Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Conversation 11, Issue #1 (1999)
Proceedings of the Eleventh Conversation, University at Albany, SUNY, June 15-19, 1999

How Do Hydrogen Bonds Contribute to Protein-DNA Recognition?

http://www.adeninepress.com/11thcon.html

Surjit B. Dixit,
Nidhi Arora†,
and B. Jayaram*
Department of Chemistry,
Indian Institute of Technology,
Hauz Khas, New Delhi-110016, India
†Present address:
Department of Molecular Biology,
The Scripps Research Institute,
La Jolla, CA 92037, USA

Abstract

The average strength of hydrogen bonding interactions at the interface of 40 protein-DNA complexes comprising ~ 1500 potential hydrogen bonds and their free energies of formation have been estimated employing some recent advances in theoretical treatments of electrostatic interactions. The hydrogen bond spatial frequency distribution shows a maximum at a proton-acceptor distance of 2.1Å. The corresponding average interaction energy and the binding free energy are computed to be -1 kcal and +4 kcal/mol.H-bond respectively. Thus hydrogen bonds do not appear to provide the driving force for the formation of specific complexes from initially separated protein and DNAbut serve to optimize the interactions in the specific complex once it is formed, via distance and angle requirements.

Introduction

Hydrogen bonds are invoked as one of the most common modes of intra and intermolecular recognition in chemical and biochemical systems. The Watson-Crick base pairing in the DNA double helix (1), the formation of alpha helical and beta sheet secondary structures in proteins via backbone hydrogen bonds as postulated by Pauling (2), are two most cited cases. The direct code proposed (3) for protein-DNA recognition and observed in some crystal structures such as in EcoRI endonuclease - DNA complex (4), is based on hydrogen bonds. These ideas originating in structural studies correlate intermolecular recognition / stability of secondary structures to the propensity for hydrogen bond formation. A clear understanding of the role of hydrogen bonds, of course requires a complementary thermodynamic assessment - as to what extent are the structurally feasible hydrogen bonds, energetically significant. This issue of structure versus energetics as it relates to hydrogen bonds, is by no means trivial because the formation of both intra and intermolecular hydrogen bonds in aqueous medium involves loss of hydrogen bonds with solvent water. Hydrogen bonds - lack of a clear context-based molecular thermodynamic perspective on their role notwithstanding - are of fundamental interest in molecular recognition and in molecular docking experiments.

Several mutational studies in both proteins and protein-DNA complexes place the strength of hydrogen bonds at ~ -1 kcal/mol H-bond (5-7). Some recent theoretical investigations based on refined electrostatic models (8,9) led to the conclusion that hydrogen bonds do not contribute to the stability of protein secondary structures as the unfavorable desolvation expense overcomes the favorable interaction energy associated with hydrogen bond formation. An analysis of the problem in terms of interaction vis-a-vis binding simplifies the scenario considerably. Structure or energy based interpretation of the role of H-bonds in the final state (as in a complex) is an interaction problem. One way to quantify this interaction energy is to identify the atoms involved in the hydrogen bond, employ an appropriate effective dielectric function and estimate the strength. This interaction energy often turns out to be negative (10,11), indicating

*Phone: 91-11-6591505;
Fax: 91-11-6862037;
E-mail: bjayaram@chemistry.iitd.ernet.in

that hydrogen bonds are favorable as they occur in the final state. To address whether hydrogen bonds provide the necessary driving force for the formation of the complex or secondary structure in the final state from the reactants or extended chain in the initial state, both initial and final states have to be considered in the analysis (8,12). The net energetics of this process which can be equated to the binding free energy, includes direct interactions as well as desolvation expense, and in principle can be either positive (unfavorable) or negative (favorable), depending upon a differential effect of the solute-solvent interactions in the initial and final states. Whether protein-DNA complex formation can occur by virtue of direct code alone falls under the purview of the binding analysis. Additional clarification of this issue has been sought here using smaller prototypical systems such as the base pairs, where in similar roles are played by hydrogen bonding interactions and the solvent while explicit solvent simulations on protein-DNA complexes are just becoming computationally tractable.

