Nature* 288, 129, 1980) forms an ideal test case for a) assessing the accuracy of theoretical calculations on nucleic acidwater systems based on statistical thermodynamic computer simulation, and b) the possible use of computer simulation in predicting the water positions in crystal hydrates for use in the further refinement and interpretation of diffraction data. Monte Carlo studies have been carried out on water molecules in the unit cell of dCpG/ proflavin, with the nucleic acid complex fixed and the condensed phase environment of the system treated by means of periodic boundary, conditions. Intermolecular interactions are described by potential functions representative of quantum mechanical calculations developed by Clementi and coworkers, and widely used in recent studies of the aqueous hydration of various forms of DNA fragments. The results are analyzed in terms of hydrogen bond topology, hydrogen bond distances and energies, mean water positions, and water crystal probability density maps. Detailed comparison of calculated and experimentally observed results are given, and the sensitivity of results to choice of potential is determined by comparison with simulation results based on a set of empirical potentials.

Introduction

The influence of environmental effects on the stability and structure of DNA has been recognized for some time, and has stimulated a number of experimental and theoretical studies on the interaction of nucleic acids and water. Considerable information on nucleic acid/water systems has been gathered from thermodynamic and spectroscopic studies, but a complete description of the aqueous hydration of nucleic acids at the molecular level by physical methods is not feasible, nor foreseen to be. Theoretical methods, wherein a complete hydration model is obtained from a computer simulation thus have a unique vantage point on the problems in this area, providing of course that reliable results can be obtained and validated. Applications of computer simulation to the aqueous hydration of biological molecules has been described by e.g. Rossky and Karplus[1] using molecular dynamics and by Clementi and coworkers[2–5] using the Monte Carlo method but only limited comparison of theoretical calculations with experimental results has been possible in these studies.

Recently, X-ray diffraction studies of certain oligonucleotide crystal hydrates have been carried out to a point of resolution such that the positions of some if not all of the water molecules can be determined with high accuracy.[6–8] The oligonucleotides reported are prototypes for the nucleic acid duplex structure in the A,B and Z forms, and for DNA drug intercalation complexes. One of the latter, the dCpG/ proflavin crystal hydrate reported by Neidle, Berman and Shieh,[8] features one of the most elaborate water networks observed to date; c.f. Figure 1. The observed network involves 100 water molecules per unit cell, in highly structured closed polygonal forms. A distinct "polygon disk" and an extensive "pentagon network" of water molecules was identified. One proflavin molecule is the counterion for each nucleic acid phosphate group, so there is no complicating ion atmosphere problem to deal with. The dCpG/proflavin system is thus an ideal prototype for theoretical study and detailed comparison between calculated and observed results. The objective of this study is to ascertain the extent and quality of correspondence between results obtained from computer simulation and experimental data on the dCpG/ proflavin crystal hydrate system.

Background

Computer simulation procedures, both the method of molecular dynamics and the Monte Carlo method, have come into wide use for the study of liquid water,[9] and other molecular liquids[10,11] as well as for aqueous solutions.[13] The application of computer simulation to the study of the organization of water in crystal hydrates is in a relatively early stage of development. The Monte Carlo method has been used by Hagler and coworkers on studies of water in lysozyme crystals[14] and in more detailed studies on crystals of a cyclic hexapeptide hydrate[15] and antamanide dodecahydrate.[16] Herman and Vacatello have recently described simulation studies of a BPTI crystal.[17] Application of molecular dynamics to the study of water in bimolecular crystals have been reported by Berendsen and coworkers.18 Goodfellow, Finney and Barnes have reported Monte Carlo studies on crystal hydrates of amino acids[19] and have work in progress on hydration of biopolymers. The results of these studies are collectively promising, although sometimes only low resolution comparison with experiment is possible, and other elements (salts, etc.) in the environment are neglected. In some cases, the waters are constrained to sequestered sites or a channel, and thus do not fully test all degrees of freedomof the simulation methodology.

Figure 1. The water network of the dCpG/proflavin crystal hydrate as described in Ref. 8.

Previous computer simulation results on water and aqueous solutions lead us to anticipate that the crystal hydrate calculations will generally be difficult.

