



FACULTY OF GRADUATE STUDIES

**Using Molecular Modeling to Identify Zinc Finger
Protein-DNA Recognition Patterns: Energy
Perspective on The Effect of Finger Linkers on
Binding**

استخدام التمثيل الجزيئي لتحديد أنماط تعرف بروتين الزنك على
الحمض النووي: تأثير سلسلة الأحماض الأمينية التي تربط أجزاء
بروتين الزنك ببعضها البعض من منظور الطاقة

This Thesis was submitted in partial fulfillment of the requirements
for the

Master's Degree in Applied Chemistry

From the Faculty of Graduate Studies at Birzeit University, Palestine

By: Roza Zaid

Supervisor: Dr. Mazen Hamed

Using Molecular Modeling to Identify Zinc Finger Protein-DNA Recognition Patterns: Energy Perspective on The Effect of Finger Linkers on Binding

By
Roza H. Zaid

This Thesis was defended successfully on the 9th of September, 2014 and approved by:

Committee Members

Signature

Dr. Mazen Hamed
Department of Chemistry, Birzeit University
Supervisor

.....

Dr. Wael Karain
Department of Physics, Birzeit University
Member of thesis Committee

.....

Dr. Salih Al-Jabour
Department of Chemistry, Al-quds University
Member of thesis Committee

.....

ACKNOWLEDGEMENTS

First and foremost, I present my gratefulness to my supervisor Dr. Mazen Hamed for his ongoing guidance, generous support & continuous motivation all the way through the project along with his knowledge which he was generous to share with me.

I would also like to express my sincere thanks to the members of the thesis committee, Dr. Wael Karain & Dr. Salih Al-Jabour for their time and effort.

Lastly but not least, my special thanks & gratitude are expressed to my family and my beloved husband.

To the souls of Palestine's martyrs.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	I
DEDICATION	II
<u>TABLE OF CONTENTS</u>	III
ABSTRACT	VI
LIST OF FIGURES	VIII
LIST OF TABLES	XIX
ملخص	XX
1. INTRODUCTION	1
1.1 DNA binding proteins	1
1.1.1 Zinc finger proteins	2
1.2 Zinc finger protein zif268	3
1.2.1 Zinc finger-DNA interactions and specificity	5
1.3 Linkers in zinc finger protein interaction with DNA	6
1.3.1 Structure of Zinc finger protein linkers	8
1.4 Applications of Zinc finger proteins in gene regulation	10
1.5 Computational approach for finding free energy of binding	13
1.5.1 Molecular dynamic software used in this work	15

1.5.2	MM-PBSA and MM-GBSA method for calculating free energy of binding of protein-DNA complex.....	17
2.	Computational methods	24
2.1	Protein data bank structure.....	24
2.2	Creating prmtop and inpcrd files.....	26
2.3	Equilibration of the solvated system.....	3.3
2.4	Production MD simulation of the solvated system.....	35
2.5	Calculating The free energy of binding for the protein-DNA complex....	36
2.6	Calculating the entropy contribution.....	37
2.7	Analyzing the results	37
2.8	Preparing mutants of Zif268.....	38.
3.	Results and Discussion.....	40
3.1	Zif268 point mutants.....	40
3.2	Analysis of simulation output files.....	41
3.3	Total Binding free energy (GBTOT) for Zif268-DNA complex compared to its ten point mutants bound to the same DNA binding site.....	48
3.3.1	Analysis of the binding free energy of Zif268 -DNA complex.....	54
3.3.2	Analysis of the binding free energy of T28A -DNA complex.....	54
3.3.3	Analysis of the binding free energy of G29P-DNA complex.....	63

3.3.4	Analysis of the binding free energy of Q30E-DNA complex.....	69
3.3.5	Analysis of the binding free energy of K31D-DNA complex.....	75
3.3.6	Analysis of the binding free energy of P32G-DNA complex.....	82
3.3.7	Analysis of the binding free energy of T56Y-DNA complex.....	87
3.3.8	Analysis of the binding free energy of G57V-DNA complex.....	94
3.3.9	Analysis of the binding free energy of E58Q-DNA complex.....	99
3.3.10	Analysis of the binding free energy of K59P-DNA complex.....	104
3.3.11	Analysis of the binding free energy of P60A-DNA complex.....	115
3.3	Mutagenesis studies of Zif268 -DNA complex.....	115
4.	Conclusion.....	122
APPENDICES		
	Appendix A: Zinc finger protein files.....	125
	Appendix B: Input files for simulation	129
	References.....	136

ABSTRACT

In this study, we investigated the effect of each amino acid in the conserved linker “TGEKP” and one of its variants “TGQKP”, on the binding free energy of Zif268 to its optimal DNA binding site (5' A GCG TGG GCG T 3'). Ten point mutants of Zif268 were created. The free binding energy for each mutant with the optimal binding site was estimated using Molecular Mechanics Generalized Born Surface Area Method (MM/GBSA).

Compared to the wild type protein Zif268, one mutant (T56Y) resulted in lower binding energy by 20.74 kcal/mol; Three mutants (Q30E, E58Q and P60A) produced considerably higher binding energy (by 25.5, 18.2, 27.6 kcal/mol, respectively); Six mutants (T28A, G29P, K31D, P32G, and G57V, and K59P) produced binding energy values within the standard deviation from the binding energy of the wild type, where T28A showed an increase in the free binding energy by 14.38 kcal/mol, whereas G29P, K31D, P32G, G57V, and K59P showed a decrease in the free binding energy by 13.44, 8, 11.42, 9.4, 15.83 kcal/mol, respectively. The free binding energy values were decomposed into their three major

components: electrostatic energy, van der Waals, and the electrostatic contribution to the solvation free energy. The binding free energy values for the ten mutants had the highest correlation with the total electrostatic energy (the sum of electrostatic energy as calculated by molecular mechanics and electrostatic contributions to the solvation free energy).

Having only one out of ten with considerably lower binding energy suggests that the main reason for the high conservation of these linkers could be due to their role in biological processes other than specific binding to the DNA. Hydrogen bond analysis revealed that each mutant affected the stability and bond lengths. The effect on hydrogen bond stability was not confined to the vicinity of the mutated amino acid, and was detected throughout the zinc finger protein.

LIST OF FIGURES

Figure 1.1 The structure of a zinc finger domain belonging to Cys₂His₂ family.....	4
Figure1.2 Conserved Hydrogen bonds in Zif268-DNA complex	6
Figure 1.3 The Krüppel-type linkers in Zif268 are too short to make zinc finger pass over one of the base pairs of DNA recognition site.....	7
Figure 1.4 Schematic representation of two strategies used to design an engineered zinc finger protein: a) Two zinc finger protein consisting of three fingers each, are linked using long linker b) Three zinc finger proteins, consisting of two fingers each, are connected using an extended Krüppel–type linker.....	11
Figure 1.5 Schematic representation of zinc finger nuclease pair.....	13
Figure 1.6 Comparison between different computational methods in terms of their accuracy and speed.Error! Bookmark not defined.	
Figure 1.7 Calculating binding free energy for a solvated system.....	18
Figure 1.8 Thermodynamic cycle demonstrating steps of calculating binding free energy using MM/GBSA method.....Error! Bookmark not defined.	

Figure 1.9 The atom as represented by Generalised Born model.....	Error! Bookmark not defined.
Figure 2.1 Loading of configuration files needed for AMBER FF99SB force field.....	27
Figure 2.2 Preparing xleap to load the ZIF268-DNA complex X-ray structure.....	28
Figure 2.3 The editor window of xleap showing the graphical representation of 1AAY-dry.pdb	
.....	Error!
Bookmark not defined.	
Figure 2.4 Neutralization of the Zif268-DNA complex by addition of sodium ions.....	31
Figure 2.5 A graphical representation for Zif268-DNA solvated system.....	32
Figure 2.6 The PyMOL viewer window showing how an amino acid can be mutated.....	39
Figure 3.1 Amino acid sequences of the two canonical linkers in Zif268 and ten of its point mutants.....	41
Figure 3.2 Density of Zif268-DNA system during equilibration runs.....	42
Figure 3.3 Temperature of Zif268-DNA system during equilibration runs.....	43
Figure 3.4 Pressure of Zif268-DNA system during equilibration runs	44
Figure 3.5 Total, kinetic and potential energy of Zif268-DNA during equilibration runs.....	45
Figure 3.6 RMSD of DNA backbone during the unrestrained equilibration run of Zif268-	

DNA complex.....	47
Figure 3.7 RMSD of protein backbone during the unrestrained equilibration run of Zif268-DNA complex.....	47
Figure 3.8 Total binding free energy for Zif268-DNA complex and for ten of its point mutants. a) The Binding free energy values with their accompanying standard deviation. b) The change in binding free energy upon point mutations.....	50
Figure 3.9 The major contributors to the binding free energy of Zif268 and its mutants to Zif268 optimal binding site.....	51
Figure 3.10 Correlation between total binding free energy and total electrostatic energy...	52
Figure 3.11 Correlation between total binding energy and the sum of van der Waals and non-polar contribution to solvation, upon point mutations in the linkers.....	53
Figure 3.12 Contributions of total electrostatic energy and van der Waals energies to the total binding energy of Zif268 and ten of its point mutants.....	53
Figure 3.13. Mutating the first linker residue threonine ²⁸ to Alanine.....	55
Figure 3.14 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and T28A.....	57
Figure 3.15 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and T28A.....	58
Figure 3.16 Percent occupancy of hydrogen bonds in the second linker over production	

runs of 1AA Y and T28A.....	59
Figure 3.17 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and T28A.....	60
Figure 3.18 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and T28A.....	61
Figure 3.19 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and T28A.....	63
Figure 3.20 Mutating the second linker residue (Gly29) to Proline.....	64
Figure 3.21 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and G29P.....	66
Figure 3.22 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and G29P.....	66
Figure 3.23 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and G29P.....	67
Figure 3.24 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and G29P.....	67
Figure 3.25 Percent occupancy of the crucial hydrogen bonds between, the conserved	

amino acids in the α-helices of the protein and bases in DNA major, over production runs of 1AAY and G29P.....	68
Figure 3.26 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AAY and G29P.....	68
Figure 3.27 Mutating the third amino acid in the first linker (Gln30) to glutamate.....	69
Figure 3.28 The third and fourth amino acids of the first linker in a) Zif268, b) Q30E.....	71
Figure 3.29 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AAY and Q30E.....	72
Figure 3.30 Average lengths of hydrogen bonds in the first linker over production runs of 1AAY and Q30E.....	72
Figure 3.31 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AAY and Q30E.....	73
Figure 3.32 Average lengths of hydrogen bonds in the second linker over production runs of 1AAY and Q30E.....	73
Figure 3.33 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AAY and Q30E.....	74

Figure 3.34 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and Q30E.....	74
Figure 3.35 Two water-mediated contacts between Lysine33 and the DNA base T5.....	75
Figure 3.36 Mutating Lysine31 in the first linker to Aspartate.....	75
Figure 3.37 The distances in Å between the oxygen in Aspartate and the two oxygen atoms in the phosphate group.....	77
Figure 3.38 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and K31D.....	79
Figure 3.39 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and K31D.....	79
Figure 3.40 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and K31D.....	80
Figure 3.41 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and K31D.....	80
Figure 3.42 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and K31D.....	81
Figure 3.43 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs	81

of 1AA Y and K31D.....	81
Figure 3.44 Mutating Proline32 in the first linker to Glycine.....	82
Figure 3.45 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and P32G.....	84
Figure 3.46 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and P32G.....	84
Figure 3.47 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and P32G.....	85
Figure 3.48 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and P32G.....	85
Figure 3.49 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major, over production runs of 1AA Y and P32G.....	86
Figure 3.50 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and P32G.....	86
Figure 3.51 Mutating Threonine56 in the second linker to Tyrosine.....	87
Figure 3.52 Hydrogen bond between Thr56 and Glu58 in 1AA Y.....	89
Figure 3.53 The distance between Tyr56 and Glu58 in T56Y mutant.....	90
Figure 3.54 Percent occupancy of hydrogen bonds in the first linker over production runs	

of 1AA Y and T56Y.....	91
Figure 3.55 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and T56Y.....	91
Figure 3.56 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and T56Y.....	92
Figure 3.57 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and T56Y.....	92
Figure 3.58 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and T56Y.....	93
Figure 3.59 Average Lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AA Y and T56Y.....	93
Figure 3.60 Mutating Glycine⁵⁷ in the second linker to Valine.....	94
Figure 3.61 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and G57V.....	96
Figure 3.62 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and G57V.....	96
Figure 3.63 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and G57V.....	97

Figure 3.64 Average lengths of hydrogen bonds in the second linker over production runs of 1AAY and G57V	97
Figure 3.65 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and G57V.....	98
Figure 3.66 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and G57V.....	98
Figure 3.67 Mutating Glutamate58 in the second linker to Glutamine.....	99
Figure 3.68 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AAY and E58Q	101
Figure 3.69 Average lengths of hydrogen bonds in the first linker over production runs of 1AAY and E58Q	101
Figure 3.70 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AAY and E58Q	102
Figure 3.71 Average lengths of hydrogen bonds in the second linker over production runs of 1AAY and E58Q	102
Figure 3.72 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and E58Q.....	103

Figure 3.73 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and E58Q.....	103
Figure 3.74 Mutating lysine ⁵⁹ in the second linker to Proline.....	104
Figure 3.75 Percent occupancy of hydrogen bonds in the first linker over production runs of 1AA Y and K59P.....	106
Figure 3.76 Average lengths of hydrogen bonds in the first linker over production runs of 1AA Y and K59P.....	107
Figure 3.77 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AA Y and K59P.....	107
Figure 3.78 Average lengths of hydrogen bonds in the second linker over production runs of 1AA Y and K59P.....	108
Figure 3.79 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and K59P.....	108
Figure 3.80 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major, over production runs of 1AA Y and K59P.	109
Figure 3.81 Mutating Proline ⁶⁰ in the second linker to Alanine.....	109
Figure 3.82 Percent occupancy of hydrogen bonds in the first linker over production	

runs of 1AAY and P60A.....	112
Figure 3.83 Average lengths of hydrogen bonds in the first linker over production runs of 1AAY and P60A.....	112
Figure 3.84 Percent occupancy of hydrogen bonds in the second linker over production runs of 1AAY and P60A.....	113
Figure 3.85 Average lengths of hydrogen bonds in the second linker over production runs of 1AAY and P60A.....	113
Figure 3.86 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AAY and P60A.....	114
Figure 3.87 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α-helices of the protein and bases in DNA major groove, over production runs of 1AAY and P60A.....	114
Figure 3.88 The experimental binding phenotype of each mutant to one of Zif268 characterized binding site (5-GCG GGG GCG-3).....	119

LIST OF TABLES

Table 1.1: Total binding free energy and the its components, for Zif268-DNA complex and for ten of its point mutants to Zif268 optimal binding site.

ملخص

في هذه الدراسة قمنا باختبار تأثير كل حمض نووي في سلسلتي الأحماض الأمينية اللتين تربطان بين أجزاء بروتين الزنك (TGEKP) و (TGQKP) على طاقة ربط بروتين الزنك بالموقع المحدد له على الحمض الريبوزي النووي المنزوع الأوكسجين. تم تكوين عشرة طفرات من بروتين Zif268 ومن ثم حساب طاقة الربط لكل طفرة على حده مع نفس الموقع على الحمض الريبوزي النووي المنزوع الأوكسجين باستخدام طريقة (Molecular Mechanics Generalized Born Surface Area).

مقارنة مع البروتين (Zif268) طفرة واحدة (T56Y) أنتجت طاقة ربط أقل ب (20.7 kcal/mol) ، ثلاث طفرات (Q30E, E58Q and P60A) أنتجت طاقة ربط اعلى ب (25.5, 18.2, 27.6 kcal/mol على الترتيب)، ستة طفرات (T28A, G29P, K31D, P32G, G57V, K59P) أنتجت طاقة ربط قريبة من طاقة ربط بروتين (Zif268) مع أخذ قيمة الانحراف المعياري بعين الاعتبار.

تم تفكيك قيم طاقة ربط البروتينات مع الحمض الريبوزي النووي المنزوع الأوكسجين الى ثلاثة من مكوناتها الأساسية : الطاقة الإلكترونية، طاقة فان دير فالس، و الإسهامات الأكتروستاتيكية لطاقة الإذابة. قيم طاقة الربط للعشر طفرات كان لها أقوى ارتباط مع مجموع الطاقة الإلكترونية.