In this study, we present an exhaustive quantification of the strength of over 1500 potential hydrogen bonds occurring at the interface of about 40 protein-DNA specific complexes and examine its consequences to protein-DNA recognition. Observations from a series of molecular dynamics simulations on base pairs in aqueous and non-aqueous media are employed to develop a molecular thermodynamic perspective on the effects of hydrogen bonding interactions and the role of the environment / solvent in biomolecular recognition mediated by H-bonds.

Materials & Methods

Crystal structures of over 40 protein-DNA complexes were adapted from NDB/PDB (13-15). These complexes were prepared for analysis as reported earlier (16). All potential hydrogen bonds weak or strong with a proton - acceptor distance of 0 to 3.0Å have been included in the analysis. This criterion results in a set of over 1500 potential hydrogen bonding interactions each of which has been examined individually and their net energetics averaged among the hydrogen bonds occurring at each separation.

Interaction Energy

Partial atomic charges and van der Waals parameters were assigned to all the interacting atoms using the most recent parameter set of AMBER force field (17,18). A modified Hingerty-Lavery distance dependent dielectric function simulating the aqueous environment developed previously (19-21) and successfully applied to assess the strength of hydrogen bonds in base pairs and alpha helices (11,12) was used to compute the electrostatic component of the H-bond interaction energy.

$$E_{elec} = q_i q_j / D(r) r_{ij} \qquad [1]$$

D(r) is a distance dependent sigmoidal dielectric function,

$$D(r) = D - \left[\left(\frac{D - D_i}{2}\right)\left(s^2 r^2 + 2sr + 2\right)e^{-sr}\right] \qquad [2]$$

where D = 78, D_i = 4, and s = 0.395. The van der Waals energy component was calculated using a (12,6) Lennard-Jones term. The estimation involved a consideration of the net interaction energy among the donor heavy atom and proton with the acceptor atom and its precedent atom. Thus in a >C=O…H-N bond, the interaction energy of the carbonyl group is computed with the NH group.

Binding Energy

The energetics of binding in a solvent medium can be estimated via a thermodynamic cycle involving four stages and three steps (Figure 1). In the first step, the donor and acceptor pairs, infinitely apart as in the initial state before complexation, are transferred from the solvent medium to vacuum. The desolvation energy accompanying this transfer process consists of three contributions. An electrostatic part that is estimated using the modified generalized Born (GB) model (22-26),

$$G^0_{el} = -166\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{m2GB}} \qquad [3]$$

where f_{m2GB} is an effective atomic size / distance parameter derived from the Born radii α_i and pair-wise distances r_{ij}. The decavitation energy, i.e. the energy gained in removing the cavity from solvent, and the concomitant loss in van der Waals interactions of the solute with solvent are estimated based on solvent accessible surface areas of the hydrogen bonding atoms. The solvent accessible surface area is calculated using the ACCESS (27) program with AMBER van der Waals radii. The cavitation and dispersion energy in water together contribute 7.2 cal/mol/Å² (24). In the estimation of the electrostatic contribution, atoms in the protein or DNA that are not involved in the hydrogen bond formation were discharged. In the second step, the hydrogen bonding pair is brought together as in the final state but in vacuum, and allowed to interact, i.e. the electrostatic and van der Waals interactions are turned on to form the complex. In the third and final step, the complex is transferred from vacuum back to the solvent and this process again involves an electrostatic term, a cavitation term and a van der Waals term signifying the corresponding interactions of the complex with the solvent. These are estimated as in the desolvation (first) step. The net average free energies of formation of hydrogen bonds at each distance of separation between the hydrogen atom and the acceptor atom together with the corresponding interaction energies in the final state are presented below.

Results and Discussion

The number of potential hydrogen bonds in all the 40 protein-DNA specific complexes considered is shown in Figure 2 as a function of separation between the hydrogen and the acceptor atom. Also shown superposed in Figure 2 are the average interaction energies (in the final state) as well as the binding free energies i.e. the free energies of hydrogen bond formation.