The statistical averaging features in computer simulation result in some properties being more accessible to reliable calculation than others, e.g., simple average quantities like internal energy converge faster than fluctuation properties such as heat capacity. Further details on convergence characteristics of computer simulation on liquid water and aqueous solutions are described in Refs. 20 and 21. The statistical mean positions of waters of hydration in a crystal are expected to be even slower to converge than any previously computed structural or energetic property, since no averaging beyond simple configurational is involved. The quantitative agreement between calculated and observed positions is also expected to be quite sensitive to the fine details of the various assumed intermolecular potential functions,[19] which are largely untested. However, if the calculation of crystal water positions in this system can be successfully accomplished, this serves to validate current simulation studies of the aqueous hydration of biological molecules and encourages further quantitative extensions. Moreover, it indicates that Monte Carlo computer simulation may be a useful additional tool for the crystallographer to use in data refinement. The ultimate hope is that computer simulations may provide a reliable point of departure for general analysis of diffraction data on water in crystal hydrates, and also serve as means for the theoretical analysis of the crystal water in terms of structural and energetic indices.

We report here preliminary results of theoretical studies of the organization of water in the dCpG/proflavin crystal hydrate, based on Monte Carlo computer simulation. We attempt to deal independently with convergence characteristics and intermolecular potential functions in the analysis of results and present further perspectives on the applicability of computer simulation to the study of this class of problems.

Calculations

Statistical thermodynamic Monte Carlo computer simulations were carried out on the waters of hydration in the dCpG/proflavin crystal hydrate at T=298^{o}K. The simulations were performed on the crystallographic unit cell of the system with lattice constants a=32.991 Å, b=21.995 Å and c=13.509 Å and $\alpha = \beta = \gamma = 90^o$. The crystal has four asymmetric units per cell and space-group symmetry P$2_1 2_1$2. There are two dCpG strands and two proflavin molecules per asymmetric unit. In the calculation, the atoms of the dCpG/proflavin complex were fixed in their crystallographically determined positions with hydrogens atoms filled in using standard bond lengths and angles. All waters of hydration were treated in the simulation as a molecular fluid. Since all waters were explicitly considered in the simulation, the space-group symmetry can serve as an indicator of convergence on the result.[22]

Coordinates of the 26 waters per asymmetric unit were reported in the crystallographic investigation, 24 of which were assigned unit occupancy and two with fractional occupancy of one half. Thus the number of ordered water molecules per unit cell was indicated to be 100, and simulations were first carried out on this basis. Calculations of the volume accessible to the water in the crystal and estimate of the total volume occupied by solvent from partial molar volumes for nucleic acids indicated that there is additional space in the crystal, presumably occupied by disordered water molecules. This prediction is consistent with new density measurements by Neidle, which indicate that there may be a total of 27 water molecules per asymmetric unit. Subsequently, simulations based on 108 water molecules per unit cell were carried out.

Two sets of calculations have been carried out on this system, one set at Hunter College (HC) and one set at Birkbeck College (BC). In the HC set, the intermolecular interactions were described by potentials developed by E. Clementi and coworkers.[23–32] The water-water interactions were modeled by the MCY potential,[30] studied in detail in several Laboratories.[9] The dinucleotide-water and proflavinwater interactions were modeled using the potential library developed for amino acid-water, nucleic acid-water and sugar-water interactions. Certain typographical errors in the original specifications were corrected with the assistance of G. Corongiu and the description of the system is thought to be identical to that used by Clementi and Corongiu in their simulation studies on the hydration of DNA.[2–5]

The MCY potential, used to describe water-water interaction is based on moderate size CI calculations. In computer simulations on liquid water the MCY potential accounts quite well for the O-O radial distribution function and thermodynamic indices such as internal energy and heat capacity. The calculated pressure with this function is well known to be too high. The solute-water potentials are representative of small basis set LCAO-SCF MO calculations of solute-water interaction energies and are the most comprehensive and interally consistent set of functions presently available for the study of hydration of biological molecules. Quantum mechanical calculations at this level are, however, susceptible to basis set super-position errors, which could lead to solute-solvent binding curves which are too deep and too narrow and equilibrium separations shorter than required. Compensating errors from Coulomb and charge transfer contributions could mitigate the difficulties,[33] so the ultimate assessment of the performance of the potentials must come from a comparison of calculated and observed quantities.

All HC simulations were carried out using the Metropolis Monte Carlo method,[34] supplemented with the force-bias sampling technique[35] for convergence acceleration. As discussed in a previous publication,[21] the use of the force-bias technique requires the recomputation of the pair energies and forces at each accepted step, since the storage of all force and torque components is impractical. In our case, however, we found that significant gain can be achieved if the crystal-water energies, forces and torques are stored since their computation is significantly more expensive than the computation of the analogous quantities for water-water interactions. The dCpG/ proflavin interactions were treated under the minimum image convention. The water-water interactions were computed using a spherical cutoff of 6.75 Å rather than the value of 7.75 Å used in previously reported simulations since the smallest edge of the unit cell is 13.5 Å.