وجود طفرة واحدة فقط من أصل عشرة لها طاقة ربط أقل ، يشير إلى أن السبب الرئيسي لحفظ تسلسل الأحماض الأمينية (TGEKP) قد يعود لأهمية هذا التسلسل في وظائف حيوية أخرى عدى عن ارتباط بروتين الزنك بموقعه المحدد على الحمض الريبوزي النووي المنزوع الأوكسجين.

Chapter 1

INTRODUCTION

1.1 DNA-binding proteins

DNA binding proteins perform many of the precise processes that take place on DNA, such as DNA replication, DNA repair, and gene expression control.¹ These proteins consist of several small domains and interact with DNA either in a sequences-specific or non-specific mode.² Some DNA-binding proteins interact with the DNA minor groove but the majority of these proteins bind DNA in the major groove, where the bases in this groove are more accessible to the protruding surfaces of these proteins.^{3,4}

There is a variety of different structures of DNA binding proteins including helix-turn-helix (HTH), leucine zipper and zinc finger proteins, this structural diversity is reflected as functional variety of different DNA-binding proteins.^{1,4}

1.1.1 Zinc Finger Proteins

About 30 years ago the first zinc finger protein was discovered in the African clawed toad *Xenopus* transcription factor (TFIIIA).⁵ This protein was found to consist of nine consecutive zinc finger domains. Each domain had a sequence of the form $X_3\text{-Cys-}X_{2-4}\text{-Cys-}X_{12}\text{-His-}X_{3-4}\text{-His-}X_4$ (where X is any amino acid).^{5,6} Since then, many other zinc finger proteins were discovered in various species, such as ADRI in yeast, GLI in humans and Zif268 in mice.⁷

The general role of zinc finger proteins can be described as “interaction modules”, since they carry out their functions through binding to various targets including DNA, RNA, as well as proteins and small molecules.⁸ It is fair to say that DNA replication and repair, transcription, translation, cell proliferation and apoptosis in addition to other vital cellular processes depend heavily on the role played by zinc finger proteins.^{8,9}

Zinc finger proteins are usually composed of one or more zinc finger domains. Each zinc finger domain contains one or more zinc ions to establish its folding pattern.¹⁰ However, zinc finger proteins were classified into different structural classes depending on how the protein folds around the zinc ion. These

structural classes include: Cys₂His₂ fingers, treble clef protein fingers, zinc ribbon and Zn₂/Cys₆ fingers. Among these classes Cys₂His₂ fingers are the most abundant and recognizable class.^{8,11,12} As the name Cys₂His₂ implies, these proteins bind the zinc ion by 4 conserved residues, 2 cysteins and two histidines. Their structure takes the shape of a finger which consists of an α -helix and two anti parallel β -sheets.¹⁰

1.2 The zinc finger protein Zif268.

Zif268 has been considered as a suitable model system for the investigation of zinc finger-DNA interactions, for two reasons. First, many studies have been conducted on Zif268-DNA complexes and succeeded to give a good picture of how Cys₂His₂ zinc finger proteins recognize their DNA targets¹³⁻¹⁹. Second, the availability of Zif268-DNA structure (with 1.6Å resolution) constitutes an adequate framework for carrying a molecular modeling study²⁰. These studies elucidated the main contacts between the zinc finger protein and the DNA bases in the major groove. The effect of mutating an amino acid in the linker region of Zif268 on the stability of the main hydrogen bonds will be investigated in this study.

Zif268 consists of three zinc finger domains. Each finger has about 30 amino acids taking the shape of an α -helix and two anti parallel β -sheets²⁰. This whole functional finger motif maintains its conformation through its coordination to a zinc ion, and its hydrophobic core composed of hydrophobic amino acids which hide the zinc ion from water²¹ (Figure 1.1).

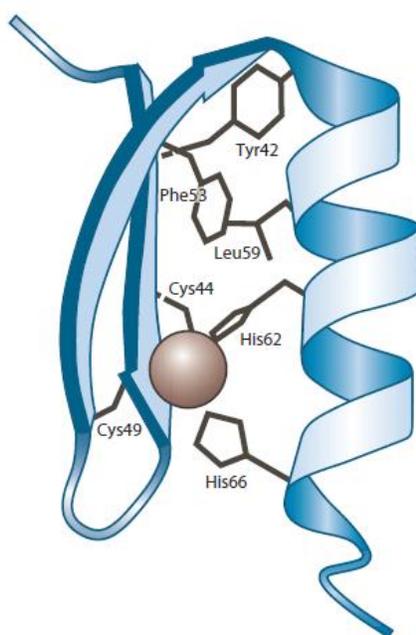


Figure 1.1 The structure of a single zinc finger domain belonging to Cys_2His_2 family. Two Histidine residues (from α -helix) and two cysteine residues (from β sheets) protruding toward the zinc ion (represented as a sphere) are shown. Three hydrophobic residues are displayed²².

1.2.1 Zif268 -DNA interactions and specificity

The identity of the side chains that project out from α -helices gives each finger certain specificity by which it can bind to certain bases on DNA. Known trends in specificity concluded from experimental mutagenesis and binding studies are : Arg (on position -1 with respect to the starting point of α helix) in fingers 1,2, and 3 interact with G bases on the primary strand of DNA. His (on position 3) in finger 2 interacts with G base on the primary strand of DNA. Glu (on position 3) in fingers 1 and 3, interact with C bases on primary strand of DNA, Arg (on position 6) in fingers 1 and 3, interact with G bases on the primary strand of DNA. Asp (on position 2) in fingers 1 and 3 interact with A bases on the secondary strand of DNA. Asp (on position 2) in finger 2 interacts with C base on secondary strand of DNA^{14,23-26} (Figure 1.2).

connect finger 1 with finger 2, and TGEKP which connects finger 2 with finger 3^{9,28,29}.

In Zif268-DNA complex the length of the linkers TGQKP and TGEKP was found to be 14.5 Å and 14.4 Å, respectively. This means that they are too short to make zinc finger motifs skip a base pair in the DNA recognition motif (Figure 1.3). Due to this shortness, in addition to the slight difference in helical periodicity of the DNA and that of the α -helix in the protein, DNA has to change its conformation and become slightly unwound in order to allow a better fit of the protein α -helix to the DNA major groove³⁰.

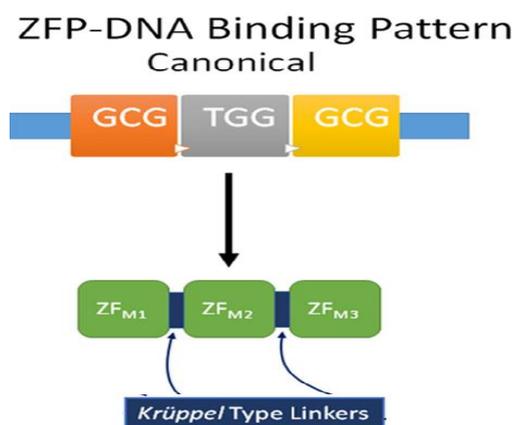


Figure 1.3 The *Krüppel*- type linkers in Zif268 are too short to make zinc finger pass over one of the base pairs of DNA recognition site³¹.

1.3.1 Role of Zinc finger protein linkers

The high conservation of the linker sequence in most of the zinc finger linkers encoded by the human genome caused a lot of curiosity about its functional role. Some early structural and mutagenesis studies tried to remove the ambiguity surrounding the role of the linkers^{28,32}.

One of the first important observation in this regard was that the consensus linkers in the first three linkers in TFIIIA zinc finger protein are dynamically disordered in solution, where the zinc finger protein is free, but exhibit a well ordered structure upon binding to DNA^{33,34}.

An experimental study was conducted on the first three fingers of TFIIIA zinc finger protein to investigate the effect of the linker sequence TGEKP between the first two fingers and the linker sequence TGEKN between the second and the third fingers²⁸. Exchanging these linkers with other linkers obtained from different proteins led to loss of binding. Single mutations in these consensus linkers led to a decrease in binding affinity of up to 24-fold in the case of G39P mutant, in which glycine in the first linker was mutated to proline. However, results of this study could not account for the reason behind the high conservation of *Krüppel* -type linkers.²⁸ In another experimental study on TFIIA the replacement of three amino

acids from the *Krüppel*-type linker with 3 amino acids of an equivalent position from another linker resulted in a decrease in its DNA binding affinity up to 8-fold³².

NMR studies on the first three fingers of TFIIIA and the first four fingers of Wilm's tumor suppressor protein confirmed the formation of α -helix capping up on binding to DNA^{35,36}. This α -helix capping is caused by hydrogen bonds between amino acids in the linker (Glycine and threonine) and amino acids from the end of the α -helix. The α -helix capping was suggested to be a stabilizing factor for Cys₂His₂ zinc finger-DNA complexes by providing an important enthalpic contribution to the binding process. Crystal structures of the Zif268-DNA complex showed a hydrogen bond between carbonyl oxygen of the third residue from the end of the helix (Arg) and the backbone amide of the second amino acid in the linker (Gly), also a hydrogen bond between the hydroxyl oxygen of Threonine (the first amino acid in the linker) and the backbone amide of the linker's third amino acid (Glutamate) was observed^{6,36}. These interactions are induced by binding to DNA and makes the α -helix become more helical. This led to an assumption that the intrinsic flexibility of the linkers is important while the zinc finger is diffusing along the DNA searching for its cognate site. After finding the binding site these linkers will be induced to "snap locks like" structures that will fix the zinc finger in

the suitable orientation to bind DNA. As a result C-capping was suggested to give a structural role of the TGEKP linker and some of its variants and account for the possible justification of the high conservation of *Krüppel* -type linkers³⁶.

1.4 Applications of Zinc finger proteins in gene regulation.

One of the characteristics that distinguish zinc finger proteins from other DNA-binding proteins is that they can be connected consecutively to recognize different lengths of specific sites on DNA. In view of this fact many engineered zinc finger proteins have been created so far by fusing various zinc fingers together²³.

At the beginning zinc fingers were fused using the conventional linkers but these short linkers weren't the best choice to produce a poly-zinc finger protein with high affinity. This is due to the increase in the mismatching between the helical periodicity of the protein and DNA, which requires an additional unwinding in DNA^{37,38}. As a result, longer linkers were used. In the first example of utilizing longer linkers to design engineered zinc finger proteins with high specificity and affinity, Zif268 was linked to another zinc finger protein namely, NER which consists of three zinc fingers³⁰ as shown in Figure 1.4. a. Another method to make

a six zinc finger protein using a variant of the consensus linker TGEKP, in which three additional amino acids were inserted, to fuse three proteins, with each protein consisting of two zinc fingers (Figure 1.4. b). The variety of possible combinations of linkers and fingers makes the design of specific artificial zinc finger proteins for different targeted genes more feasible²².

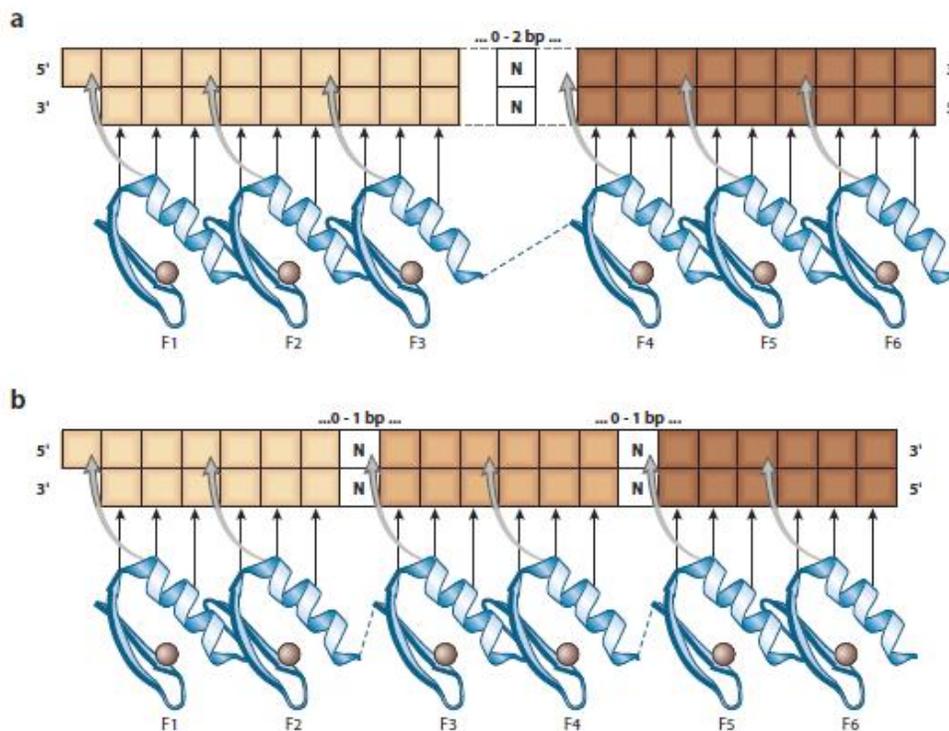


Figure 1.4 Schematic representation of two strategies used to design an engineered zinc finger protein: a) Two zinc finger protein consisting of three fingers each, are linked using long linker b) Three zinc finger proteins, consisting of two fingers each, are connected using an extended *Krüppel*-type linker²².

A further utilization was achieved by fusing engineered zinc finger proteins with transcription activators or repressors, thus producing highly specific transcription factors. Taking it to the next level, zinc finger proteins were connected to DNA cleavage domain, producing zinc finger nuclease which is considered to be an efficient tool, used to repair defective genes. In principle, zinc finger nucleases reach the defective gene via the engineered zinc finger proteins which then binds to a locus adjacent to the defective gene. This directs cleavage domain to the site where it should introduce the break in DNA. This process takes place on the other strand of DNA, but in an opposite direction by another zinc finger nuclease, eventually producing a DNA double-strand break which would then be mended via a natural process called homology-directed repair^{22,39} (Figure 1.5).

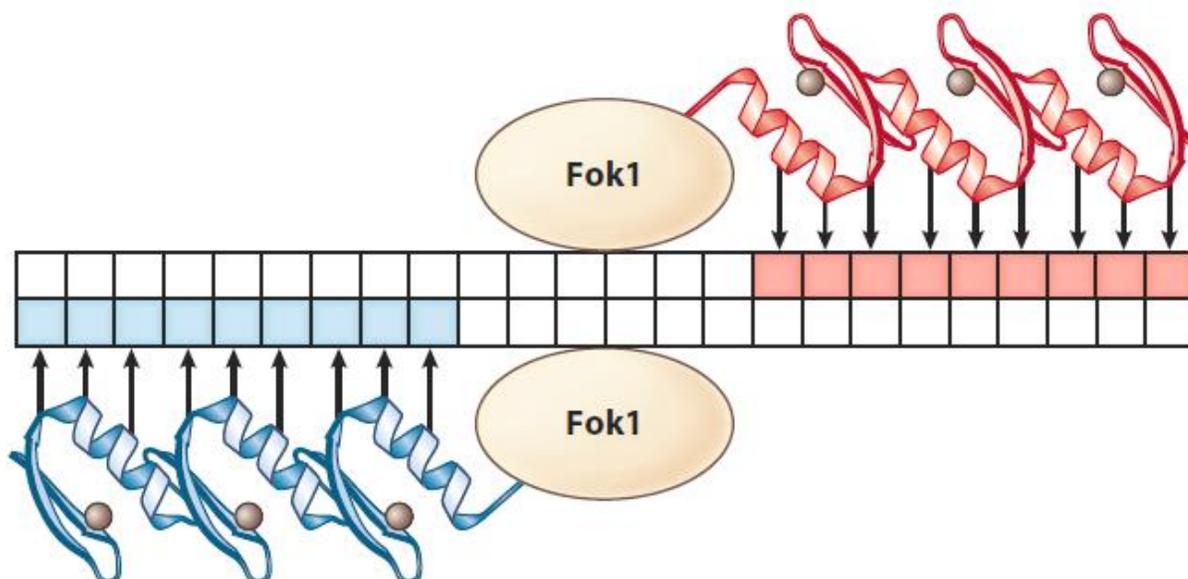


Figure 1.5 Schematic representation of zinc finger nuclease pair. Each one consists of an engineered zinc finger protein linked to a nonspecific cleavage domain of the Fok1 type II restriction enzyme²².

1.5 Computational approach for finding free energy of binding.

Computational methods have emerged as very effective tools for predicting free energy of binding (ΔG)⁴⁰⁻⁴⁷. These methods facilitated the understanding of the factors accounting for binding affinity in protein-protein interactions, protein-DNA interactions and drug actions.⁴⁸ Several computational methods are available to perform this task. Some are rigorous but computationally expensive such as, the

thermodynamic integration (TI) and the free energy perturbation (FEP) methods⁴⁹⁻⁵². Others are less accurate but more computationally efficient methods. These include, molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) and molecular mechanics/ Generalized Born surface area (MM/GBSA) methods⁵³ (Figure 1.6).

