A COMPUTATIONAL STUDY OF DNA-PROTEIN INTERACTIONS

SHAYONI DUTTA

DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY DELHI

October 2016

A COMPUTATIONAL STUDY OF DNA-PROTEIN INTERACTIONS

by

SHAYONI DUTTA

Department of Biochemical Engineering and Biotechnology

Submitted
in fulfilment of the requirements of the degree of Doctor of Philosophy
to the

Indian Institute of Technology Delhi

October 2016

CERTIFICATE

This is to certify that the thesis entitled **"A computational study of DNA-Protein interactions"** being submitted by **Ms. Shayoni Dutta** to the Indian Institute of Technology Delhi for the award of the degree of "**Doctor of Philosophy**", is a record of the bonafide research work carried out by her, which has been prepared under my supervision in conformity with the rules and regulations of the Indian Institute of Technology Delhi. The research reports and the results presented in this thesis have not been submitted for any degree or diploma in any other University or Institute.

Dr. D. Sundar

Department of Biochemical Engineering and Biotechnology
I.I.T Delhi

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Dr. D Sundar for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Gopal Agarwal, Dr. Ritu Kulshreshtha, and Dr. Gitanjali Yadav (NIPGR, New Delhi), not only for their insightful comments and encouragement, but also for the hard questions which motivated me to widen my research from various perspectives.

I thank my fellow lab-mates especially Jaspreet, Spandan and Harsh for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four and half years. Also my special thanks to my good friend and lab mate Shashank for his innumerable contribution and support as well as brainstorming discussions, truly without whom I would have never been able to complete my Ph.D.In particular, I am grateful to Vidhi and Anjani for ensuring that I never lost hope and in keeping the lab environment full of energy, love and fun.

Last but not the least; I would like to thank my family: my parents and my brother for supporting me spiritually throughout writing this thesis and my life in general. My mother in fact has been the biggest support in my life, without whom my existence holds no worth.

Shayoni Dutta

ABSTRACT

Technology for specifically correcting 'faulty' bases or mutations within genes of humans has been the 'Holy Grail' in genetic medicine. Gene targeting using zinc finger nucleases (ZFNs) - proteins custom-designed to cut at specific DNA sequences – appears to make this possible. These "artificial" proteins combine the non-specific endonuclease activity of *Fok*I restriction enzyme with the ability of zinc-finger (ZF) motifs to specifically recognize a DNA triplet sequence. The Cys_2His_2 ZF motif binds specific sequences in DNA by virtue of its unique *modular* 30 amino acid structure (stabilized by a zinc ion), the α-helix inserting into the major groove of the DNA double helix. Amino acids involved in DNA recognition within the α-helix of the ZF motifs can be changed while maintaining the remaining amino acids as a consensus backbone to generate ZF motifs with new triplet sequence specificities. Normally, three such ZF motifs are linked together in tandem to generate a ZF protein (ZFP) that binds to a 9-bp DNA target site, which is a composite of the individual DNA triplet sub-sites recognized by each of the three ZF motifs. Binding of two three-finger ZFN monomers, each recognizing a 9-bp DNA target inverted site is necessary because dimerization of the *Fok*I cleavage domain is required to produce a DSB (double stranded break). Therefore, three-finger ZFNs effectively have an 18-bp recognition site, which is long enough to specify a unique genomic address in plants and mammals including humans. ZFNs thus offer a general mechanism to introduce a site-specific DSB within a plant or a mammalian genome (Berg 1993).

For ZFN-mediated gene targeting to become an efficient and powerful genome engineering tool for specific proven biological and biomedical applications, rapid

design and generation of ZFPs that determine the sequence specificity of the ZFNs is essential. The designed ZFPs appear to have the highest affinity and sequence-specificity for their targets only when the individual ZF designs are chosen in the context of their neighbouring fingers. The computational work in this thesis focuses on understanding DNA-binding specificity in zinc finger proteins (i) through analysis of the physicochemical nature behind the ZFP-DNA interactions (ii) by investigating aspects like desolvation and DNA deformation based on simulations and free energy profile data that revealed a consensus in correlating affinity and specificity as well as stability of ZFP-DNA interactions, and (iii) by development of novel prediction algorithms that were evaluated for their performance against experimental data.

CONTENTS

Title **Page No.**

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ZFP	zinc finger proteins
ZFN	zinc finger nucleases
ns	nanosecond
μNN	micro neural network
IHBE	interfacial hydrogen bond energy
FEP	free energy of perturbation
MD	molecular dynamic simulations
CoDA	context-dependent assembly
OPEN	oligomerized pool engineering
SVM	support vector machines
RMSD	root mean square deviation
ANN	artificial neural network
BN	Bayesian network