Three simulation studies of the dCpG/proflavin crystal hydrate have been carried out to date at HC. Two simulations, HC-I and HC-II assumed 100 water molecules and differ only in the way the initial configuration in the simulation was chosen. In simulation HC-I, we chose an initial configuration based on the crystallographically observed water oxygen positions, and placed waters on each site with random orientations. Equilibration was begun with water positions fixed for the first 200K configurations to obtain reasonable orientations, which brought the energy of the system to the range of -1400 kcal/mol (of cells). A subsequent 500K equilibration, with positional and orientational coordinates sampled, carried the energy to around -1870 kcal/mol. From this point, 4000K more configurations were sampled and ensemble averages were formed over the last 2400K. This run length can be compared to molecular dynamics calculations using the findings of Pangali et al[36] that for the ST2 water, one force-biased Monte Carlo step is equivalent to 10^{-15}/N sec (N is the number of waters). Thus our 2400K long run is effectively equivalent to 24 psec of dynamical sampling. Simulation HC-II, of comparable length, was carried out from a starting configuration obtained by arranging the requisite number of waters arbitrarily in the crystallographic unit cell. Simulation HC-III involved 108 water molecules with the experimentally obtained oxygen positions used as in HC-I, with eight more waters added to the cell in such a way that serious molecular overlaps were avoided. Ensemble averages in simulation HC-III were formed over 2000K configurations. The thermodynamic indices of the system and generic mean positions36 of water molecules were computed.

The BC simulations were carried out on the unit cell with 108 water molecules using the standard Metropolis procedure and the potential functions as described in Ref. 19. At the time of this writing, results for a preliminary 1000K simulation were available, 800K equilibration and 200K production. A fuller description of this calculation will be published after the realization is extended.

Results and Discussion

The convergence characteristics of the mean energy for simulation HC-I is shown in Figure 2. The energy is seen to be stable within 1-2% for the final 2600K configurations. Comparison of corresponding plots for liquid water[20] and aqueous solutions[21] shows that convergence in the crystal hydrate case is considerably slower, and as a consequence the length of the realization was extended to approximately twice that found to be satisfactory for liquids. The other simulations in this study, HC-II and HC-III, produced convergence profiles of an essentially similar appearance to that of HC-I, and are not reproduced here.

Figure 2. Convergence profile (control function) of the mean energy for simulation HC-I.

The calculations were carried out on the crystallographic unit cell, and the extent to which the simulation results develop space group symmetry in the asymmetric units can also be used as a convergence test in the calculations. The calculated mean water positions in each asymmetric unit were compared, and in simulation HC-I 60% of the calculated sites passed an asymmetric unit test requiring mutual symmetry correspondence of four sites yielding an average tolerance of 0.6 Å. Substantially similar results were obtained for HC-II and HC-III, 48% and 63%, respectively. Symmetry convergence of mean positions is thus significant but far from complete in spite of the reasonable convergence realized in the computed mean energy.

The comparison of calculated and observed mean positions must be carried out in context of the results on convergence described above. To facilitate this an initial comparison is made in terms of the scattergram plotted in Figure 3. Here the abscissa is a measure of deviation from convergence on a position by position basis, realized by plotting the deviation in corresponding mean positions in simulations HC-I and HC-ll. Perfectly converged results would show total agreement in mean positions from HC-I and HC-II, since the results should be independent of initial guess. Deviations from perfect agreement are thus a measure of deviation from convergence. Deviation of the calculated mean position from experiment are measured along the ordinate. The resulting scattergram, Figure 3, has a cluster of points around the origin, shows clearly a strong correlation between convergence of calculated mean positions and agreement between calculated and observed values, and gives a general idea of the range of correspondance achieved between theory and experiment up to this point. The assignment of theoretically obtained sites to the experimentally determined sites was solved to a good approximation here by integer programming.

Figure 3. Plot of deviations between mean positions obtained in simulation HC-I and the experimentally obtained sites vs deviations of the mean positions between simulation HC-I and simulation HC-II.