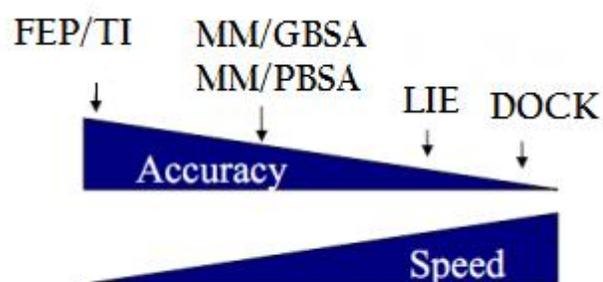


Figure 1.6 Comparison between different computational methods in terms of their accuracy and speed.

Both (TI) and (FEP) deal with the solvent explicitly and calculate the binding free energy of Protein-DNA complex depending on simulations taken at intermediate stages between the two end states (unbound and bound states of a protein and its target DNA). These principles adopted by (TI) and (FEP) explain

the high accuracy and time consuming characteristics of the two rigorous methods.⁵⁴

The simpler (MM/PBSA) and (MM/GBSA) are end point methods (i.e. they predict the binding free energy depending on simulations of only two end states, the unbound and the bound states of a protein and its target DNA). These methods treat the solvent implicitly and for that they are called continuum solvent methods⁵⁴.

1.5.1 Molecular dynamic software used in this work.

Through performing molecular dynamics one is actually trying to mimic experimental procedure for reading a certain property over certain time interval. In the experiment, one first prepares the sample to be studied, and then the sample is to be connected to a device that takes the measurement of the property over particular time duration. Any statistical noise that may accompany the measurement can be eliminated by longer averaging time⁵⁵.

In an analogous way, when using molecular dynamics a model system composed of N particles is chosen. Newton's motion equation are integrated and

solved for the system until the system equilibrates, and the properties of the system reach a steady state. at this point the desired measurement can be taken⁵⁵.

Various software suites have been used for running (MD) simulations. One of the most used packages is AMBER^{56,57}. This acronym stands for assisted model building and energy refinement. The use of AMBER to perform MD simulations is not limited to giving the equilibrium property but also the transport property of the system.⁵⁸ Using AMBER to carry out molecular dynamic simulation and to calculate binding free energy is an important approach that complements experimental work since in some cases binding affinities calculated through experimental procedure were reported with 40-50% error²⁸.

AMBER consists of many programs that collaborate to carry out the molecular dynamic simulations. Examples of these programs are: LEap, Sander, Ptraj, Nmode. These programs are created to help in various tasks^{56,59}:

(i) Leap is designed to help in setting up and adjusting structures of biomolecules and to prepare input files that can be used by other molecular dynamics programs within AMBER suit .

(ii) Sander is designed for conducting minimization and other molecular dynamic steps.

(iii) Ptraj can be used to analyze the output files.

(iv) Nmode can be used to carry out normal mode calculations.

(v) MM-GBSA: enables calculating the binding free energy of bio-molecule complexes.

AMBER also stands for the set of force fields required for performing molecular dynamic simulations of biomolecules. These force fields contain a large collection of parameters that define the constituents of biomolecules and several relevant solvents. This family of software programs, together with a diverse set of force fields, constitute a suitable environment to carry out MD simulations of biomolecules⁶⁰.

1.5.2 MM-GBSA method for calculating free energy of binding of protein-DNA complex.

The protein-DNA binding free energy can be calculated as the difference between the free energy of protein-DNA complex and the free energies of protein and DNA, in separate, as illustrated in the diagram and the equation (1.1) below⁵³.

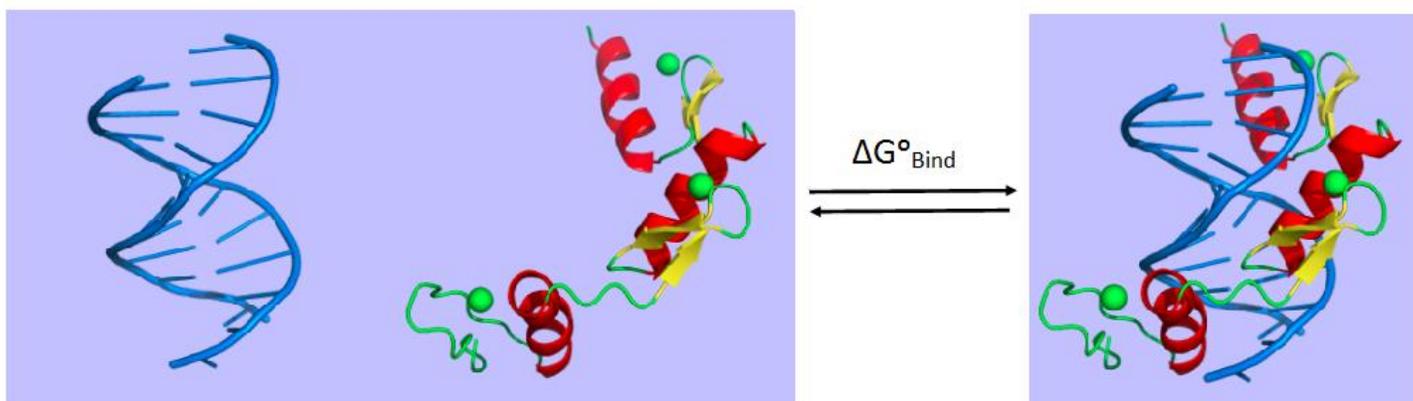


Figure 1.7 Calculating binding free energy for a solvated system as the difference in free energy between bound and unbound states. Light blue indicates solvent shell.

$$\Delta G^{\circ}_{\text{Bind}} = G^{\circ}_{\text{Protein-DNA complex}} - (G^{\circ}_{\text{protein}} + G^{\circ}_{\text{DNA}}) \quad (1.1)$$

Nonetheless, solvent-solvent interactions are the major contributors in the simulations of solvated systems, and the fluctuations in total energy is about ten times greater than the binding energy. This will result in taking excessive time before total energy converges to an acceptable tolerance. Therefore, a more efficient approach is to partition the calculation^{53,56,61}, as illustrated in the thermodynamic cycle below (Figure 1.8).

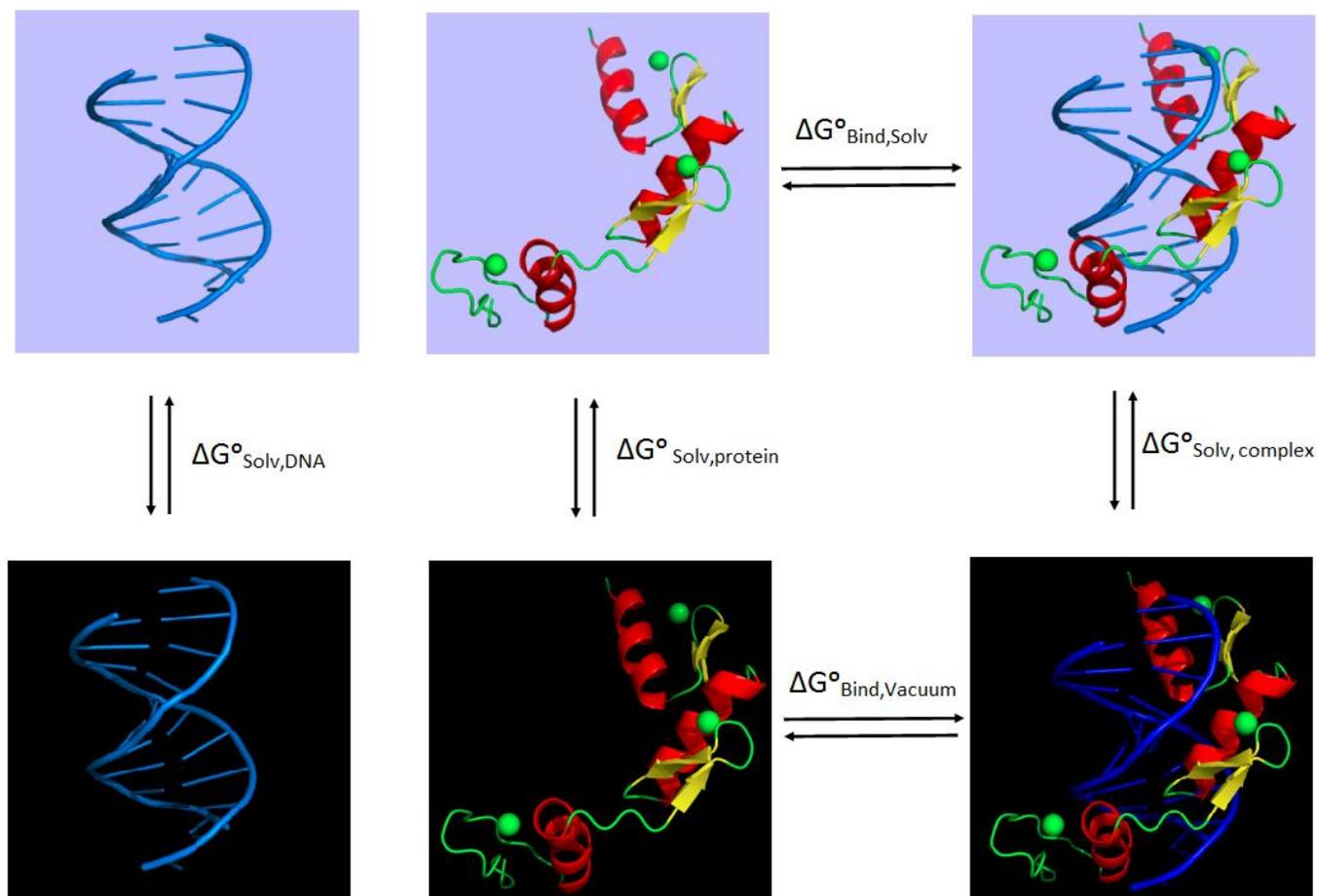


Figure 1.8 Thermodynamic cycle demonstrating steps of calculating binding free energy using MM/GBSA method. Light blue indicates solvent shell, while the black background indicates vacuum.

It is clear from the thermodynamic cycle that the binding free energy can now be calculated as in equation (1.2).

$$\Delta G^{\circ}_{\text{Bind, Solv}} = \Delta G^{\circ}_{\text{Bind, Vacuum}} + \Delta G^{\circ}_{\text{Solv, complex}} - (\Delta G^{\circ}_{\text{Solv, DNA}} + \Delta G^{\circ}_{\text{Solv, protein}}) \quad (1.2)$$

Computing solvation free energy ($\Delta G^{\circ}_{\text{solv}}$) requires encompassing of two components, a polar and non polar contribution as in equation (1.3).

$$\Delta G^{\circ}_{\text{solv}} = \Delta G^{\circ}_{\text{polar}} + \Delta G^{\circ}_{\text{non polar (Hydrophobic)}} \quad (1.3)$$

In order to gain a deep understanding of solvation free energy, one should first visualize how implicit solvent methods deal with the solvation process. Implicit solvation methods account for the following steps in solvating a charged solute: (i) a cavity is formed in the solution to host the solute while hypothetically assuming that this solute is hydrophobic by turning off all atomic charges (ii) the atomic charges on the solute are turned on again⁵³.

The first step in the solvation process accounts for the non polar contribution to solvation free energy which is assumed to be proportional to solvent accessible surface area, and given by the following equation (1.4)

$$\Delta G^{\circ}_{\text{solv (non polar)}} = \gamma (\text{SASA}) + \beta \quad (1.4)$$

Where γ and β values depend on which method and solvation model is applied.⁶²

The second step in the solvation process accounts for the polar component of solvation free energy. This polar component comes from the difference in free energy resulting from transporting the charged solute from gas phase with low dielectric constant ($\mathcal{E} = 1$) to a solvent which has a high dielectric constant ($\mathcal{E} = 80$), equation (1.6).

$$\Delta G^{\circ}_{\text{solv (polar)}} = G^{\circ}_{\text{electrostatic, } \epsilon=80} - G^{\circ}_{\text{electrostatic, } \epsilon=1} \quad (1.5)$$

The polar component can be calculated by solving the Generalized Born equation by estimating the work needed to turn on all atomic charges in the system giving equation (1.5)^{56,63}.

$$\Delta G_{el} \approx -\frac{1}{2} \left(1 - \frac{1}{\epsilon_w}\right) \frac{q^2}{\rho} \quad (1.6)$$

In the Generalized Born model, each atom is represented as a sphere with radius ρ_i and charge q_i . The inner part of the sphere is filled with a material with low dielectric constant ($\mathcal{E} = 1$), and the atom it exists in a solvent with high dielectric constant ($\mathcal{E} = 80$ for water at 300K).

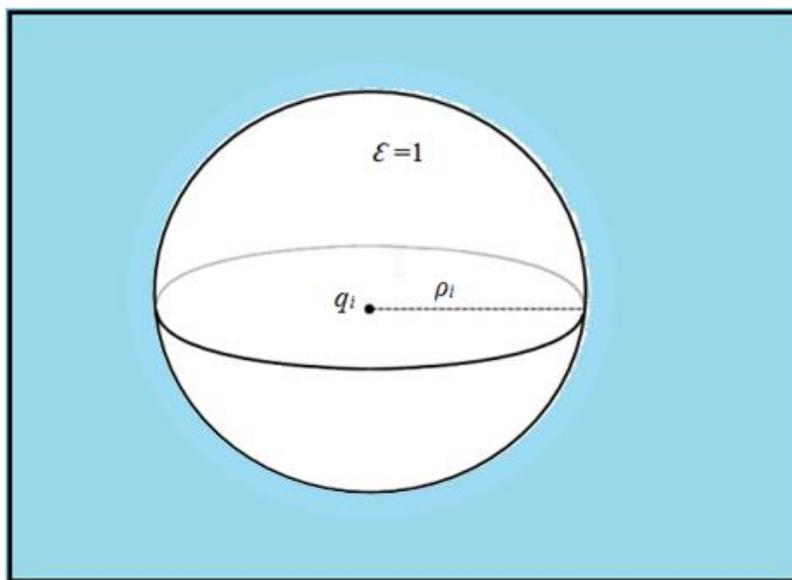


Figure 1.9 The atom as represented by Generalised Born model. Light blue indicates solvent with high dielectric constant.

Thus the solvation free energy ($\Delta G^{\circ}_{\text{solv}}$) can be given by the following overall equation (1.7).⁵⁴

$$\Delta G^{\circ}_{\text{solv}} = G^{\circ}_{\text{electrostatic, } \epsilon=80} - G^{\circ}_{\text{electrostatic, } \epsilon=1} + \Delta G^{\circ}_{\text{non polar}} \quad (1.7)$$

The binding free energy in vacuum ($\Delta G^{\circ}_{\text{Bind, Vacuum}}$) equals the summation of the average interaction energy between the protein and DNA in gas phase ($\Delta E^{\circ}_{\text{Molecular mechanics}}$) and the change in configurational entropy through binding, as shown in equation (1.8).

$$\Delta G^{\circ}_{\text{Bind, Vacuum}} = \Delta E^{\circ}_{\text{Molecular mechanics}} - T \cdot \Delta S^{\circ}_{\text{Normal mode analysis}} \quad (1.8)$$

The average interaction energy between the protein and DNA in gas phase can be estimated using molecular mechanics through applying force field functions and parameters. This term is composed of two types of energy. The first is the covalent energy represented by bonds, angles and dihedral energies. The second is non-covalent energies consisting of electrostatic and van der waals energies. These different energy contributions are summed up in equation (1.9) ⁵⁴.

$$E_{\text{MM}} = E_{\text{Bond}} + E_{\text{Angle}} + E_{\text{Torsion}} + E_{\text{van der Waals}} + E_{\text{Electrostatic}} \quad (1.8)$$

The Change in configurational entropy can be calculated using *nmode*. However, in many cases this term is ignored especially when calculating relative binding free energy, where the protein is binding to similar targets in each time. This is due to the fact that calculating entropy is a computationally demanding process that adds little information when ligands bound to the protein are similar^{53,54}.

Chapter 2

Computational Methods

2.1 Protein data bank structures

The Zif268 zinc finger-DNA complex X-ray structure (PDB code: 1AAY) was obtained from the RCSB Protein Data Bank. The names of some atoms and residues were changed to fit those in the amber topology files. The connectivity data at the end of the file were removed, and ‘TER’ cards were added between the different parts of the complex. Water molecules were removed from the PDB file.

