# Theoretical Studies on the Interaction and Binding Energies in Protein-DNA Complexes

by

**Dixit Surjit Bhima Rao**

Department of Chemistry

Submitted

in fulfillment of the requirements of the degree of

**Doctor of Philosophy**

to the

**Indian Institute of Technology, Delhi**

**May 1999**

# CERTIFICATE

This is to certify that the thesis entitled “Theoretical Studies on the Interaction and Binding Energies in Protein-DNA Complexes”, being submitted by Mr. Surjit B. Dixit to the Indian Institute of Technology, Delhi for the award of the degree of Doctor of Philosophy in Chemistry is a record of bonafide research work carried out by him. Mr. Surjit B. Dixit has worked under my guidance and supervision, and has fulfilled the requirements for the submission of this thesis, which to my knowledge, has reached the requisite standard.

The results contained in this dissertation have not been submitted in part or full to any other university or Institute for the award of any degree or diploma.

Date: May 13, 1999

**B. Jayaram**
Associate Professor
Department of Chemistry
Indian Institute of Technology, Delhi

# Acknowledgement

This thesis was prepared under the tutelage of Dr. B. Jayaram. His consistent and practical advice, endless patience, and confidence were essential for completion of this thesis. He provided a great environment for active research. For all his guidance, attention, insight and support over the years, I wish to express my sincerest appreciation. His effort / good nature to talk through my ideas with great benefit to my clarity of mind has been extremely beneficial to my research activity.

My thanks to the Head of Department – Chemistry, IITD for providing me full access to the departmental facilities. Without his cooperation and willingness to be helpful no such research as my own is possible.

I am indebted to all my teachers for their goodwill and role in providing me with the bigger perspective of science and research. I wish express my special appreciation for Dr. Nalin Pant, my first teacher of biochemistry, for all the useful discussions and for impressing upon me the nature of a chemist to "see" molecules.

For the timely help and support – acting as my sounding board, my collegues at the Lab and friends have helped me in innumerable ways, especially Achintya Das. His diligence in helping me through my struggles with the computing issue was immensely helpful. My sincere thanks to Nidhi Arora, Piyush Shukla, Parul Kalra, Harish Hirani, Shashank Deep and Patrick Pancreas.

I would like to take this opportunity to thank my family especially parents for giving me advice and support during all these years. They had to endure my moods while I was frustrated and confused when the research didn't go quite as expected. To them, I dedicate this thesis.

Lot of work in this research thesis was done using the Amber program suite. Dr. Bill Ross and Prof. Peter Kollman, UCSF, were kind enough to provide me this wonderful tool at a crucial juncture when I needed it the most – I thank them for the same. I sincerely thank Prof. W. F. van Gunsteren, ETHZ, for providing me the GROMOS '96 package.

I would like to acknowledge the scholarship from the Indian Institute of Technology, Delhi that supported me during the course of this research. The financial support received from the Indo-French Center for the Promotion of Advanced Research (Center Franco-Indien pour la Promotion de la Recherche Advancee) in the form of a grant to Dr. B. Jayaram for creating research facilities used in this thesis work is gratefully acknowledged.

Surjit Dixit

# ABSTRACT

Specific recognition of DNA base sequences by proteins to form complexes is the hallmark of regulated gene expression. From a thermodynamic perspective, the process of protein-DNA recognition is associated with a free energy of binding. Development of methodologies for a theoretical determination of the binding free energies is a challenging problem but finds applications in several promising areas of research and development such as structure based drug design. In addition to a methodological development for a rapid estimation of the binding free energies of protein-DNA complexes, this thesis is aimed at proffering some insights into the energetic features that provide the impetus for specific recognition, complementing the structural perspective. This thesis is divided into five chapters with the contents therein detailed below.

The introduction to the thesis in chapter 1 is based on published literature and surveys the developments in our understanding of protein-DNA interactions. Results from structural studies of co-crystal complexes, experimental studies on binding, and issues related to theoretical and computer simulations of biomolecules are presented.

Information on molecular level specificity in protein-DNA systems so far has been mainly confined to structural analyses of the complexes i.e. the final state. As an essential first step towards developing an energetic perspective of recognition, a computational strategy is presented in chapter 2. A final state analysis of some protein-DNA complexes containing the *helix-turn-helix* motif has been presented based on an easy to implement empirical potential energy function mainly to quantify the different contributions to the net interaction energy in the final state. The strength of hydrogen bond interactions at the interface of 40 protein-DNA complexes has also

been studied using this potential function. The protein-DNA intermolecular interaction energy and the desolvation energy effects act in a compensating fashion. The effective van der Waals energy changes as a result of binding are negative and favor binding while the net electrostatics are strongly base sequence / system dependent and tend to be unfavorable in general. The average strength of the hydrogen bonds at a proton-acceptor separation of 2.1 Å where the maximum number of hydrogen bonds are observed is computed to be about –1 kcal/mol, close to a value elicited from experimental studies.

