**Manish Datt (2010).** ***Modeling protein-DNA recognition -an in Silico approach.*** Ph.D. Thesis. CSIR-IMTECH, Chandigarh/ Jawaharlal Nehru University, New Delhi: India.
**Supervisor: Dr. Balwinder Singh** **Acc. No. : Th-225**

## SUMMARY OF THE THESIS

DNA binding proteins exhibit exquisite form and flexibility. Different structural features are adopted to recognize diverse DNA sequences. For binding to palindromic stretch of DNA, proteins generally dimerize to bind two sites on the palindrome. Various transcription factors viz. Gal4, Leu3, Hap1, etc. have been shown to homodimerize for binding to target binding site (Schjerling and Holmberg 1996; Todd and Andrianopoulos 1997; Akache, Wu et al. 2001). Dimerization and DNA binding can be achieved in two possible ways – either as monomer or dimer in solution, such that in former case homodimerize only in presence of cognate binding site while in case of latter a preformed dimer binds to DNA. It is intriguing to investigate the mechanistic details of protein-DNA interaction when the protein dimerizes on DNA. Particularly, to check for differences in binding of each monomer to their respective sites and the events leading to the final complex. In this study, an attempt has been made to address these issues working with a homodimer bound to palindromic stretch of DNA. The structural details of this complex has been studied using X-ray diffraction and has suggested that the protein exist as monomer in solution while dimerize on binding to DNA (Wang, Grant et al. 2001).

Analysis of crystal structure having zinc finger containing dimeric protein bound to palindromic stretch of DNA has shown no conformational differences between the two monomers. Also, the half sites of the palindrome are structurally similar. This proves that both the monomers can interact in similar manner with each half site. But, interestingly, there are subtle differences in the interactions of two monomers with the DNA. These include additional non-bonded contacts between residues of one of the

monomer with DNA which are absent for corresponding residues of other monomer. These differences raises some questions like (a) what is the source of these different binding mode of the two monomers, and (b) the events leading to dimerization of monomeric protein on DNA. It is very difficult to analyze these events using experimental techniques and therefore *in silico* approaches are well suited for these studies. Molecular dynamics simulations have been performed to elucidate differences between the interactions of two monomers with DNA. Four different trajectories have been generated starting with different combinations of monomers of protein and DNA. The strategy adopted here presumes that, being a monomer in solution, the protein binds to DNA in a step-wise manner. That is, in the first step, one of the monomer binds to one of the half site of palindromic DNA and subsequently the second monomer freezes the complex by binding to the other half site on DNA. Simulation trajectories have been analyzed to find structural and thermodynamical differences between interactions of the two monomers with DNA. The results, envisage the possible events leading to the final protein-DNA complex.

The two monomers bind to DNA one after another – implying that first one binds to DNA while the second one binds to protein-DNA complex. It is reasonable to assume that the first monomer makes stable contacts with DNA so as to facilitate binding of second monomer. RMSD calculations of DNA backbone during the course of different simulations testify this point. The structure of DNA in simulation of DNA alone and that of DNA with both monomers are similar to trimer_G and trimer_H simulations, respectively. It has been observed that structure of DNA with one of the monomer (monomer H) remains similar to that in the final complex. The monomer H possibly

binds first and presents DNA conformation for the final complex with the second monomer. This is substantiated by the fact that variations in RMSD of trimer_G simulations have been similar to that of free DNA. In trimer_G simulation, absence of monomer H leads to loss of conformation of DNA as observed in the final complex.

Interactions of the monomers with DNA have been analyzed over the course of trajectory. This has been done in two parts – 1) backbone interactions between protein and DNA, and 2) variations in interface of protein and DNA. Backbone interactions between protein (Cα atoms) and DNA (P atoms) have been monitored in trimer_G and trimer_H simulations at a threshold distance of 7Å. It has been observed that the trimer_H simulation has more number of persistent interactions between P and CA atoms than trimer_G simulation. The persistencies in interactions have been quantified by calculating their occupancy over the course of trajectory. These observations clearly suggest that the monomer H forms more stable contacts with DNA as compared to monomer G. Interface alignment of trimer_G and trimer_H snapshots with the crystal structure further corroborate this hypothesis. IAS calculated for trimer_H simulations are closer to native IAS than IAS calculated for trimer_G simulation. IAS for complex of monomer G and DNA with itself has been 460 while for complex of monomer H and DNA with itself is 450. During simulation, the monomer G and DNA interface gets disrupted as evident by the drastic fall in IAS. IAS for trimer_H simulation also decrease but to lesser extent. These observations appears to advocate that monomer G require presence of monomer H for proper binding to DNA since the removal of monomer H disturbs the association of monomer G with DNA.

Molecular dynamics of Zif268 with selected peptide extension bound to palindromic DNA has been performed to study the dimerization on DNA. By simulating protein-DNA complex and individual components, the events that lead to complex formation can be substatiated. Based on the MD results and thermodynamics of protein-DNA interaction, it may be concluded that one of the monomer (chain H) binds to DNA first followed by binding of other monomer (chain G). The interaction between the two monomers and DNA explains the thermodynamic stability of complex of monomer (chain H) with DNA as compared to binding of other (chain G).