Several observations emerge out of this figure on the nature of

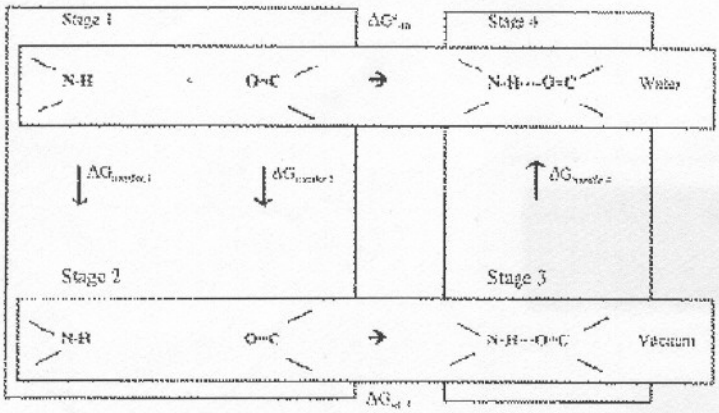

Figure 1: The thermodynamic cycle used for the estimation of the free energy of hydrogen bond formation. In Stage 1, the hydrogen bonding atoms are infinitely apart in solvent. The functional groups are transferred to vaccum (Step 1) accompanied by transfer free energies, $\Delta G_{transfer,1}$ and $\Delta G_{transfer,2}$. In Step 2, the hydrogen bonding groups interact ($\Delta G_{int,3}$) in vaccum to form the complex. Step 3 is attained on transfer of the complex back to solvent accompanied by $\Delta G_{transfer,4}$.

hydrogen bonds in protein-DNA specific complexes. The shortest hydrogen bonds occur at a proton-acceptor distance of 1.8Å. At this distance, the interaction energy is around -1.5 kcal/mol.H-bond. The maximum number of hydrogen bonds occur at a seperation of 2.1Å. The corresponding average interaction energy is ~ -1 kcal/mol.H-bond. Beyond 2.8Å, the interaction energies associated with the potential hydrogen bonds are close to zero in conformity with expectations.

The computed binding energies (free energy of hydrogen bond formation) are however positive in almost all cases. At 2.1Å, it is around 4 kcal/mol.H-bond. The unfavorable cost of the desolvation component in the process of creating a hydrogen bond strongly overrides the favorable interactions that are derived by the formation of a hydrogen bond. The inference that follows is that hydrogen bonds do not provide the necessary driving force for the formation of protein-DNA complexes in aqueous medi-

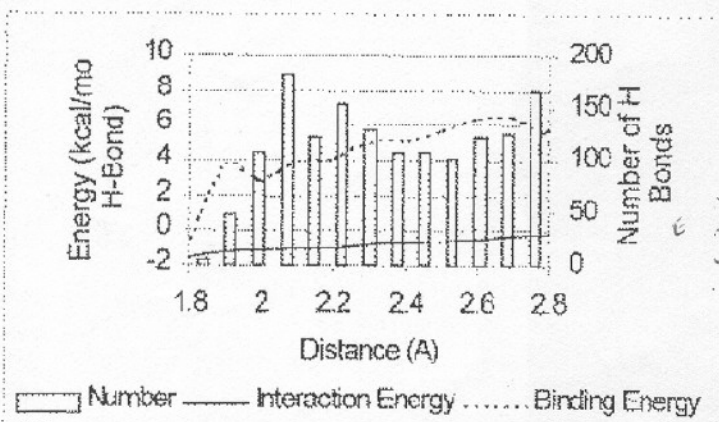

Figure 2: Hydrogen bonding at the protein-DNAinterface: Average properties of 1520 hydrogen bonds at the protein-DNAinterface with proton-acceptor atom distance less than 2.8Å are plotted as a function of the seperation. The primary Y axis carries the average binding and interaction energies and the secondaryY axis represents the number of hydrogen bonds (bar). The maximum number of hydrogen bonds are observed near 2.1Å. The interaction energies at this distance are about -1 kcal/mol and the binding energies are about 4 kcal/mol. The strength of these interactions tend to zero beyond 2.8Å.

um. Some caution however needs to be exercised in interpreting this result. Only the dipolar interactions and their desolvation are considered in estimating the binding free energies as described in figure 1 and in the methodology section. Thus hydrogen bonds formed by charged residues such as Asp, Glu, Arg, Lys and the phosphate group may exhibit different energetic characteristics when full charges are included in the theoretical treatment. Related studies on the total (all atom) energetics of protein-DNA complexation indicate that overall electrostatics is marginally unfavorable to binding but other factors such as the van der Waals and hydrophobic interactions make a favorable contribution to the net binding free energies (16).