Zif268 contains three zinc ions. Each zinc ion is coordinated to two cysteine and two histidine residues. Therefore each one of the coordinated residues should be in the deprotonated form (i.e. as negatively charged amino acids). The Zif268 file obtained from the protein data bank is not written in a way that enables xleap to discriminate bonded from non-bonded amino acids. Thus, these amino acids were manually modified to

enable them to bond the zinc ion.

Cystein residues in positions 5, 10, 35, 38, 63, and 66 were edited as follows:

- (i) The atom named HG was erased from each residue.
- (ii) The residue name for all atoms contained in that residue was converted from CYS to CYM, where CYM is the residue name used by *xleap* for deprotonated cystein.

Histidine residues number 23, 27, 51, 55, 79, and 83 were edited as the following:

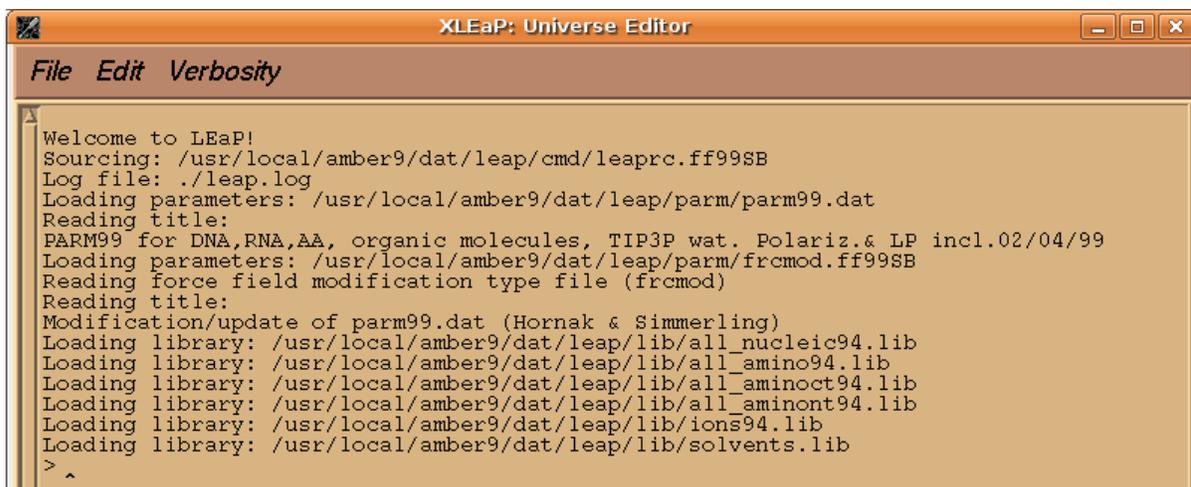
- (i) The atom named HE2 was erased from each residue.
- (ii) The residue names for all atoms contained in that residue was changed from HIE to HID, where HID is the residue name used by *xleap* for deprotonated Histidine in delta position. This edited PDB file was saved as 1AAY-dry.pdb

2.2 Creating prmtop and inpcrd files.

The Prmtop and inpcrd files are the molecular topology/parameter and coordinate files, respectively. These files are necessary for running molecular dynamics simulation using *Sander*. The xleap program was used to build these files. In order to open xleap the following command was typed in a terminal:

```
usr/local/amber9/exe$ xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB
```

This command not only starts xleap, but also loads the configuration files needed for AMBER FF99SB force field⁶⁴ as shown in Figure 2.1. However, the zinc parameters required for our complex are not included in the AMBER FF99SB force field.



```

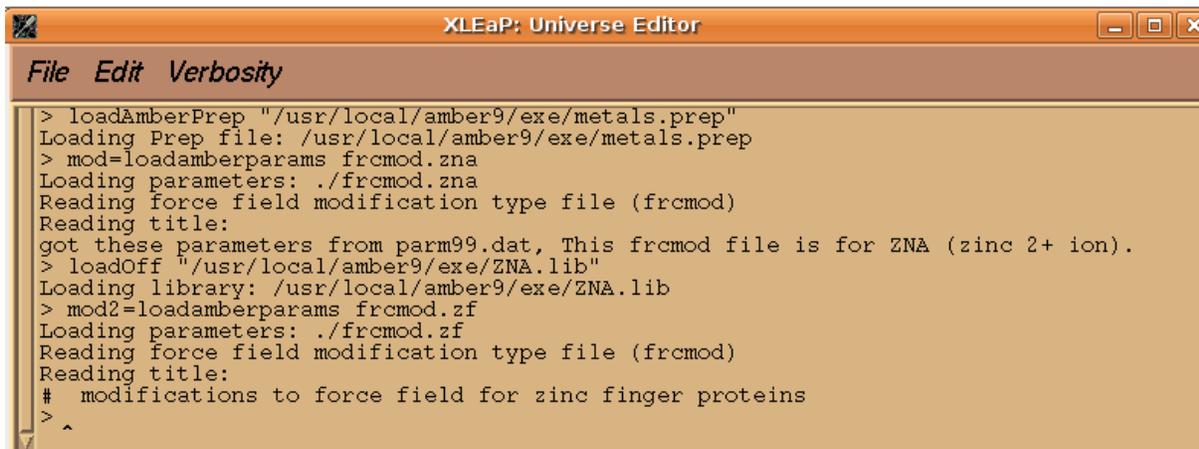
XLEaP: Universe Editor
File Edit Verbosity
Welcome to LEaP!
Sourcing: /usr/local/amber9/dat/leap/cmd/leaprc.ff99SB
Log file: ./leap.log
Loading parameters: /usr/local/amber9/dat/leap/parm/parm99.dat
Reading title:
PARM99 for DNA,RNA,AA, organic molecules, TIP3P wat. Polariz.& LP incl.02/04/99
Loading parameters: /usr/local/amber9/dat/leap/parm/frcmod.ff99SB
Reading force field modification type file (frcmod)
Reading title:
Modification/update of parm99.dat (Hornak & Simmerling)
Loading library: /usr/local/amber9/dat/leap/lib/all_nucleic94.lib
Loading library: /usr/local/amber9/dat/leap/lib/all_amino94.lib
Loading library: /usr/local/amber9/dat/leap/lib/all_aminoc94.lib
Loading library: /usr/local/amber9/dat/leap/lib/all_aminont94.lib
Loading library: /usr/local/amber9/dat/leap/lib/ions94.lib
Loading library: /usr/local/amber9/dat/leap/lib/solvents.lib
> ^

```

Figure 2.1 Loading of configuration files needed for AMBER FF99SB force field

In order to add the zinc missing parameters, the AMBER PREP input, PARMSET and OFF library files for the zinc ion are loaded before loading the Zif268 zinc finger-DNA complex X-ray structure in xleap, (see Appendix A).

The following command lines were typed in the main xleap window to load these files:



```

XLEaP: Universe Editor
File Edit Verbosity
> loadAmberPrep "/usr/local/amber9/exe/metals.prep"
Loading Prep file: /usr/local/amber9/exe/metals.prep
> mod=loadamberparams frcmod.zna
Loading parameters: ./frcmod.zna
Reading force field modification type file (frcmod)
Reading title:
got these parameters from parm99.dat, This frcmod file is for ZNA (zinc 2+ ion).
> loadOff "/usr/local/amber9/exe/ZNA.lib"
Loading library: /usr/local/amber9/exe/ZNA.lib
> mod2=loadamberparams frcmod.zf
Loading parameters: ./frcmod.zf
Reading force field modification type file (frcmod)
Reading title:
# modifications to force field for zinc finger proteins
> ^

```

Figure 2.2 Preparing xleap to load the ZIF268-DNA complex X-ray structure.

After these steps, xleap was ready to load the 1AAY-dry.pdb file without having problems in recognizing ZNA residue. 1AAY-dry.pdb file was loaded into xleap after setting it to a new unit called “a” and using the command loadPdb. The command edit was then used to look at the unit named “a”. As a result, the editor window of xleap appeared showing the graphical representation of 1AAY-dry.pdb (Figure 2.3).

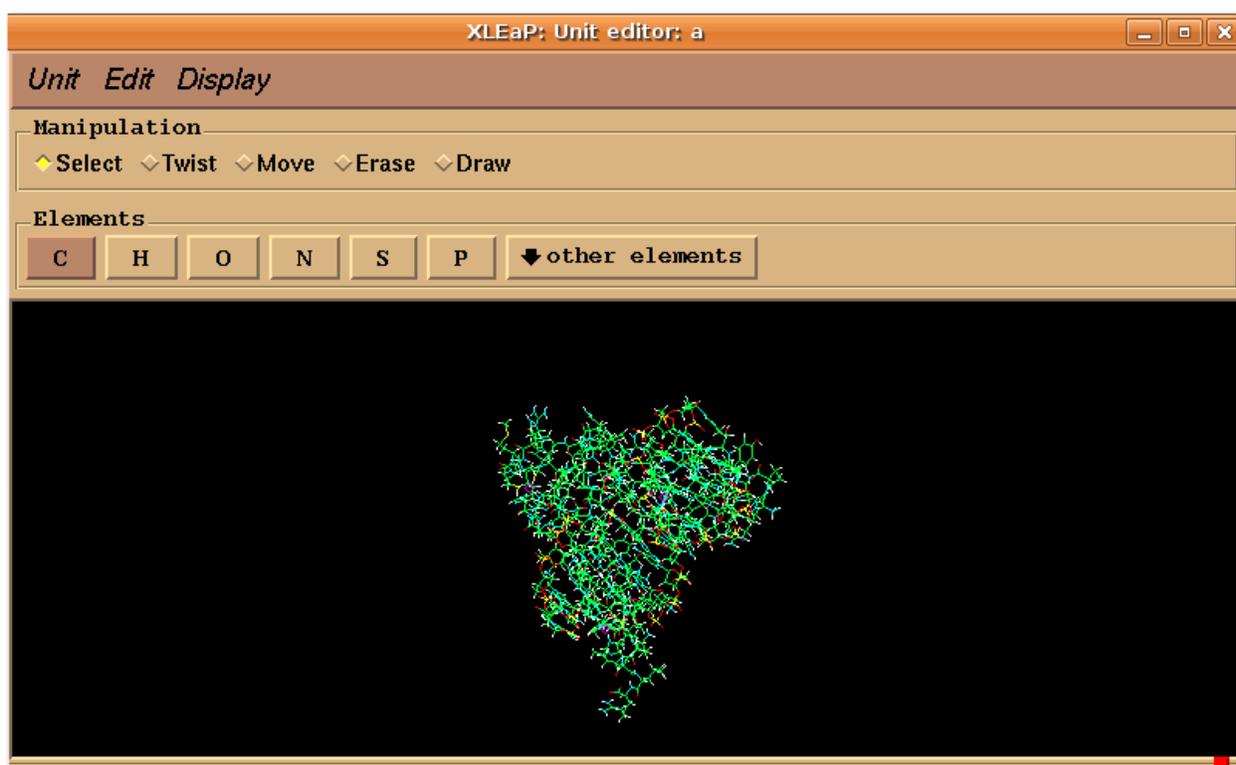


Figure 2.3 The editor window of xleap showing the graphical representation of 1AAY-dry.pdb.

The structure of ZIF268-DNA complex was examined using the command check and the it was found to be Ok. In order to bond the zinc ion with the four residues, the command bond was used as the following:

```
bond a C.5.SG C.108.ZN
bond a C.10.SG C.108.ZN
bond a C.23.NE2 C.108.ZN
bond a C.27.NE2 C.108.ZN

bond a C.35.SG C.109.ZN
bond a C.38.SG C.109.ZN
bond a C.51.NE2 C.109.ZN
bond a C.55.NE2 C.109.ZN

bond a C.63.SG C.110.ZN
bond a C.66.SG C.110.ZN
bond a C.79.NE2 C.110.ZN
bond a C.83.NE2 C.110.ZN
```

A library file was then saved to add the previous parameters to the force field, to avoid repeating all the previous steps each time we load a zinc finger protein PDB file. Instead, we only need to load this library file together with the previous AMBER PREP input, PARMSET and OFF library files each time we start xleap and before loading the 1AAY.pdb or the Zif268 mutant's PDB files , which will be mentioned later. The library file was saved using the following order:

```
saveoff a zincbond.lib
```

Then the prmtop and inpcrd files for the 1AAY-dry.pdb file were saved as 1AAY-dry.top and 1AAY-dry.crd, respectively using the saveamberparm command. The next step was to neutralize the system by adding counter ions. The total charge was found to be -5. Thus, 5 Na⁺ ions should be added to counteract the charge of the system. This task was accomplished using the following order:

```
addions Na+ 0
```

This order actually causes a columbic potential on a grid of 1Å resolution and then puts the counter ions simultaneously at the points of lowest/greatest electrostatic potential (Figure 2.4).

```
File Edit Verbosity
> addions a Na+ 0
5 Na+ ions required to neutralize.
Adding 5 counter ions to "a" using 1A grid
Grid extends from solute vdw + 1.87 to 7.97
Resolution: 1.00 Angstrom.
grid build: 0 sec
(no solvent present)
Calculating grid charges
charges: 1 sec
Placed Na+ in a at (36.52, 44.11, 46.58).
Placed Na+ in a at (27.52, 45.11, 35.58).
Placed Na+ in a at (13.52, 26.11, 30.58).
Placed Na+ in a at (33.52, 21.11, 22.58).
Placed Na+ in a at (14.52, 31.11, 34.58).

Done adding ions.
> ^
```

Figure 2.4 Neutralization of the Zif268-DNA complex by addition of sodium ions.

Finally, the system was solvated using this command:

```
solvatebox a TIP3PBOX 10.0
```

A rectangular parallelepiped box of TIP3PBOX water model was created around the solute. The number 10.0 is called the buffer argument. It indicates the distance in angstroms separating the edge of the solute from the closest atom in the solute. In this case the buffer argument consists of one number and thus the buffer distance equals 10.0 Å in x, y, and z directions (Figure 2.5).

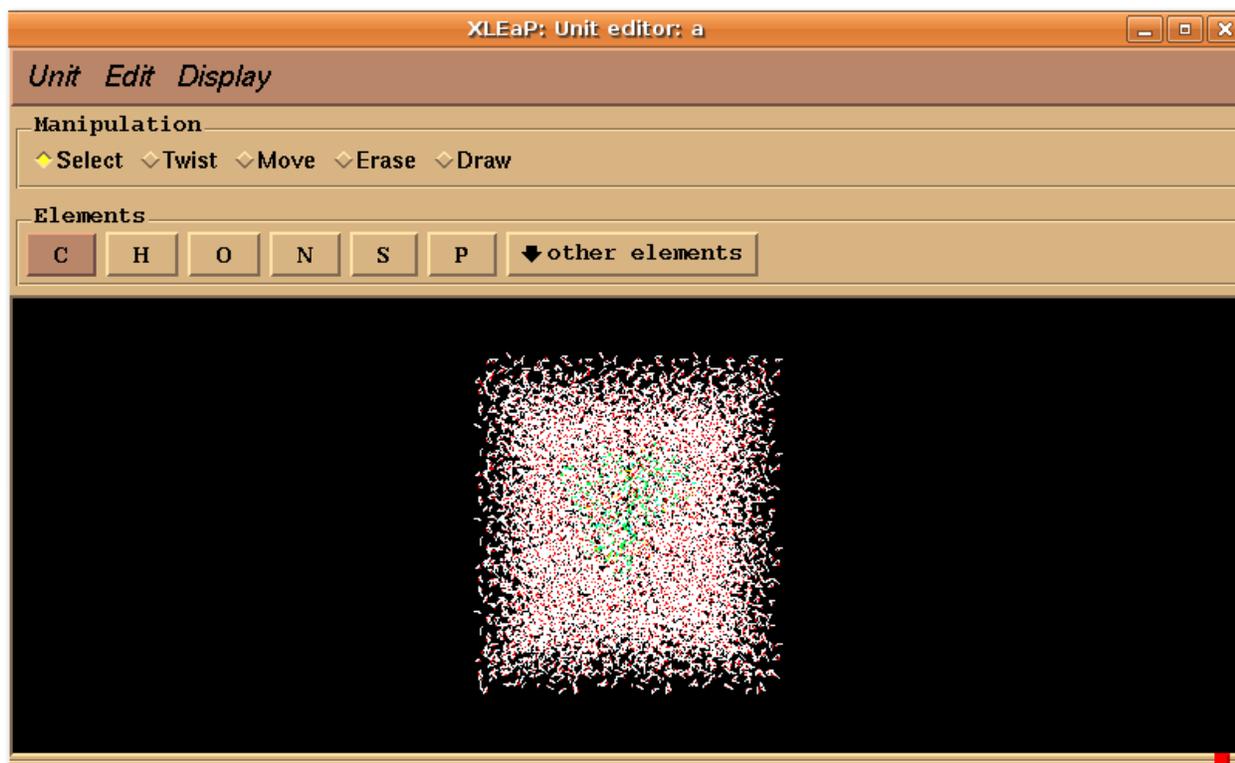


Figure 2.5 A graphical representation for Zif268-DNA solvated system.