A more graphic view of the correspondence between calculated and observed water positions in the dCpG/proflavin crystal hydrate is shown in Figure 4 for simulation HC-III (108 waters). Here the observed water positions of an asymmetric unit of the crystal are displayed as small circles relative to the four dinucleoside phosphate strands. In the analysis presented here, each of the four asymmetric units in the crystallographic unit cell was examined separately for correspondence between calculated and observed mean positions. Agreement within a tolerance of 0.6 Å is signalled by encircling the observed mean position.

Full agreement, correspondence between calculated and observed values, is achieved for five positions. Partial agreement is indicated for another 11, with then some level of agreement achieved for 64% of the sites. At a more detailed level, one can see from Figure 4 that best agreement between calculated and observed positions is found in the region of the polygon disk, whereas the pentagon network is not well established in the simulation. The proflavin hydration is fairly well described, but agreement on hydration of the nucleotide bases, and guanine in particular, is lacking. Analysis of the solute water interatomic separations, the distance between solute heavy atoms and water oxygens, shows an overabundance of short solute-water contacts. There are 160 separations less than 3.3 Å in the calculated results, compared with 92 observed. The sensitivity of the results to the choice of the potential has been further elucidated by the BC study currently being carried out using an alternative set of functions for intermolecular interaction energy.[19]

Figure 4. Site by site comparison of the experimentally obtained sites with the sites obtained in simulation HC-HI. The number of circles give the number of asymmetric units where agreement within 0.6 Å was found.

Analysis of the initial production segment of the BC simulation indicates the number of solute-water separations less than 3.3 Å is 72. The distance between each solvent molecule and all neighbouring atoms within a 3.75 Å sphere is predicted within an average tolerance of 0.6 Å. However at this stage of the simulation, the same nearest neighbour distances differed by an average of 0.45 Å between the four asymmetric units in the simulated unit cell. The water structure in the pentagon network appears to be reproduced. The convergence behaviour of the BC simulation is similar to that described above: 24% of the sites passed the asymmetric unit test requiring mutual symmetry correspondence of four sites with a tolerance of 0.6 Å. This percentage increased to 72% with a tolerance of 1.0 Å. When each asymmetric unit is examined separately, 36% of the sites showed partial agreement with only 32% of the sites showing no correspondence within the 0.6 Å tolerance.

Summary and Conclusion

The extent of agreement between observed mean positions in the dCpG/proflavin crystal hydrated and Monte Carlo calculations has been presented, along with a characterization of the level of convergence achieved in the calculations. Comparison of calculated and observed mean water positions indicates that well converged sites tend to correspond more closely to experiment than those with convergence problems. The polygon disk and proflavin hydration in the water network are better described in the HC calculation than the pentagon network or nucleotide base hydration while calculated solute-solvent intermolecular separations turned out to be shorter overall. This may indicate either that the intermolecular potential functions assumed for the calculation may be overestimating the solute-solvent attraction or that the high pressure of the MCY water caused an extension of the simulated water network at the expense of the crystal-water distances.

Crystal water simulations can be expected to present more difficult convergence problems than water and aqueous solutions and it is likely that methodological improvements are necessary to achieve results commensurable in quality to those obtained for dilute aqueous solutions. Follow-up studies on the performance of the theoretical potential function vis-a-vis experiment is also indicated to determine if further use of these functions is advisable. Further details from the BC simulation will be available shortly, and will permit a detailed consideration of the sensitivity of results to the choice of potential. At this point we feel that simulation procedures have produced a significant but only partial account of the water network in the dCpG/proflavin crystal hydrate, and the predictive value of simulation for use in analysis and interpretation of crystal hydrates remains to be established.

Acknowledgements

This research was supported by NIH grants GM-12149 (DLB) and GM-21589 (HMB), a NATO Collaborative Research Grant, No. 010.82. and a CUNY Faculty Reserach Award. HMB acknowledges as well NIH grants CA-06927 and RR-05539.

References and Footnotes

1. P.J. Rossky and M. Karplus, *J. Am. Chem. Soc., 101,* 1913 (1979).

2. G. Coronqiu and E. Clementi, *Biopolymers, 20,* 551 (1981).

3. G. Coronqiu and E. Clementi, *Biopolymers, 20,* 2427 (1981).

4. G. Coronqiu and E. Clementi, *Biopolymers,* in press.

5. E. Clementi and G. Coronqiu, in *Biomolecular Stereodynamics,* R.H. Sarnia, ed., Adenine Press, New York 1983.

6. R.E. Dickerson, H.R. Drew, B.N. Connor, R.M. Wing, A. Fratini, and M.L. Kopka. *Science, 216,* 475 (1982).

7. A.H-J. Wang, GJ. Quigley, F.J. Kolpak, G. van der Marel, J.H. van Bloom and A. Rich, *Science, 211,* 171 (1981).

8. S. Neidle, H.M. Berman, and H.S. Shieh, *Nature 288,* 129 (1980); H.S. Shieh, H.M. Berman and S. Neidle, *Nucleic Acids Res. 8,* 85 (1980).