In chapter 3 the total free energy change accompanying the protein-DNA binding process has been estimated using a thermodynamic cycle comprising four main stages split further into seven steps and 26 well defined components. The model involves explicit all-atom accounts of the energetics of van der Waals and electrostatic interactions between the protein and the DNA, electrostatic, hydrophobic and van der Waals components of desolvation, plus consideration of condensed counter ions, the ion release on complexation, added salt effects and structural adaptation. Additionally, estimates of entropy loss due to decreased translational, rotational and vibrational degrees of freedom in the complex relative to the unbound species are included. The computations take advantage of some recent theoretical developments such as the second generation force fields and a compatible generalized Born solvation treatment.

The net binding free energies of the EcoRI endonuclease - DNA complex and 43 other complexes have been evaluated. The net binding free energies thus obtained are seen to be a resultant of large compensating factors. The calculated standard free energy of complex formation agrees with experiment to within 5 kcal/mol for the EcoRI system. Overall, the van der Waals interactions and water release favor complexation, while electrostatic interactions, considering both intramolecular and solvation effects, prove unfavorable. A detailed sketch of the methodology, the results on each system followed by a consensus view of protein-DNA complexation is presented.

Discussions of specificity invariably implicate hydrogen bonds. It thus becomes interesting to ask whether hydrogen bonds can drive specific complex formation. In chapter 4 of the thesis the role of hydrogen bonds in complex formation is discussed. The formation of hydrogen bonds at the protein-DNA interface in aqueous medium involves not only favorable interactions of the donor and acceptor functional groups (in the final state) but also a loss of interactions of these groups with the solvent water (as present in the initial state). The contribution of about 1500 potential hydrogen bonds occurring at protein-DNA interfaces to the net binding free energy is investigated employing the generalized Born and the finite difference Poisson-Boltzmann treatment of solvation.

The average hydrogen bond contribution to the binding free energy at a proton-acceptor atom distance of 2.1 Å is computed to be about +4 kcal/mol using the generalized Born treatment. A similar trend is observed on using the finite difference Poisson-Boltzmann method for estimation of the hydrogen bond strength. This result taken together with the computed favorable interaction energies discussed in chapter 2, supports the emerging view that hydrogen bonds do not drive protein-DNA complex formation from initially separated protein and DNA by virtue of the unfavorable cost of desolvation but serve to optimize the interactions in the complex once it is formed, via distance and angle requirements.

The distinct energetic characteristics of hydrogen bonds in interaction and binding in protein-DNA complexes necessitate further studies on the role of hydrogen bonds in biomolecular recognition - in particular the regions where hydrogen bond interactions are favorable and conditions under which the unfavorable desolvation effects take over. In chapter 5, the issue of hydrogen bonding is revisited with extensive molecular dynamics (MD) simulations on base pairs in chloroform and water as prototypes. The MD simulations are then utilized for a *post facto* free energy component analysis. The MD trajectory of an A-T base pair in water shows that the Watson-Crick hydrogen bonds are disrupted and the base pair prefers the stacked conformation in water while the hydrogen bonded configuration is maintained in chloroform. While the strength of association of the base pair in water at the minimum in the free energy curve is about –2.0 kcal/mol (corresponding to the stacked rather than a paired configuration), the minimum association free energy of the base pair in chloroform is estimated to be about –8.0 kcal/mol, qualitatively matching the reported experimental values. This study strongly suggests that hydrogen bonds do not drive complex formation in the presence of a polar solvent like water, they however preferentially promote molecular association in a nonpolar environment.

Solvation plays a crucial role in the thermodynamics of recognition. The finite difference Poisson Boltzmann methodology offers an accurate approach for the calculation of solvation free energies at the continuum solvent level. As a precursor to the protein-DNA problem and to appreciate force field dependence in the calculated energetics, the transfer free energies of the 20 amino acid side chains were characterized with charges and size parameters from AMBER, CHARMM, CVFF, GROMOS, OPLS, PARSE and an *ab initio* derived charge set and is presented in an appendix to the thesis. A general agreement between the theoretical predictions, emerging from each of the parameter sets, and experiment is discernible. A detailed discussion of the sensitivity of the results to the choice of the force field is presented.

The thesis provides a new perspective on the factor governing specificity in protein-DNA complexes on the basis of the energetics calculated using some state-of-art theoretical molecular simulation techniques. Apart from offering insights into the contribution and role of the diverse energetic components to binding, the thesis points to the need for further explorations to comprehensively understand these complex biomolecules and their processes.

# CONTENTS