THEORETICAL STUDIES ON THE ENERGETICS OF SPECIFICITY IN SOME PROTEIN-DNA SYSTEMS

by

NIDHI ARORA
DEPARTMENT OF CHEMISTRY

SUBMITTED
IN FULFILMENT OF THE REQUIREMENTS
OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

to the

INDIAN INSTITUTE OF TECHNOLOGY, DELHI

INDIA

March 1997

To

My Husband

&

My Parents

CERTIFICATE

This is to certify that the thesis entitled, **"THEORETICAL STUDIES ON THE ENERGETICS OF SPECIFICITY IN SOME PROTEIN-DNA SYSTEMS"**, being submitted by **Ms. NIDHI ARORA** 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 her. Ms. Nidhi Arora 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: April 2, 1997

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

ACKNOWLEDGEMENTS

I was like a child on the crossroads, unable to decide which road was the right one. It was Dr. B. Jayaram, who by his effusive enthusiasm and clear understanding of the subject guided me in the right direction in research. He moulded my ideas, clearing my concepts, removing the cobwebs and in the process constantly encouraging independent thought and work. His mature criticism was one of the key factors that transformed the vague concepts in my mind into this presentable form. I wish to express my deep sense of gratitude and respect to Dr. B. Jayaram for his patience, understanding and able guidance throughout this work.

It was Prof. M. N. Gupta who introduced me to the field of research, built in me the confidence for independent work and to accept no compromises in research. I am grateful to him for his constant encouragement throughout my academic career at IIT Delhi.

My sincere thanks to the Head, Department of Chemistry for providing the necessary facilities and to all faculty members for their valuable advice and cooperation.

This work could not have been carried out without the help and support of my friends especially Sandhya Jain, Achintya Das, Jyotsna Rikhi, Sandhya Sanghi, Sadhna Sharma and Surjit B. Dixit who gave me a patient hearing during my stressful days and infused me with enthusiasm. It will be too little to say "thanks" to them.

I am immensely thankful to Satish Bhardwaj, Ravi Kiran, Anil Kumar, E. Rajasekaran, Rajni Madan, Renu Batra, Shashank Deep, B. K. Pradhan, Pramod Sharma, Abhishek Upadhyay, Sunil Garg and Kaushik Dutta who have been very helpful to me in numerous ways.

I owe a great deal to my in-laws, my husband, my brother, my sister and my parents for being a sturdy support at all times and for their immense cooperation during the course of this work. Words are too little to appreciate their efforts.

Many a thanks are due to all members of the department staff for their kind cooperation throughout the course of my work.

Financial assistance from the Department of Science and Technology (DST) and the Indian Institute of Technology, Delhi is gratefully acknowledged.

Nidhi Arora

Nidhi Arora

ABSTRACT

Protein-DNA interactions occuring at all levels of DNA expression and replication are crucial determinants for the coordinated and controlled function of cells, and are therefore extremely important in the life of a cell. The remarkable specificity with which restriction endonucleases bind DNA and cut it at specific sites, the affinity of repressor to its cognate operator, and the identification of lesions in DNA followed by their repair are but a few instances of the biochemical processes which are intriguing and are, of late, responsible for a surge in experimental and theoretical studies directed towards obtaining a complete picture of the interactions taking place at the molecular level. In this regard, both theory as well as experiment have benefitted significantly with the availability of atomic coordinates of a number of protein-DNA complexes. These coupled with the availability of ever increasing computer power and with advances made in obtaining numerical solutions to biomolecular problems have enhanced our understanding considerably, but no rules for specific recognition have hitherto been defined. The current status on protein-DNA specificity is reviewed in Chapter I. The present work takes a closer look at specificity at the molecular level by forming an energetic perspective of these interactions.

Estimation of the thermodynamic parameters of binding requires a potential of mean force between the two interacting macromolecules. Several forcefields perform well with proteins but not with DNA because of the dominance of electrostatic interactions in the latter and the indispensability of a satisfactory account of solvent and counterions. For an expeditious and accurate estimation of the energetics of nucleic acid systems, a distance dependent dielectric function, D(r) in the electrostatic term of the forcefield, with a proper

calibration, has been recommended by several workers, in lieu of explicit molecular solvent representation. Based on the collective experience in modeling small molecules and DNA with and without molecular solvent emerging from diverse Laboratories, we have put together a computationally simple forcefield exclusively for protein-DNA energetics. The methodology adopted is detailed in Chapter II. The minimum that is expected of such a dielectric function is to reproduce the base pairing energies in solution. Base pairing thus provided a convenient testing ground.