9. D.L. Beveridge, M. Mezei, P.K. Mehrotra, F.T. Marchese, G. Ravishanker, T.R. Vasu and S. Swaminathan, in *Molecular-Based Study and Prediction of Fluid Properties*, J.M. Haile and G.A. Mansoori, eds., American Chemical Society, Washington, D.C. (1983).

10. W.L. Jorgensen, *J. Am. Chem. Soc., 100,* 7824 (1978).

11. W.L. Jorgensen, *J. Am. Chem. Soc., 101,* 2016 (1979).

12. D.L. Beveridge, M. Mezei, S. Swaminathan and S.W. Harrison in *Computer Modelling of Matter,* P.G. Lykos, ed., American Chemical Society (1978).

13. D.L. Beveridge, M. Mezei, P.K. Mehrotra, F.T. Marchese, T.R. Vasu and G. Ravishanker, *Ann. N. Y. Acad. Sci. 367,* 108 (1981):

14. A.T. Hagler and J. Moult, *Nature, 272,* 222 (1978).

15. A.T. Hagler, J. Moult and D. Osguthorpe, *Biopolymers,* in press.

16. V. Madison and A.T. Hagler, submitted.

17. J. Herman and M. Vacatello, in *Water in Polymers,* S.P. Rowland, ed., ACS Symp. Ser. 127, American Chemical Society, Washington, D.C. (1980).

18. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunterer and J. Hermans, in *Intermolecular Forces,* B. Pullman, ed., D. Reidel Publ. Co., Dodrecht (1981).

19. J.M. Goodfellow, J.L. Finney and P. Barnes, *Proc. R. Soc. Lond., B214,* 213

(1982).

20. M. Mezei, S. Swaminathan and D.L. Beveridge *J. Chem. Phys., 71,* 3366 (1979).

21. P.K. Mehrotra, M. Mezei, and D.L. Beveridge, *J. Chem. Phys., 78,* 3156 (1983).

22. In principle a simulation could be constructed on a fraction of the unit cell, considering the symmetry relations between the asymmetric units. This would reduce the number of waters to be considered explicitly, but could introduce artifacts whereby areas are left artificially void or waters are held fixed unnecessarily due to the close proximity of images at symmetry axes.

23. E. Clementi, F. Cavallone, R. Scordamaglia, *J. Am. Chem. Soc., 99,* 5531 (1977).

24. R. Scordamaglia, F. Cavallone and E. Clementi, *J. Am. Chem. Soc. 99,* 5545 (1977).

25. L. Carozzo, G. Corongiu, C. Petrongolo, E. Clementi, *J. Chem. Phys., 68,* 787 (1978).

26. M. Ragazzi, D. Ferro, E. Clementi, *J. Chem. Phys., 70,* 1040 (1979), Erratum: 71, 3552 (1979).

27. E. Clementi, G. Corongiu, F. LejI, *J. Chem. Phys., 70,* 3726 (1979).

28. O. Matsuoka, C. Tbsi and E. Clementi, *Biopolymers, 17,* 33 (1978).

29. G. Corongiu, E. Clementi, *Gazz. Chien. Ital., 108,* 273 (1978).

30. O. Matsuoka, E. Clementi and M. Yoshimine, *J. Chem. Phys., 64,* 1351(1976).

31. E. Clementi, *Liquid Water Structure,* Springer Verlag, New York, 1978.

32. E. Clementi, *Computational Aspects for Large Chemical Systems,* Springer Verlag, New York, 1980.

33. S. Scheirer, in *Aggregation. Processes in Solution,* E. Wyn-Jones and J. Gormally, eds., Elsevier North-Holland.

34. N. Metropolis, A.W. RosenbIuth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys., 21,* 1087 (1953).

35. C.S. Pangali, M. Rao, and B.J. Berne, a. *Chem: Phys. Lett. 55,* 413 (1979); b. M. Rao, C.S. Pangali, and B.J. Berne, *Mol. Phys. 37,* 1773 (1979).

36. The computation of the generic mean positions used an approximate procedure designed at HC that takes account of the exchange of the water molecules among different sites during the simulation.