The prmtop and inpcrd files for the solvated system were saved as the following:

```
saveamberparm a 1AAY.top 1AAY.crd
```

2.3 Equilibration of the solvated system.

Before performing MD production stage, the system was equilibrated in the following four steps: minimization, heating, density equilibration and unrestrained equilibration.

(i) Minimizing the solvated system.

Minimization was performed using 500 steps of the steepest descent method, followed by 500 steps using the conjugate gradient algorithm. Constant volume periodicity was applied. It is necessary to minimize the system before conducting molecular dynamic simulations. This is needed to take the structure to the nearest local minima in the force field being used. Minimization helps in getting rid of the greatest stress in the system that can be caused by unfavorable van der Waals and electrostatic interactions. When our system was solvated, the solute was surrounded by a box of pre-equilibrated TIP3P water. This process may leave some gaps between the solvent and solute. In addition to this, the pre-equilibrated box of water did not sense the effect of the solute. Minimizing the solvated

system would allow water molecules to be more comfortable with other components of the system, thus resulting in a more stable and relaxed solvated structure.

Performing minimization using *sander* requires three files: *prmtop*, *inpcrd*, and *mdin* files. The *mdin* file is the input file for minimization that contains parameters to control the run. The file titled **min.in** was used to perform minimization (see Appendix B).

(ii) Heating the solvated system.

The system was heated gradually to 300k using langevin dynamics. Harmonic potential with 2 kcal/mol-Å² weight was imposed on heavy atoms of the complex. The file titled **heat.in** was used to perform heating (see Appendix B).

(iii) Density equilibration.

A 50 ps of density equilibration at 300 k with constant pressure periodic boundary and positional restrains of 2 Kcal/mol-Å² weight harmonic potential was applied to adjust the density of system. The file titled **density.in** was used to perform this step (see Appendix B).

(iv) Unrestrained equilibration:

500 ps of unrestrained equilibration at 300 K and constant pressure was performed to further equilibrate the system throughout the last three steps. The SHAKE method was applied to hold all hydrogen-heavy atom bond distances. The file titled **equil.in** was used to perform this step (see Appendix B).

2.4 Production MD simulation of the solvent system.

2 ns of production runs were performed at the same condition of the preceding equilibration step in order to avoid any sudden jump in the potential energy, which would result from the heterogeneity in conditions of the system between subsequent equilibration runs. Snapshots of the MD trajectories were output every 10 ps. The 2 ns of production were carried out over four consecutive steps using the file **prod.in** (see Appendix B).

Binding free energy of Zif268-DNA complex was calculated using the MM/GBSA method. The value of ionic strength was determined to be 65mM as it is in experimental conditions. SURFTEN and SURFOFF values which are utilized

to estimate the non polar contribution to solvation energy were set to .00542 Kcal/A2.mol and 0.92 Kcal/mol , respectively.

During all the previous MD simulations, harmonic constrains with a force field of 50 Kcal/mol/A2 were applied on three zinc ions. The Particle Mesh Ewald (PME) method was utilized with 10Å cutoff for long-range interactions.

2.5 Calculating the free energy of binding for the protein-DNA complex

In order to calculate the free energy of binding for the protein-DNA complex, the following steps were performed:

- (i) The PDB file of the dry protein was split into two new files, one containing the DNA and the other containing the zinc finger protein.
- (ii) snapshots from the coordinates files of the production runs were extracted using the input file **extract.mmpbsa** (see Appendix B). The files needed to carry out the extraction were the PDB files for the desolvated complex, DNA and zinc finger protein. In addition to the production coordinate files.

- (iii) The total binding free energy was calculated from the extracted snapshots and using the input file **binding-energy.mmpbsa** (see Appendix B).

2.6 Calculating the entropy contribution.

The entropy contribution was calculated using *nmode*. The file titled **calculate-entropy.in** was used to do this (see Appendix B).

2.7 Analyzing the results

The energies were extracted from output files using a perl script. After that several summery files were plotted using *xmgrace*. RMSD values of the DNA backbone and protein backbone were calculated by supplying input files **DNA.calc_rms** and **Zif.calc_rms** to *Ptraj*.

Ptraj was used to follow up the percent occupancy and bond lengths of hydrogen bonds through the trajectory. The file **Hbond.ptraj** was used as input file in *Ptraj* to perform this task. Hydrogen bonds with percent occupancy greater than 60% were considered stable hydrogen bonds. The length of a hydrogen bond was defined as the distance between donor and acceptor atoms. Based on this

definition, the bonds with lengths (2.2-2.5 Å) were considered strong, the bonds with lengths (2.5-3.2 Å) were considered moderate and the bonds with lengths (3.2-4.0 Å) were considered weak⁶⁵.

2.8 Preparing mutants of Zif268.

Different single point mutations were introduced using PyMOL⁶⁶, each time targeting different amino acid in the linkers region through the following steps:

- i. The Zif268-DNA complex PDB file (1AAY.PDB) was loaded into PyMOL.
- ii. From the **Display** menu **sequence** was selected. This choice displays the sequences of amino acids, nucleic bases, and zinc ions composing the complex.
- iii. From the **Wizard** menu, **mutagenesis** was chosen.
- iv. Certain amino acid residue was chosen from the sequence displayed in the PyMOL viewer window.
- v. The choice **No mutation** was clicked and the resultant residue was selected as illustrated in Figure 2.6.

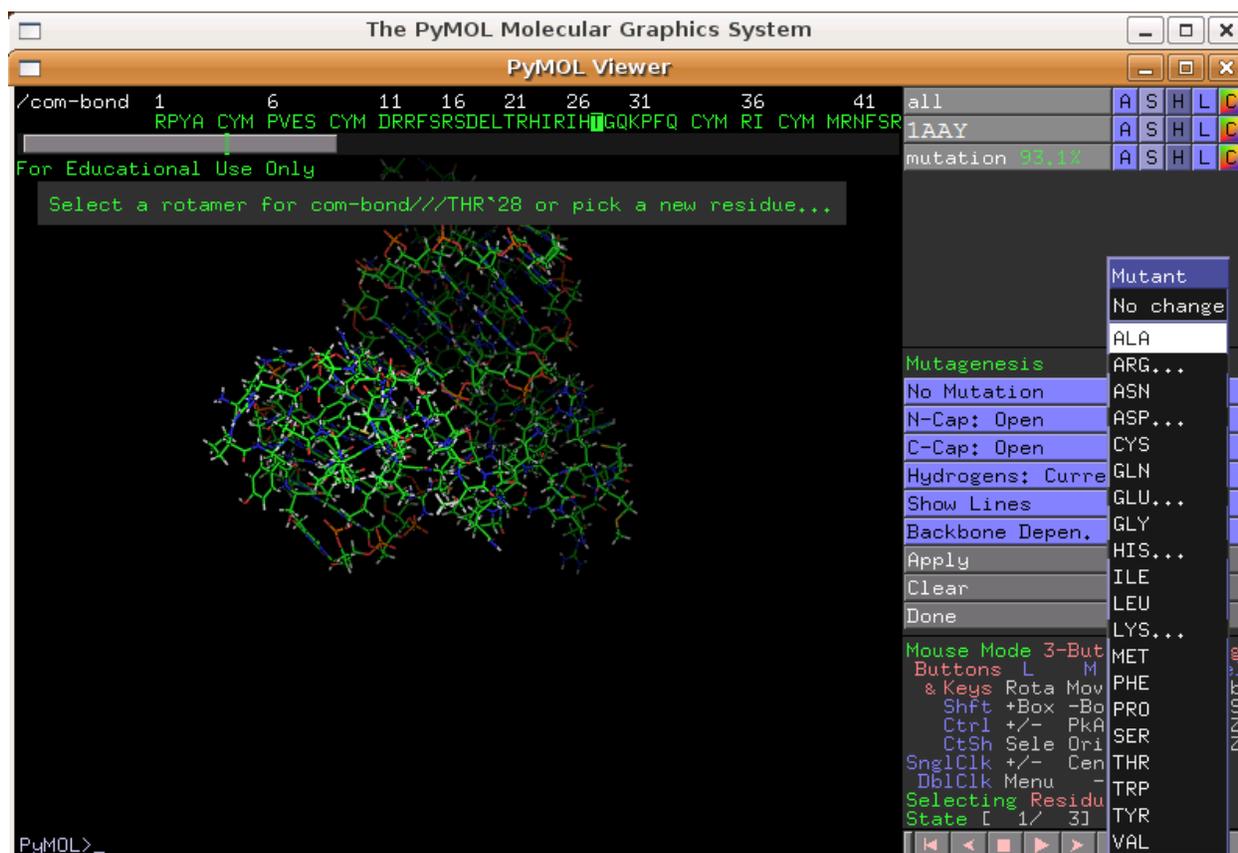


Figure 2.6 The PyMOL viewer window showing how the amino acid Threonine (T), in position 28 in Zif268, can be selected from the sequence bar and then mutated by choosing the resultant amino acid from the list after clicking “No Mutation” option.

The file of Zif268 mutant was saved as follows: the single letter abbreviation of the mutated amino acid, followed by its position, followed by the single letter abbreviation of the resultant amino acid. Each mutant PDB file was treated as 1AAJ, starting from minimization, and finishing with calculating the binding free energy for each mutant of Zif268 with DNA.

Chapter 3

RESULTS & DISCUSSION

3.1 Zif268 point mutants

In this work, each amino acid in the linker region of the zinc finger protein was mutated separately, producing ten point mutants of the zinc finger protein (Figure 3.1). Subsequently, the free energy of binding of each mutant to the binding site was calculated using MM/GBSA. Zif268 zinc finger protein with its optimal binding site (5' A GCG TGG GCG T 3')²⁰ was used as a model system for this study.

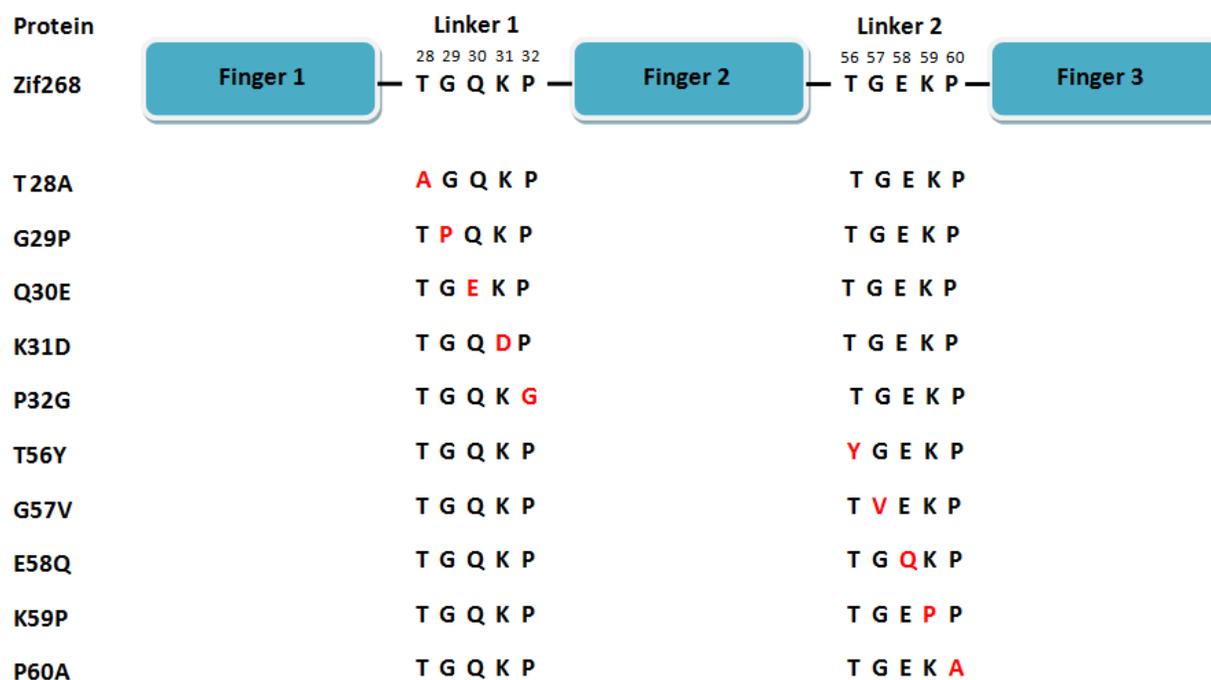


Figure 3.1 Amino acid sequences of the two canonical linkers in Zif268 and ten of its point mutants. Positions of the amino acids in the linkers with respect to the first amino acid in finger 1 are indicated above each residue. Amino acids are represented by their one letter abbreviation. Amino acids in red show the resultant amino acid in each mutant.

3.2 Analysis of simulation output files

Over different stages of simulations, the system under study is supposed to reach an equilibrated state and subsequently maintain this state. Accomplishing this state should be checked by monitoring how different properties of the system

changed during the simulation⁶⁷. This type of analysis is important in order to make sure everything went correctly, and no sudden jumps of the system properties took place. As such, we extracted the system properties from the output files of the different equilibration runs, and here we present plots of the density, temperature, pressure, and energy versus time. Figure 3.2 to 3.4 shows the different plots for Zif268-DNA complex.

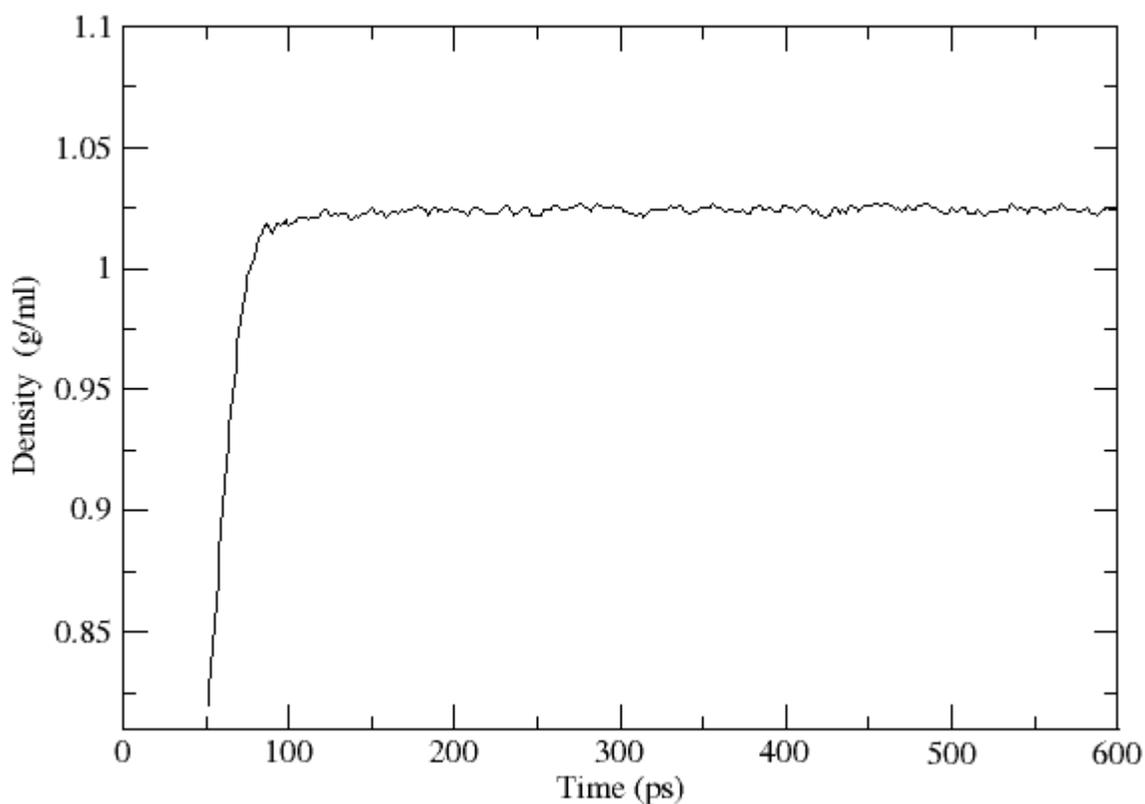


Figure 3.2 Density of Zif268-DNA system during equilibration runs.

The first 50 ps of the simulation represents the heating stage for the system, which was carried out under constant volume, thus no density data were recorded. After that, the density increased to about 1.024 g/ml and stayed around that over the last 550 ps of the simulation. The value of the equilibrated density corresponds to the density of pure water (1g/ml) to which the protein-DNA complex has been added, leading to the slight rise in the density of the system.