We have performed some exploratory molecular dynamics studies on base pair formation in aqueous and non-aqueous media with the aim of understanding the effect of solvent on molecular association and hydrogen bonding. Results from the molecular dynamics simulations - performed with initial configurations involving base pairs at different separations - indicated that while the bases prefer to remain in the vicinity of each other exhibiting different orientations in aqueous environment, the bases predominantly stack. It is a well-accepted opinion that stacking is affected by van der Waals and hydrophobic / cavitation interactions between the planar molecules (28-31). This observation which is well received in the case of base pair interactions also appears to hold true in the case of protein-DNA interactions (16). In chloroform however, where the hydrophobic component is not operational and there is little electrostatic desolvation expense because of the nonaqueous and low dielectric nature of the medium, the base association is dominated by the electrostatic / hydrogen bonding interactions between the bases. Thus base paired configurations involving hydrogen bonding interactions are formed. In the context of protein-DNA complexation, the release of the first layer of water from the protein and DNA surface results in a more nonpolar environment that is conducive to establishing favorable hydrogen bond interactions. The role of hydrogen bonds in protein-DNAcomplexes thus seems to be not in driving the specific complex formation but in stabilizing the complex via favorable interactions once the final state is attained / complex is formed. Specificity due to hydrogen bonds manifests itself via an optimization of the distance and angle requirements resulting in favorable interactions. However, this picture derived from a continuum model of solvent is not extendable to water mediated hydrogen bonds reported in the literature.

Taken together with some earlier studies on protein-DNArecognition (7,16,19,32), our present study enables the formulation of a working hypothesis for protein-DNArecognition from an energetic viewpoint. Non-specific association could be driven by simple diffusion or favorable long range electrostatic interactions. Non-specific to specific complex formation among other contributors (16), involves desolvation (7), the electrostatics of which is unfavorable (16,33). It is here that the van der Waals (packing/steric) and hydrophobic interactions contribute favorably. After the complex is formed hydrogen bonds are expected to stabilize the final state in case of specific complexes.

Conclusions

Based on an analysis of over 1500 hydrogen bonds at the protein-DNA interfaces in 40 co-crystal structures of protein-DNAcomplexes, it is observed that the maximum number of hydrogen bonds occur at a proton acceptor atom separation of 2.1Å. On the basis of an empirical energy function, the interaction energy at this separation is estimated to be ~ - 1 kcal/mol.H-bond. The corresponding free energy of hydrogen bond formation between dipolar groups turns out to be ~ + 4 kcal/mol.H-bond, mainly due to the unfavorable nature of the electrostatic component in the hydrogen bonding group desolvation. Although hydrogen bonding does not appear to drive complex formation, the presence of hydrogen bonds at the protein-DNAinterface facilitates recognition in specific complexes due to the favorable interaction energy between the hydrogen bonding groups.

Acknowledgements

Funding from the Indo-French Centre for the Promotion of Advanced Research (Centre Franco-Indien pour la Promotion de la Recherche Avancee) is gratefully acknowledged. SBD wishes to thank the Council for Scientific and Industrial Research and the Department of Science and Technology, Govt. of India, for travel grant to attend the X1 Conversation in ~~Discipline~~ Biomolecular Structure and Dynamics. The authors wish to express their gratitude to Prof. D. L. Beveridge, Prof. Barry Honig and Mr. K. McConnell for many helpful discussions and useful suggestions on the subject matter.