Date Received: July 12, 1983

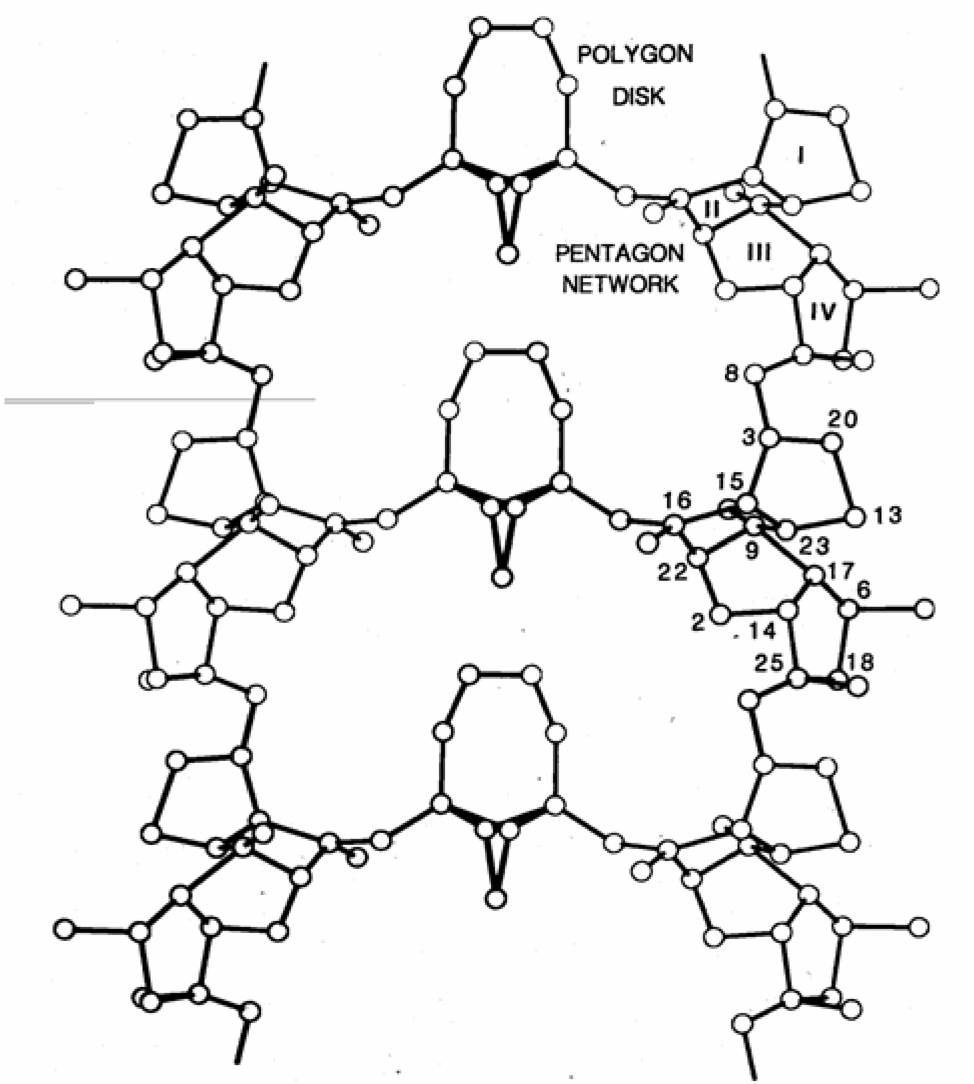

Figure 1. The water network of the dCpG/proflavin crystal hydrate as described in Ref. 8.