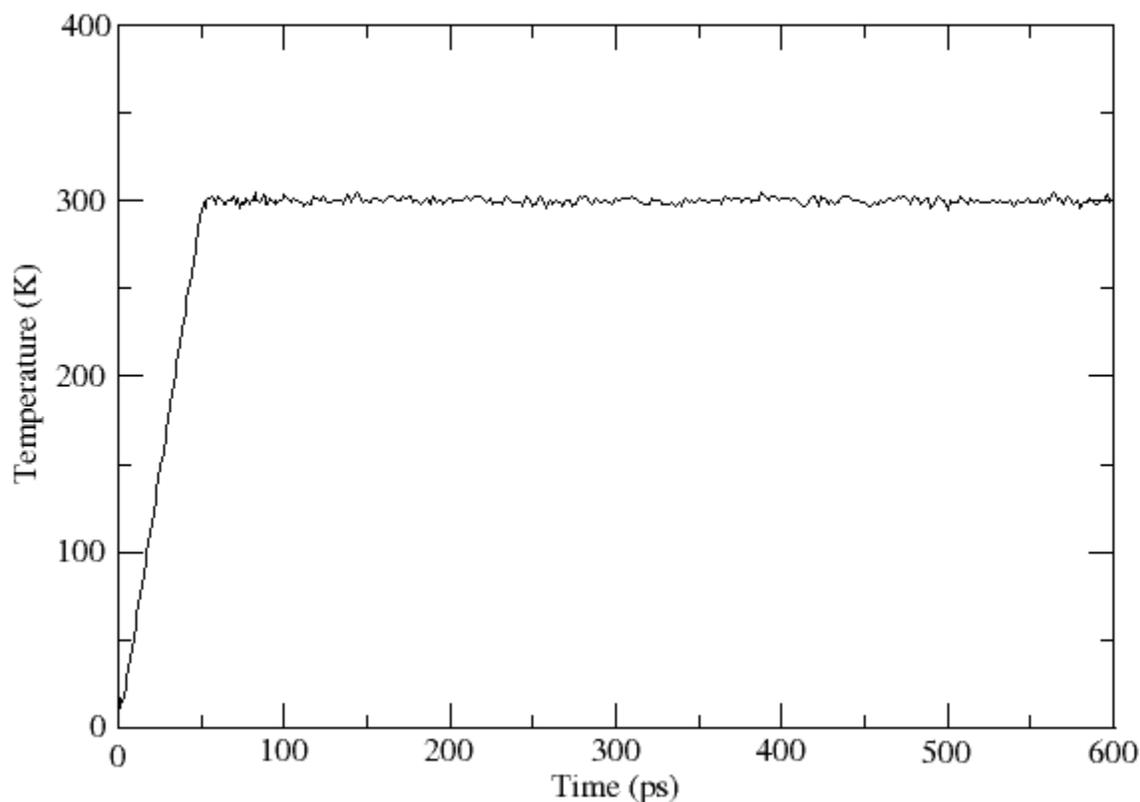


Figure 3.3 Temperature of Zif268-DNA system during equilibration runs.

In the plot of temperature versus time (Figure 3.3), the temperature rises

regularly from 0 K to 300K, which is consistent with the heating stage. After that, the temperature is equilibrated around 300 K over the rest of the simulation, indicating that langenvin dynamics worked effectively.

At the beginning of the simulation, during the heating step, pressure data were not collected because the system was under constant volume. Over the remaining part of the simulation the system was kept under constant pressure. However, the plot in Figure 3.4 shows that in the time between 50 ps to 100 ps the pressure decreased slightly then it fluctuated around a mean value of 1 atm. This kind of behavior was reported as adequate to prove that the system reached the required equilibrium state⁶⁸.

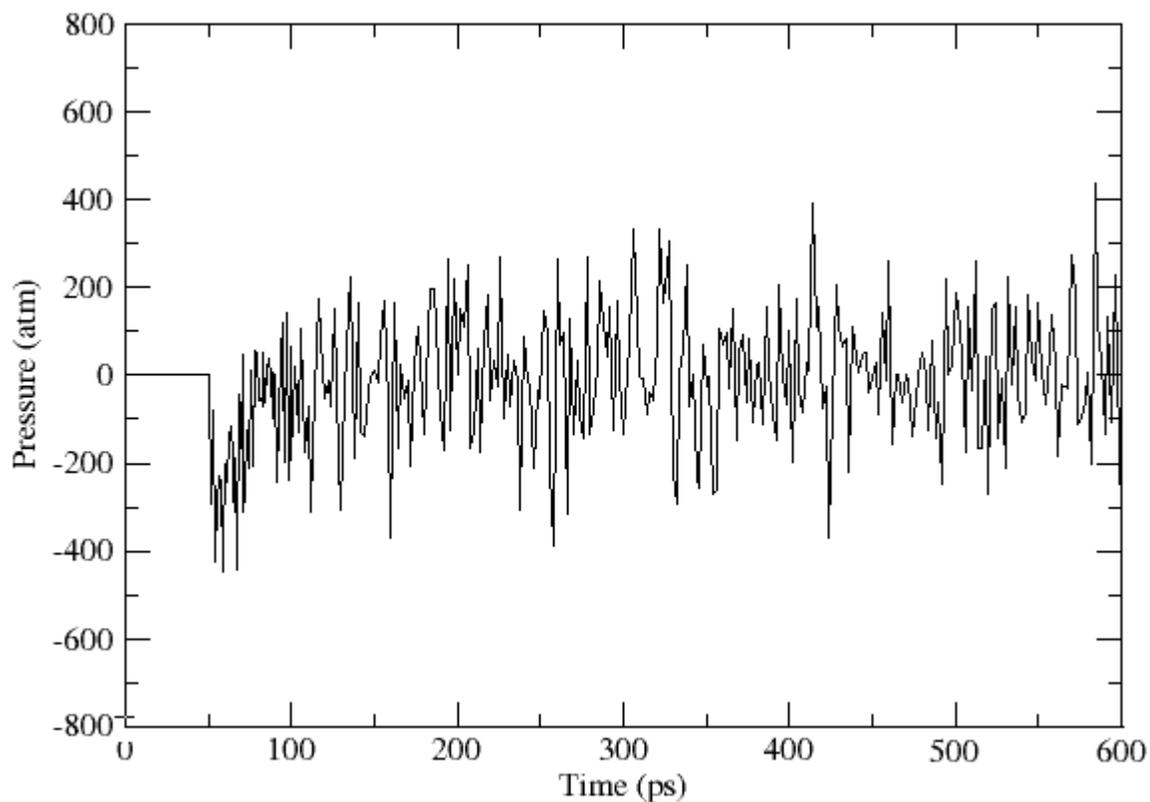


Figure 3.4 Pressure of Zif268-DNA system during equilibration runs

It is worth mentioning that negative pressure values represent a “force” trying to reduce the volume of the water box, whereas the positive values represent a “force” trying to make the water box larger⁶⁸.

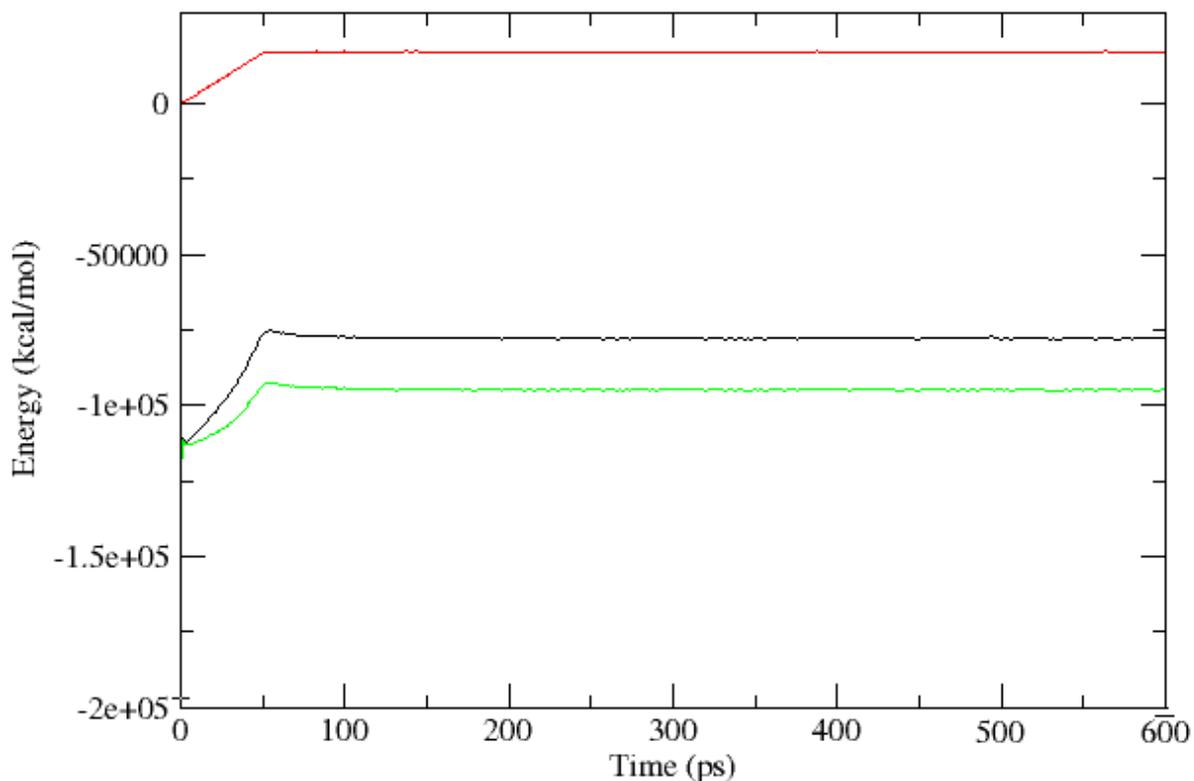


Figure 3.5 Total, kinetic and potential energy of Zif268-DNA during equilibration runs. The red, black and green plots represent kinetic, total and potential energies, respectively.

The energy plots (Figure 3.5) show a rise in both forms of energy during the first 50 ps of simulation. After that the kinetic energy remained constant indicating successful performance of temperature thermostat which affects the kinetic energy. On the other hand, the potential energy plot shows a slight decrease that corresponds to system relaxation due to simulation under constant pressure,

followed by potential energy stability through the rest of the simulation. Total energy is the sum of kinetic and potential energy. As such, it shows a plot consistent with the behavior of the two forms of energy.

RMSD variation for DNA backbone and protein during the equilibration run from their starting structures was calculated for each one separately. Figure 3.6 shows that the RMSD value for the DNA backbone increased rapidly in the first 75 frames, then it fluctuated steadily around 1.4 Å, which is an acceptable value. Figure 3.7 indicates that after the first 100 frames, the RMSD values fluctuated around 1.2 Å, which indicates that conformational changes in the protein backbone were acceptable.

Analysis of output files for all of the mutants-DNA complexes showed similar results.

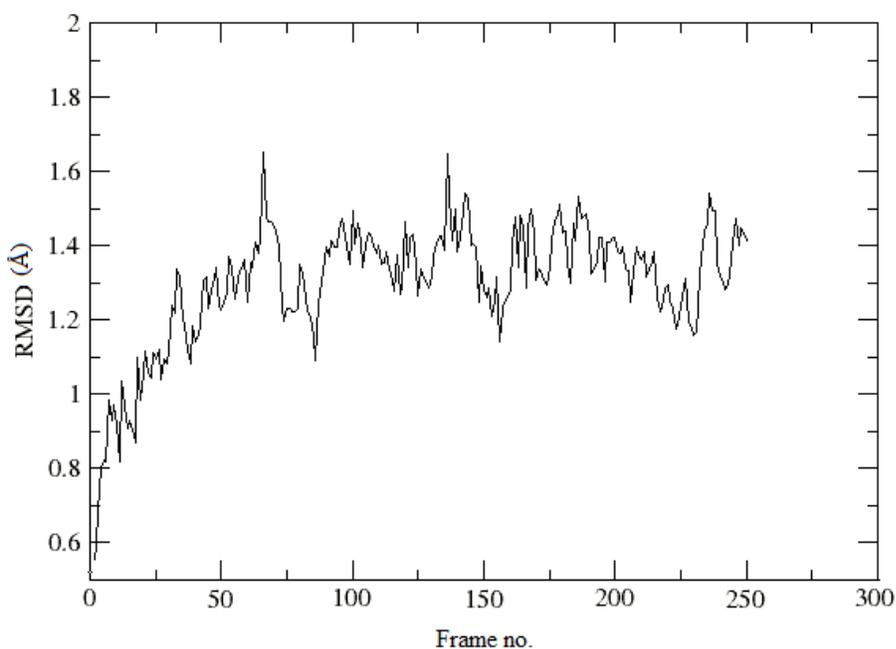


Figure 3.6 RMSD of DNA backbone during the unrestrained equilibration run of Zif268-DNA complex.

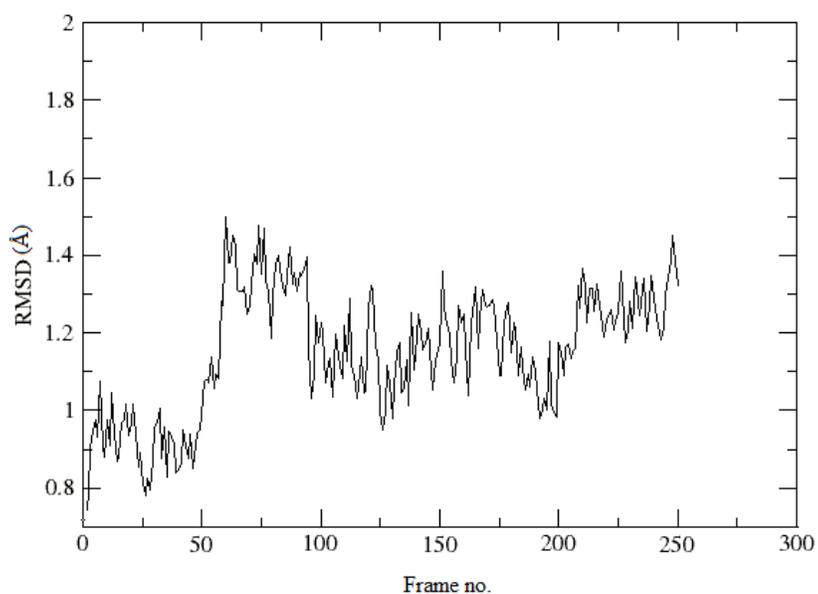


Figure 3.7 RMSD of protein backbone during the unrestrained equilibration run of Zif268-DNA complex.

3.3 Total binding free energy (GBTOT) for Zif268-DNA complex compared to its ten point mutants bound to the same DNA binding site.

In this study, we present the values of the binding free energy of Zif268-DNA complex and ten single point mutants of this complex (Table 1) calculated using MM/GBSA method.

Table 3.1: Total binding free energy, electrostatic energy as calculated by molecular dynamics(ELE), van der Waals energy as calculated by molecular dynamics(VDW), electrostatic contribution to solvation free energy (GBCAL) and non polar solvation free energy(GBSUR) for Zif268-DNA complex and for ten of its point mutants to the same binding site. The unit of all energy terms are in kcal/mol.

Protein	GBTOT	STD	ELE	VDW	GBCAL	GBSUR
1AA Y	-179.39	16.89	-5894.6	-139	5878.44	-24.1
T28A	-193.77	10.04	-5942.1	-136.18	5907.82	-23.34
G29P	-165.95	11.86	-5781.6	-125.22	5763.14	-22.26
Q30E	-204.9	12.14	-5704.4	-147.58	5671.71	-24.66
K31D	-171.4	10.61	-5072.8	-126.5	5050.51	-22.65
P32G	-167.97	9.62	-5883.9	-120.22	5855.91	-22.08
T56Y	-158.65	11.88	-5778.3	-136.04	5778.32	-22.58
G57V	-170	9.8	-5803	-137.04	5794.85	-24
E58Q	-197.55	9.89	-6282.7	-131.22	6239.02	-22.7
K59P	-163.56	10.3	-5496.5	-124.26	5480.08	-22.91
P60A	-207.01	10.02	-5931.6	-129.03	5877.08	-23.47

The binding free energy of mutants varies according in the to the type of amino acid mutated in the wild type Zif268-DNA complex, most pronounced in P60A, T56Y, Q30E and E58Q which experienced a considerable change in their binding free energy upon mutation, where one

mutant (T56Y) resulted in lower binding energy by 20.74 kcal/mol; whereas the other three mutants (Q30E, E58Q and P60A) produced considerably higher binding energy (by 25.5, 18.2, 27.6 kcal/mol, respectively). The effect of mutation was less pronounced for the rest Six mutants (T28A, G29P, K31D, P32G, and G57V, and K59P) which produced binding energy values within the standard deviation from the binding energy of the wild type , where T28A showed an increase in the free binding energy by 14.38 kcal/mol, whereas G29P, K31D, P32G, G57V, and K59P showed a decrease in the free binding energy by 13.44, 8, 11.42, 9.4, 15.83 kcal/mol , respectively. (Figure 3.8.a). The binding free energy for the proteins with the optimal binding site vary in the order P60A > Q30E > E58Q > T28A > 1AAY > K31D, P32G, G57V > T56Y (Figure 3.9.b).