References and Footnotes

1. J.D. Watson and F.H.C. Crick, *Nature 171*, 737-738 (1953).
2. L. Pauling and R.B. Corey, *Proc. Natl. Acad. Sci., USA 37*, 235 (1951).
3. N.C. Seeman, J.M. Rosenberg and A. Rich, *Proc. Natl. Acad. Sci. USA 73*, 804-808 (1976).
4. J.A. McClarin, C.A. Frederick, B.-C. Wang, P. Greene, H.W. Boyer, J. Grabel and J.M. Rosenberg, *Science 249*, 1307-1309 (1986).
5. R. Fersht and L. Serrano, *Curr. Opin. Struct. Biol. 3*, 75-83 (1993).
6. J.S. Thorson, E. Chapman and P.G. Schultz, *J. Am. Chem. Soc. 117*, 9361-9362 (1995).
7. L. Jen-Jacobsen, *Biopolymers 44*, 153-180 (1997).
8. A.Yang, and B. Honig, *J. Mol. Biol. 252*, 351-365 (1995).
9. A. Yang, B. Hitz and B. Honig, *J. Mol. Biol. 259*, 873-882 (1997).
10. B. Honig and A. Nicholls, *Science 268*, 1144 (1995).
11. N. Arora and B. Jayaram, *J. Comp. Chem., 18*, 1245-1252. (1997)
12. N. Arora and B. Jayaram, *J. Phys. Chem. B, 102*, 6139-6144 (1998).
13. H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A.R. Srinivasan and B. Schneider, *Biophys. J., 63*, 751-759 (1992).
14. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr, M.D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol. 112*, 535-542 (1977).
15. NDB Codes of complexes studied: e001; e002; e003; e009; e0115; e017; e020; e022; e025; eb08; r001; r004; r008; r009; r010; r011; r012; r015; r016; r021; r023; r027; rc01; rc03; t003; t004; t005; t007; t009; t011; t014; t016; t017; t028; t029; t031; t038; t039; t040; tb41
16. B. Jayaram, K. McConnell, S.B. Dixit and D.L. Beveridge, *J. Compu. Phys. 151*, 333-357 (1999).
17. A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheathem III, D. M. Ferguson, G.E. Seibel, U.C. Singh, P.K. Weiner, and P.A. Kollman, (1995) Amber 4.1, University of California, San Francisco.
18. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *J. Am. Chem. Soc. 117*, 5179-5197 (1995).
19. B. Jayaram, A. Das and N. Aneja, *J. Mol. Struc. (THEOCHEM) 361*, 249-258 (1996).
20. E. Hingerty, R.H. Ritchie, T.L. Ferrel and J E. Turner, *Biopolymers. 24*, 427-439 (1985).
21. D. Flatters, K. Zakrzewska and R. Lavery *J. Comp. Chem. 18*, 1043-1055 (1997).
22. B. Jayaram, Y. Liu and D.L. Beveridge, *J. Chem. Phys. 109*, 1465-1471 (1998).
23. B. Jayaram, D. Sprous, D.L. Beveridge, *J. Phys. Chem. 102*, 9571-9576 (1998).
24. W.C. Still, A. Tempczyk, R.C. Hawley and T.J. Hendrickson, *Am. Chem. Soc. 112*, 6127-6129 (1990).
25. G.D. Hawkins, C.J. Cramer and D.G. Truhlar, *J. Phys. Chem. 100*, 19824-19839 (1996).
26. G.D. Hawkins, C.J. Cramer and D.G. Truhlar, *Chem. Phys. Lett. 246*, 122-129 (1995).
27. B.K. Lee and F.M. Richards, *J. Mol. Biol. 55*, 379-406 (1971).
28. R.A. Freidman and B. Honig, *Biopolymers 32*, 145-159 (1988).
29. P. Cieplak and P.A. Kollman, *J. Am. Chem. Soc. 110*, 3734-3739 (1988).
30. J. Norberg and L. Nilson, *Biophys J. 74*, 394-402 (1998).
31. J. Šponer, H.A. Gabb, J. Leszczynski and P. Hobza, *Biophys. J. 73*, 76-87 (1997).
32. P.H. von Hippel, *Science 263*, 769-770 (1994).
33. V.K. Misra, J.L. Hecht, K.A. Sharp, R.A. Friedman and B. Honig, *J. Mol. Biol. 238*, 264-280 (1994).