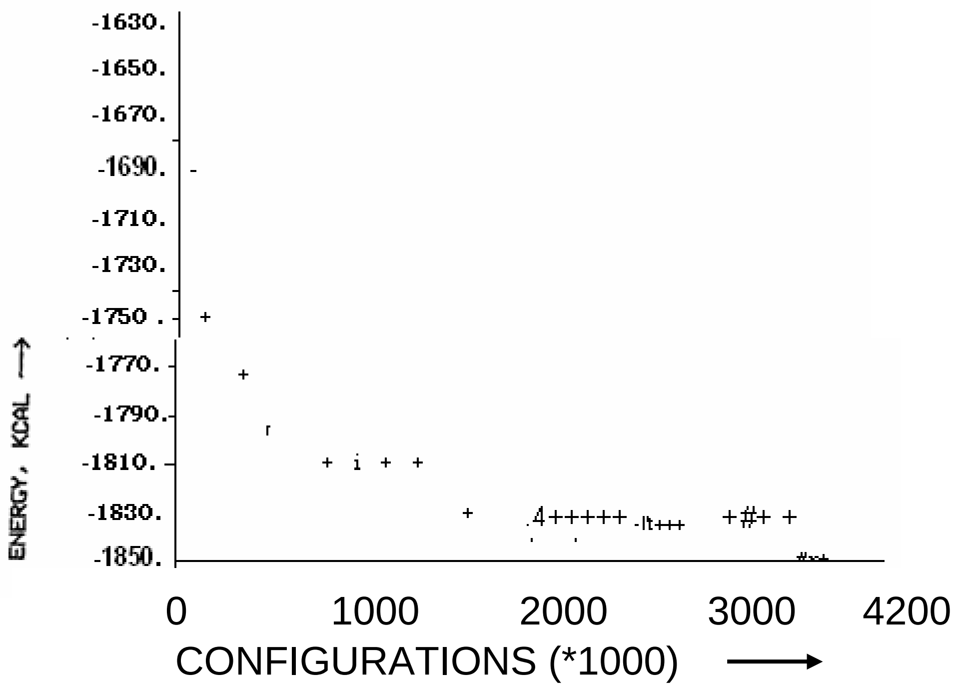

Figure 2. Convergence profile (control function) of the mean energy for simulation HC-I.

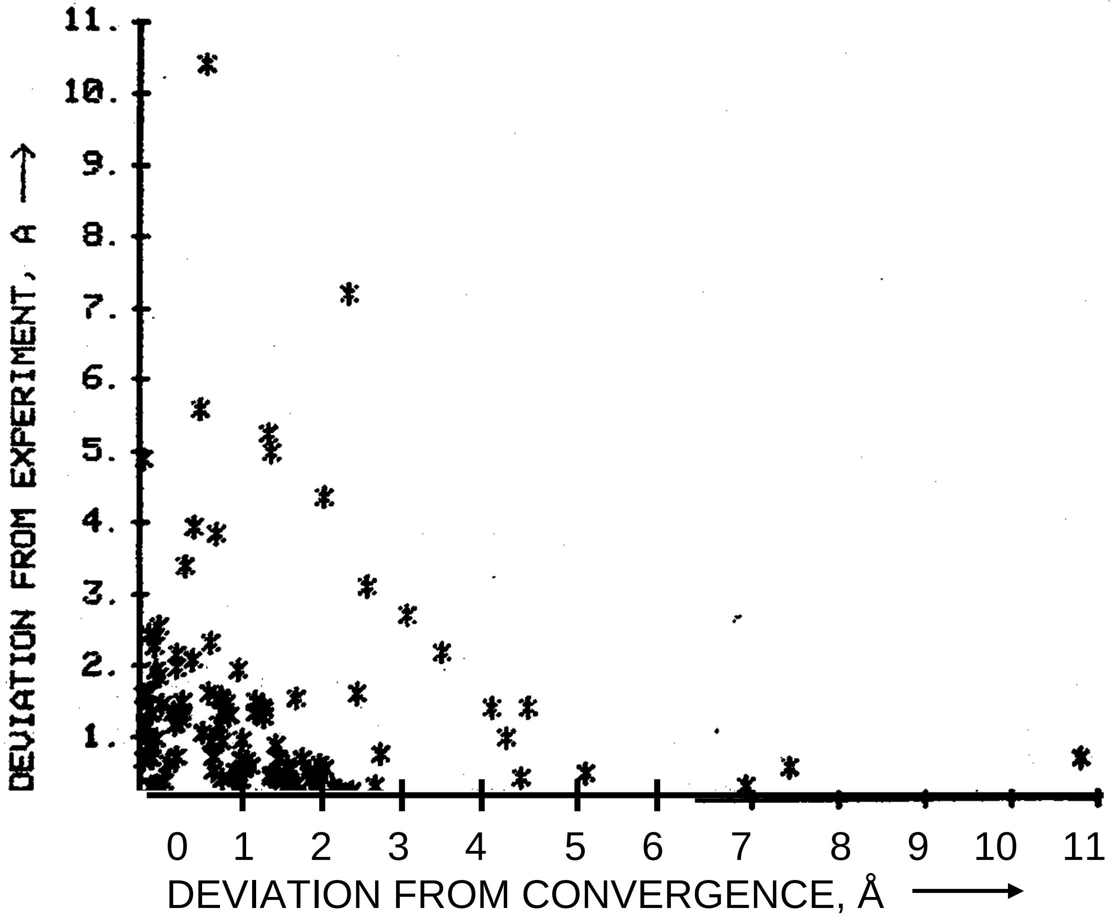

Figure 3. Plot of deviations between mean positions obtained in simulation HC-I and the experimentally obtained sites vs deviations of the mean positions between simulation HC-I and simulation HC-II.