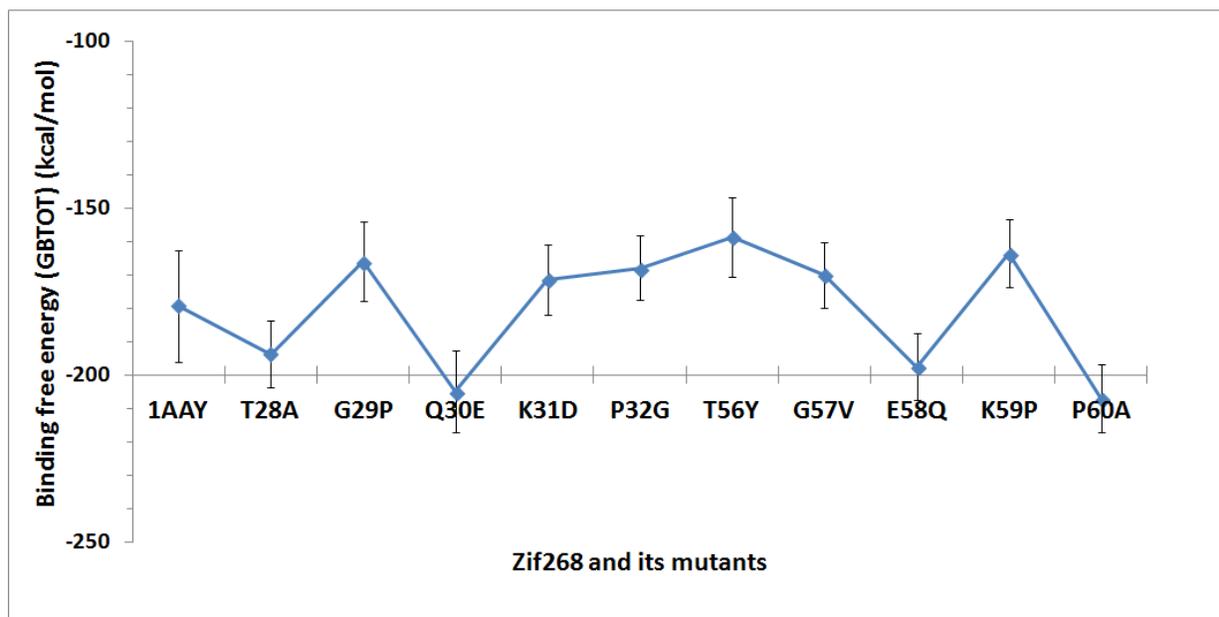


Figure 3.8.a The Binding free energy for Zif268-DNA complex and for ten of its point mutants to Zif268 optimal binding site .

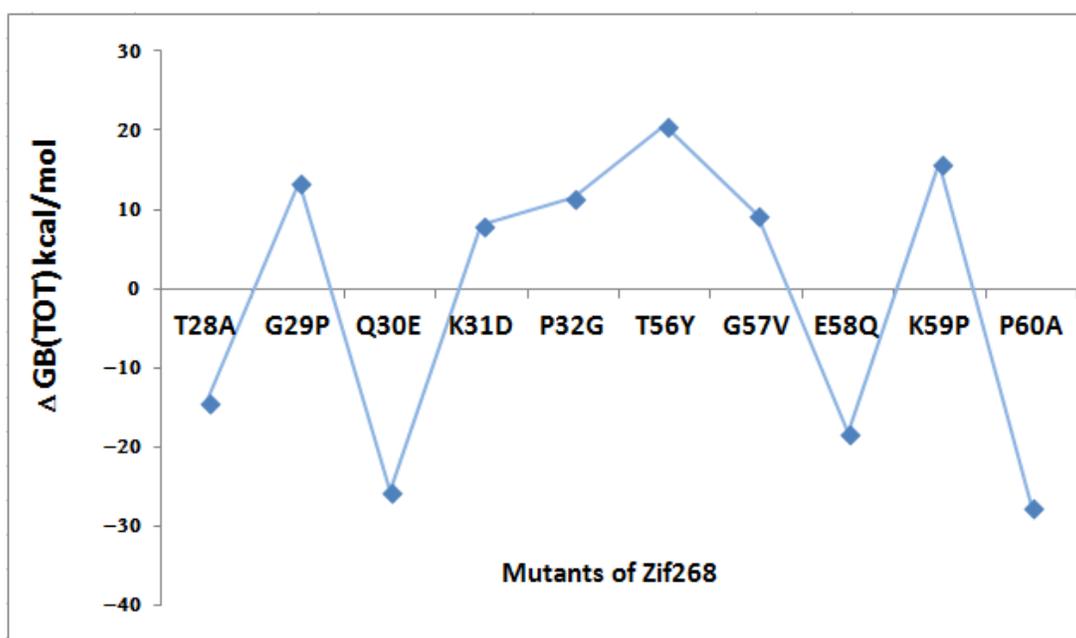


Figure 3.8.b The change in binding free energy upon point mutations in the linker region of Zif268.

The major contributor to the binding free energy for all of the complexes are the electrostatic energy, as calculated by molecular mechanics (ELE), and the electrostatic contribution to the solvation free energy, as calculated by the GB method (GBCAL). Nonetheless, these two predominant values have opposite signs and thus cancel each other out (Figure 3.9), allowing the van der Waals contribution calculated by molecular mechanics (VDW) to have a significant impact on the value of binding free energy⁶⁹.

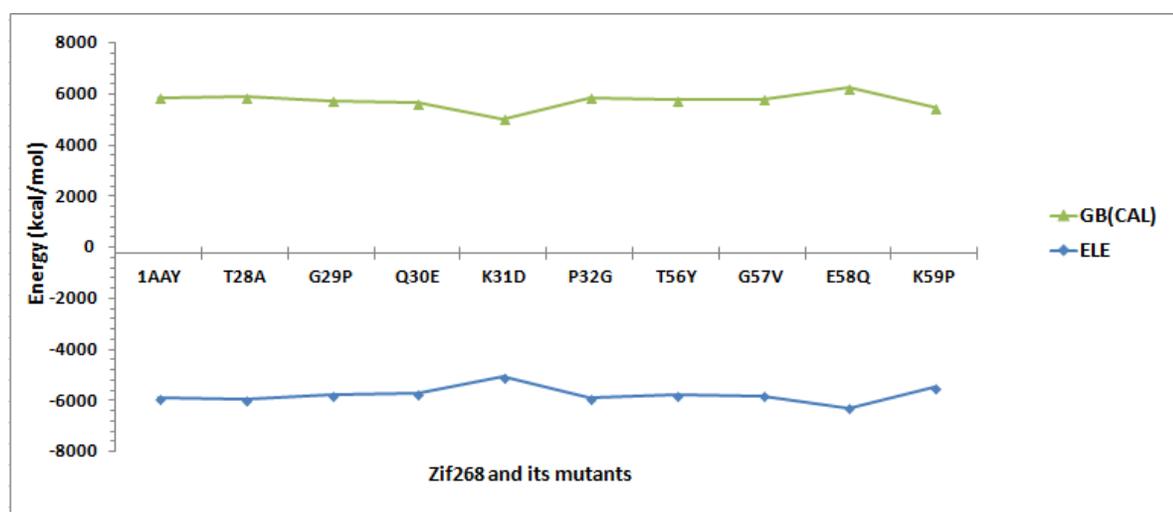


Figure 3.9 The major contributors to the binding free energy of Zif268 and its mutants to Zif268 optimal binding site: the electrostatic energy as calculated by molecular mechanics (ELE) and the electrostatic contribution to the solvation free energy calculated by GB method (GBCAL).

The binding free energy values are well correlated with the total electrostatic energy (i.e. the sum of the electrostatic energy as calculated by molecular mechanics (ELE) and the electrostatic contribution to the solvation free energy calculated by GB method (GBCAL)). The correlation coefficient in this case equals 0.86, (Figure 3.10).

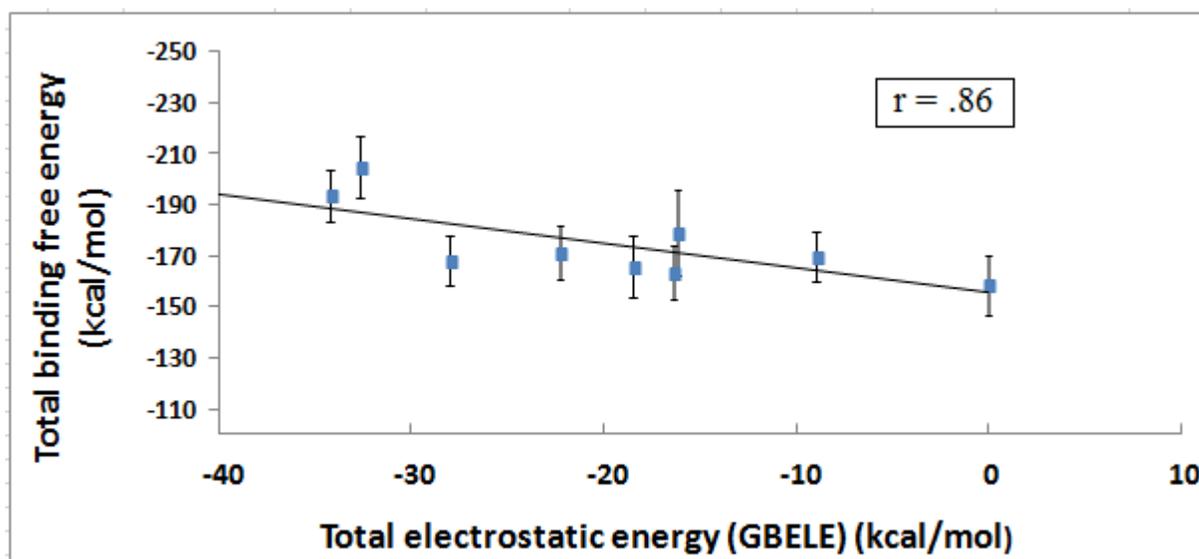


Figure 3.10 Correlation between total binding free energy and total electrostatic energy, upon point mutations in the linkers.

To the contrary, the correlation between the sum of van der Waals and non-polar contributions, and the binding free energy, is weaker (Figure 3.11). This suggests that the small difference between the two opponent energies (i.e. the

electrostatic interactions (as calculated by molecular mechanics and the electrostatic contributions to solvation) plays a vital role in deciding the value of the total binding free energy (Figure 3.12).

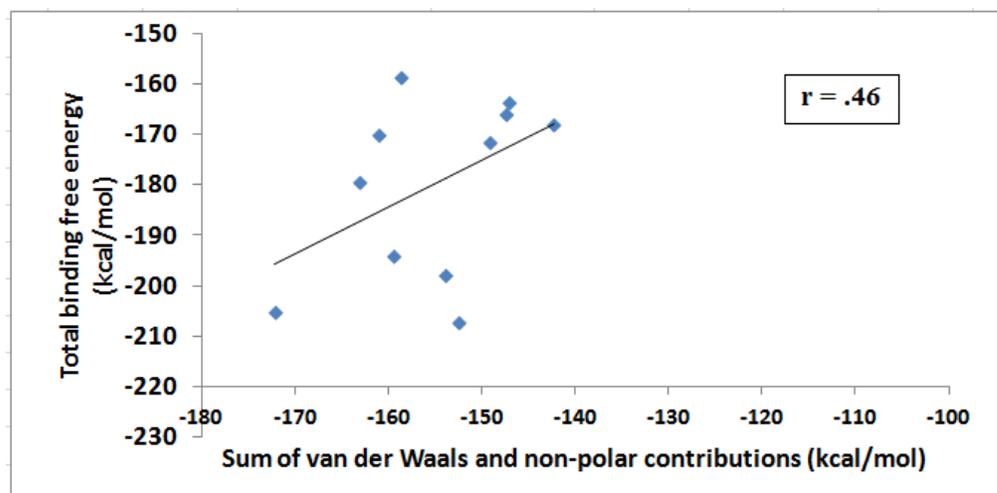


Figure 3.11 Correlation between total binding energy and the sum of van der Waals and non-polar contribution to solvation, upon point mutations in the linkers.

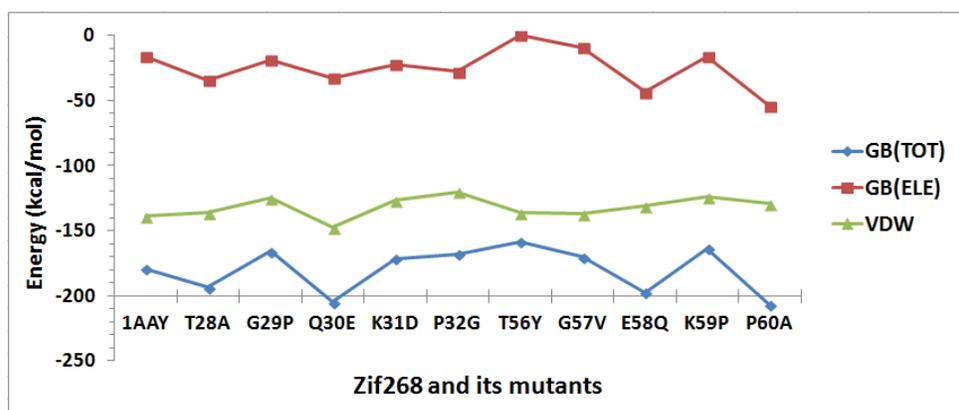


Figure 3.12 Contributions of total electrostatic energy and van der Waals energies to the total binding energy of Zif268 and ten of its point mutants to the same DNA binding site.

3.3.1 Analysis of the binding free energy of Zif268-DNA complex.

The total binding free energy (GBTOT) for Zif268-DNA complex is -179.39(\pm 16.89) kcal/mol. The negative value is indicative of favorable protein-DNA binding in water. This binding affinity reflects contributions from several interactions, including specific interactions between amino acids and DNA bases, non-specific interactions between amino acids with DNA backbone, and interactions between amino acid side chains.^{69,70}

3.3.2 Analysis of the binding free energy of T28A-DNA complex:

Crystal structure of Zif268-DNA complex showed that the first linker residue (Thr28) is involved in two types of interactions.⁶ The first is a hydrophobic interaction by the Threonine's methyl group, which was suggested to play an important role in stabilizing the finger by aiding in shielding the zinc coordination sphere from the solvent. The second interaction is a hydrogen bond between the hydroxyl group in Threonine and the back bone amide of the third amino acid in the linker. A similar hydrogen bond in the zinc finger protein TFIIA was found to help in forming DNA-induced capping of the α -helix of its first finger³⁶. Mutating the first amino acid linker (Threonine) to Leucine in TFIIA resulted in a noticeable reduction in affinity of this zinc finger protein with its DNA binding site. In order

to study the effect of these two interactions on the free energy of binding, Threonine was mutated to Alanine.

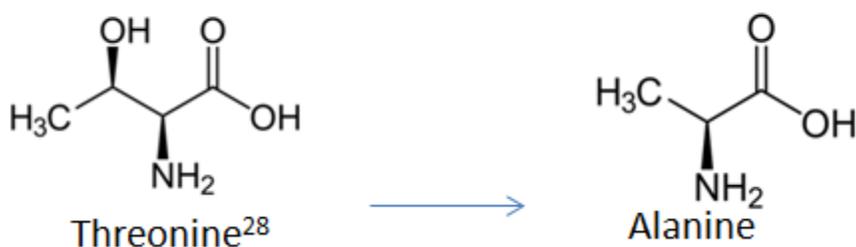


Figure 3.13. Mutating the first linker residue threonine²⁸ to Alanine.