The base pairing energies in poly (dA)-poly(dT) and poly (dG)-poly(dC) homopolymers have been examined (Chapter III) using parameters from recent versions of AMBER, CHARMM, GROMOS and OPLS forcefields and employing different dielectric continuum models namely D(r) = 1., 4., 80., r and MHLF. An analysis of the base pairing energetics in canonical B-DNA suggests that a dielectric constant of unity {D(r)=1} or a distance dependent dielectric function {D(r)=r}mimic gas phase quite well while a modified Hingerty-Lavery function (MHLF) is suitable for solution conditions. Other uniform dielectric functions are not so satisfactory. The finite difference Poisson-Boltzmann calculations and the MHLF results indicate that the electrostatic contribution to the base pairing energy in B-DNA in solution is in the range of -2.0 to -3.0 kcal/mol per H-bond.

After characterizing our approach with the latest forcefields by estimating the base pairing energetics, it was deemed important to gain the same level of confidence in estimating intra-protein energetics before proceeding to the more complex protein-DNA systems. The CO...HN interaction energies for topological neighbours involved in *(i, i+4)* H-bond interactions were estimated for a nonapeptide of poly (L-alanine) in α_R

conformation (Chapter IV). The total H-bond energetics is calculated as a sum of the contributions due to electrostatic and van der Waals interactions between the carbonyl of the i^{th} residue and amide of the $(i+4)^{th}$ residue employing parameter sets from AMBER, GROMOS, CHARMM, OPLS and ECEPP force fields using different dielectric continuum models. All the forcefields capture the energetics in solution (~1.0 kcal/mol) well within the experimentally predicted value of -0.5 to -2.0 kcal/mol. These are also supported by the FDPB values indicating the predominant role played by electrostatics in H-bonding. The modified Hingerty-Lavery function estimates a H-bond strength of ~ -1.0 kcal/mol in α-helices while the same function leads to a value of -2.0 kcal/mol in base pairing without any modifications or extra parameterization, indicating the role of contextual effects and geometry on the H-bond energies.

As the interactions in both DNA and protein in solution were captured in close correlation with experiment, we applied our methodology to the well characterized λ repressor-OL1 system as a test case for studying protein-DNA interactions using OPLS parameters (Chapter V). While confirming most of the H-bond interactions noted in the X-ray studies, this investigation led to an identification of the turn-helix3-turn rather than the helix2-turn-helix3 as being more important energetically for specific binding. Lock and key hypothesis puts forth structural complementarity as a hall mark of biomolecular recognition and this, as supported by the significantly high van der Waals contribution superposed on a hydrogen bond matrix, appears to be an essential mode of recognition in the λ repressor-operator system. A partitioning of the energetics into contributions due to phosphodiester group, sugar and bases reveals that the phosphate backbone plays a significant role in guiding the protein towards the DNA. In fact the maximum number of H-bonds are with the phosphodiester group which contribute largely to the total

electrostatic interaction. The sugars however show negligible contribution. Binding energetics implicate base atoms as being critical determinants of specificity with a van der Waals contribution of ~ 58% to the total van der Waals interaction energy. A similar partitioning for the protein subunits reveals that the N-terminal arms make a large electrostatic contribution (~ 82% of total electrostatic interaction energy) to the total energetics besides significant van der Waals interactions. The h-t-h motif interacts mainly through van der Waals (~ 48% of the total van der Waal energy) with relatively minor electrostatic and hydrophobic contributions. The t-h-t motif (containing the recognition helix and flanking turns) however emerges as the most important contributor to van der Waals interactions (~ 58 %) also aided in binding by the favourable electrostatic and hydrophobic interactions.

Intermolecular interaction energies of 11 protein-DNA complexes have been evaluated (Chapter VI) with a view to determining some common principles, if any, which may be responsible for specific recognition of certain DNA sequences in preference to others, from an energetic point of view. Results reveal that irrespective of the number of residues or bases involved in complexation, the interaction energies in all the complexes fall in a very close range implicating the presence of restricting factors that define protein-DNA interactions. An important conclusion that emerges is that packing effects at the protein-DNA interface manifested via the van der Waals interactions contribute equally or more strongly than electrostatics.

A partitioning of the total energy into contributions due to direct and indirect codes reveals that more than the direct code (base-residue interactions) that contributes significantly but it is the amino acid side chain - DNA backbone interactions which are

responsible for a maximum contribution to the total interaction energy. Over all, these two components together contribute a significant 70-90% to the total energetics in diverse systems studied here with different force field parameters. The conclusion that the turn beyond the recognition helix is important for recognition is not only a feature of λ repressor -operator complex but a feature of all the helix-turn-helix proteins considered here. Hydrogen bonds are routinely implicated in processes involving molecular recognition. The strength of the hydrogen bonds at the protein-DNA interface is determined here to be of the order of ~ -1.0 kcal/mol/H-bond.

CONTENTS