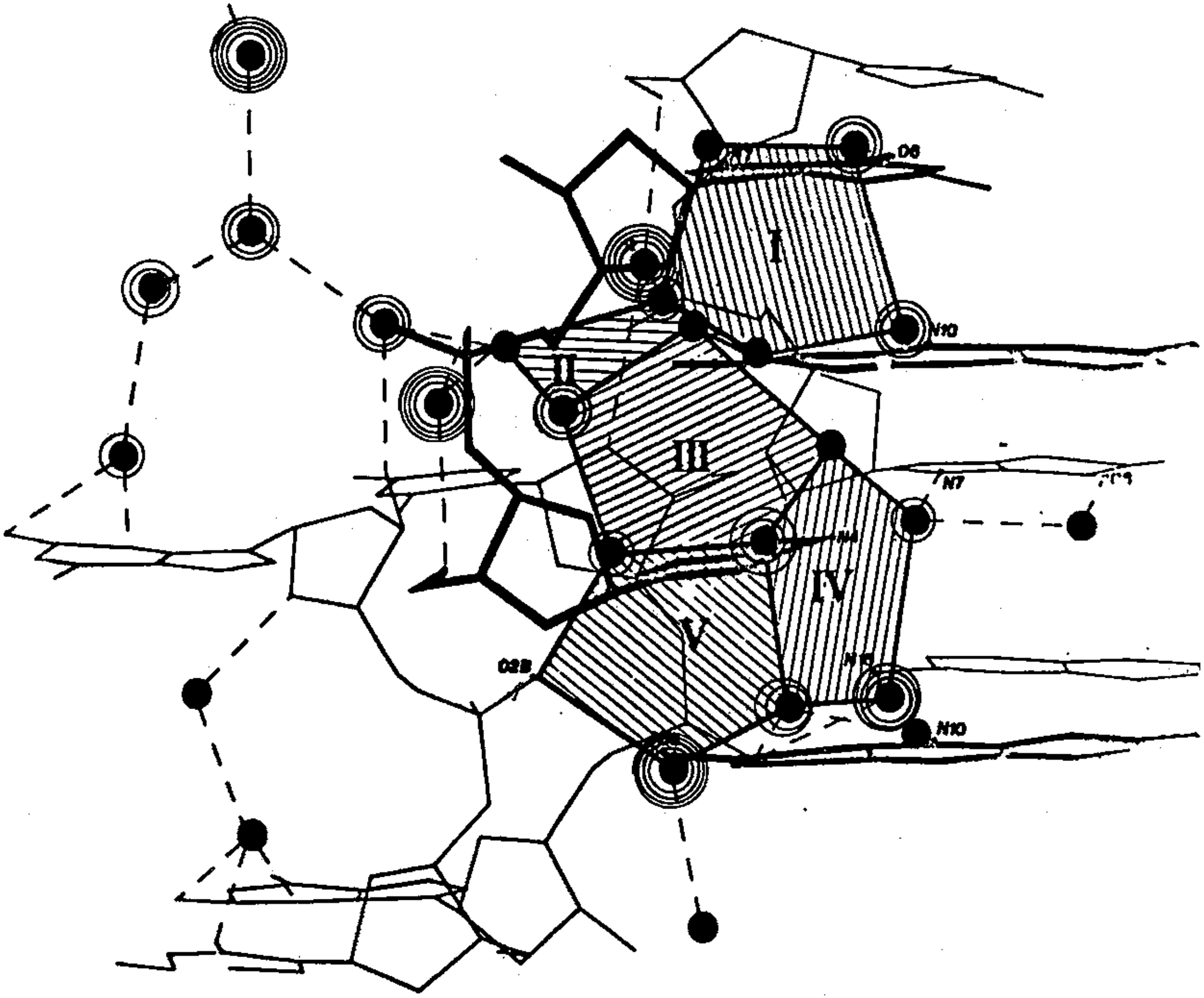

Figure 4. Site by site comparison of the experimentally obtained sites with the sites obtained in simulation HC-HI. The number of circles give the number of asymmetric units where agreement within 0.6 Å was found.