The binding free energy for T28A-DNA complex as calculated using MM/GBSA equals -193.77 ± 10.04 kcal/mol. In view of this result T28A shows a slightly greater binding free energy than the wild type peptide. This result indicates that the loss of the hydrogen bond between Thr²⁸ and backbone amide of Gln³⁰ didn't cause a significant reduction in binding affinity as it did in the zinc finger protein TFIIIA²⁸. The reason behind this increase is unclear but our computational analysis of the factors contributing to binding free energy suggests that this mutation may cause a decrease in some unfavorable electrostatic interactions. One possible explanation of this result is that the replacement of Threonine with Alanine, which has a smaller side chain, reduced the steric hindrance between the

atoms of this amino acid and the neighboring atoms, taking into consideration that this amino acid is cited in the α -helix of finger 1.

The hydrogen bonds between Zif268-DNA and the mutant T28A-DNA were studied using *Ptraaj*. Subsequently, the percent occupancy during the production runs and the average lengths of these bonds were compared. Emphasis was on the first linker, second linker, and crucial hydrogen bonds between conserved amino acids in the α -helices of the protein and bases in the major groove of DNA. The occupancy of hydrogen bonds over the trajectory of the production runs reflects the stability of these bonds during the production simulations. The changes in the percent occupancy and lengths of hydrogen bonds from the first linker in T28A-DNA and Zif268-DNA complexes are shown in Figures 3.14 and 3.15.

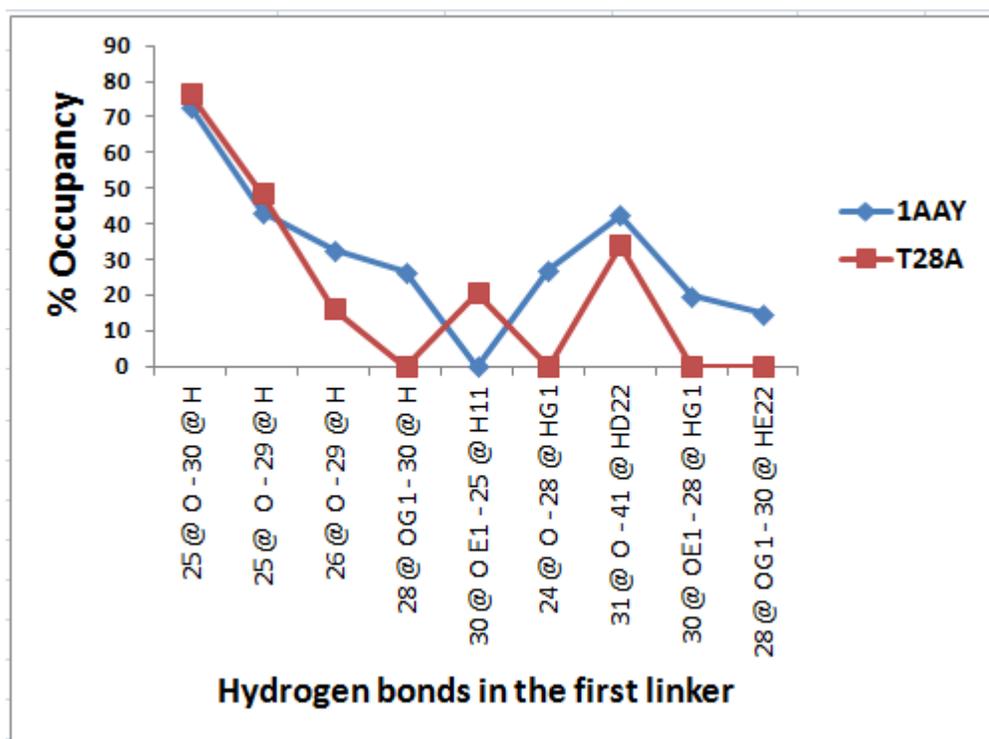


Figure 3.14 Percent occupancy of hydrogen bonds over production runs of 1AAY and T28A. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered. Each hydrogen bond is represented as (residue no. @ atom of the donor – residue no. @ atom of the acceptor).

As can be seen from Figure 3.14, mutating Thr²⁸ to Ala caused a loss of the hydrogen bond between Thr²⁸ and Gln³⁰, in addition to losing three unstable hydrogen bonds. This mutation also caused a decrease in the percent occupancy of hydrogen bond between the backbone amides of Ile²⁶ and Gly²⁹. On the other hand, it created a new hydrogen bond between Arg²⁵ and the carbonyl oxygen of Gln³⁰. The hydrogen bond between the carbonyl oxygen of Arg²⁵ and backbone amide of

Gly²⁹ wasn't affected upon mutation, this bond is responsible for forming “Gly C-cap” that help in terminating the α -helices^{6,36}. The lengths of these hydrogen bonds were compared and no significant changes due to this mutation were observed.

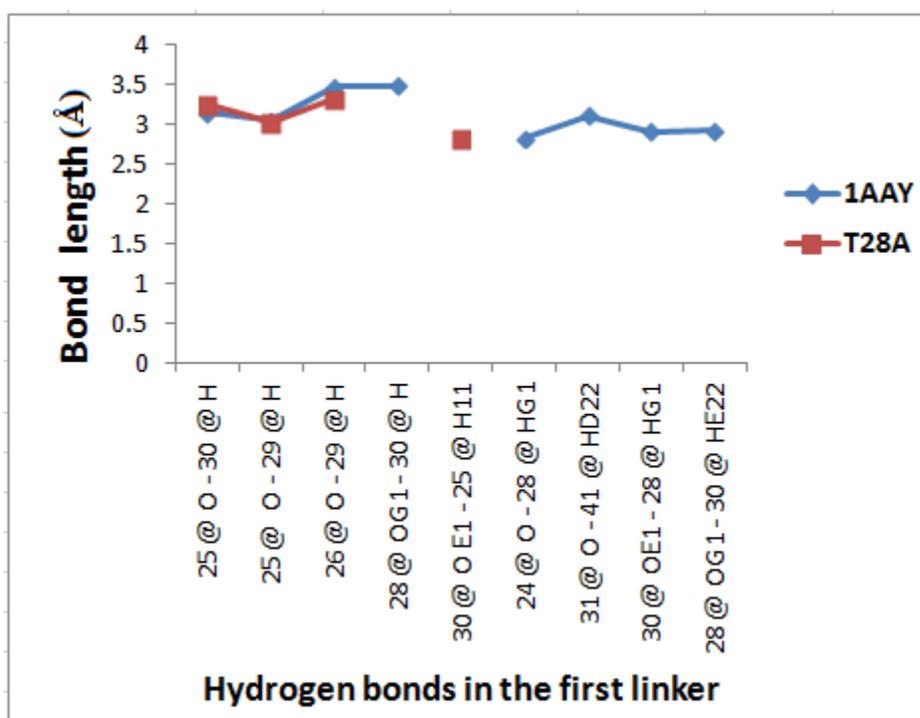


Figure 3.15 Average lengths of hydrogen bonds over production runs of 1AAV and T28A. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered. Each hydrogen bond is represented as (residue no. @ atom of the donor – residue no. @ atom of the acceptor).

Figure 3.16 shows the percent occupancy of hydrogen bonds in the second linker, as can be seen the stability in T28A mutant increased for some hydrogen

bonds and decreased for the others. Thus no certain trend can be observed concerning the stability changes of hydrogen bonds due to the mutations. The length of hydrogen bonds, however, remained stable (Figure 3.17).

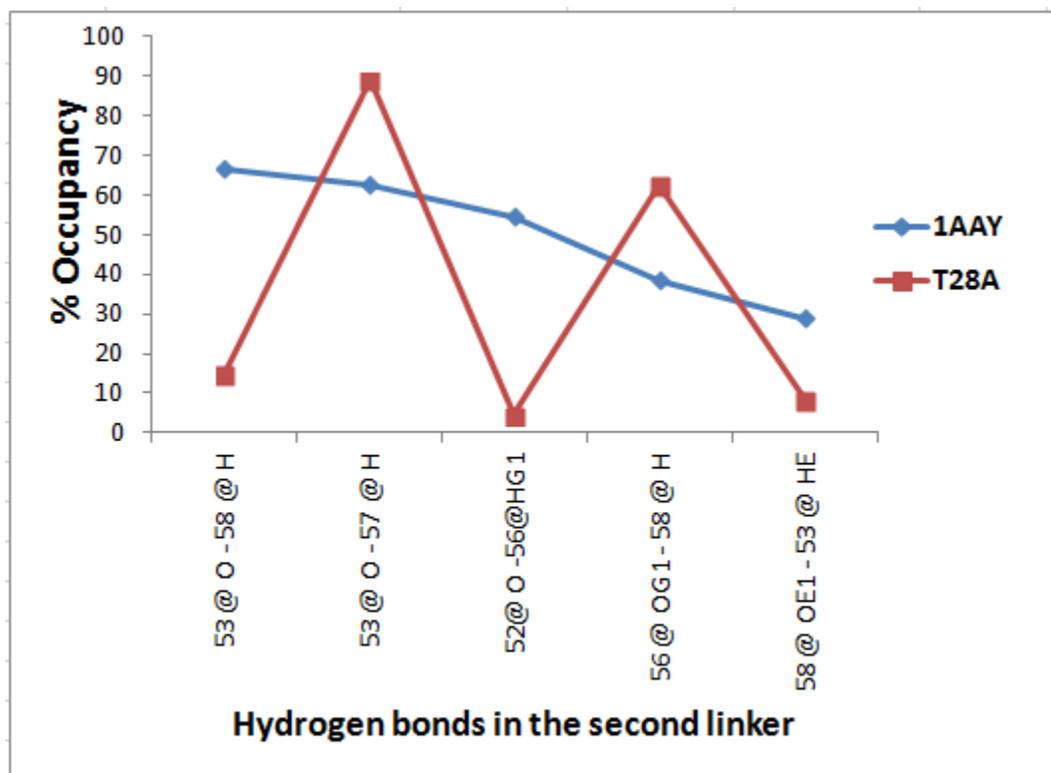


Figure 3.16 Percent occupancy of hydrogen bonds over production runs of 1AAY and T28A. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered. Each hydrogen bond is represented as (residue @ atom of the donor - residue @ atom of the acceptor).

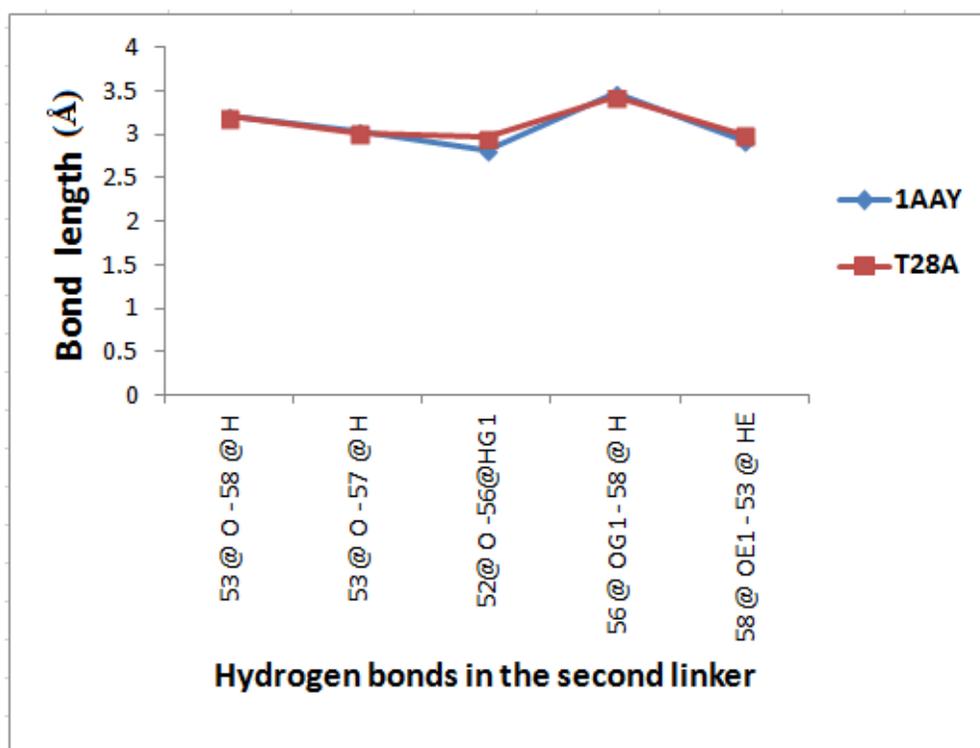


Figure 3.17 Average lengths of hydrogen bonds over production runs of 1AAY and T28A. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered. Each hydrogen bond is represented as (residue no. @ atom of the donor – residue no. @ atom of the acceptor).

The occupancy of the crucial hydrogen bonds, connecting side chains of conserved amino acids in the α -helices of the protein with bases in DNA major groove, were monitored and analyzed in the Zif268-DNA and T28A-DNA complexes (Figure 3.18). The highly stable hydrogen bonds (occupancy $\geq 80\%$) in the T28A-DNA complexes did not suffer any noteworthy changes in their stability due to mutation, except for two hydrogen bonds. The first, is significant destabilization of the a crucial hydrogen bond connecting two conserved amino

acids, Arg¹⁶ and Asp¹⁸. This bond is one of two bonds that are supposed to stabilize and fit Arg¹⁸ to specifically interact with the guanine base (G10)⁶. Moderately stable Hydrogen bonds (those with percent occupancy: 60% < % occupancy < 80%) showed destabilization of most of the bonds. However, the less stable hydrogen bonds were stabilized to further extent after the mutation.

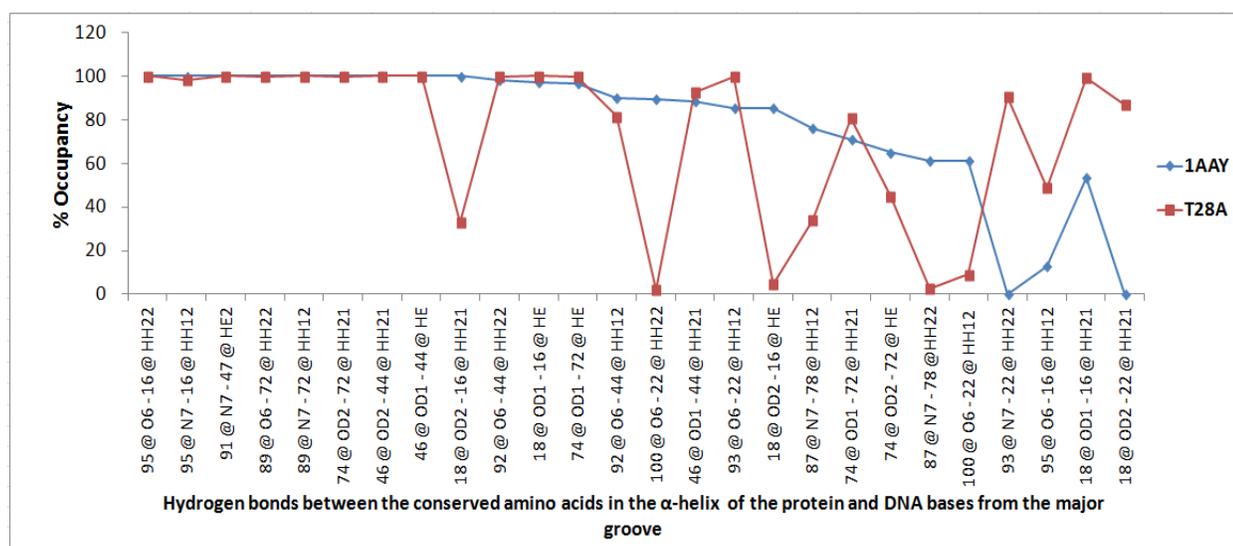


Figure 3.18 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and T28A. Each hydrogen bond is represented as (residue @ atom of the donor - residue @ atom of the acceptor).

The bond lengths analysis shown in Figure 3.19, indicates a measurable change in the length of two hydrogen bonds only. These are the hydrogen bonds

formed between the two conserved amino acids, Arg¹⁶ and Asp¹⁸, mentioned previously. The first bond (18 @ OD2 - 16 @ HH21) experienced an increase in its length of 0.59 Å which is consistent with the observed destabilization of this bond after the mutation. The second bond (18 @ OD1 - 16 @ HH21) suffered a decrease in its length of about 0.55 Å which is consistent with the observed increase in its stability. These two details suggest a potential sliding of the Asp carboxylate group, which brings the carboxylate oxygen (OD1) closer to Arg hydrogen atom (HH21), and simultaneously shifts the other Asp carboxylate oxygen (OD2) away from the Arg hydrogen (HH21).

These results suggest that mutating Thr²⁸ to Alanine altered the stability of hydrogen bonds and probably other electrostatic interactions all over the protein without limiting this effect on the neighboring region.

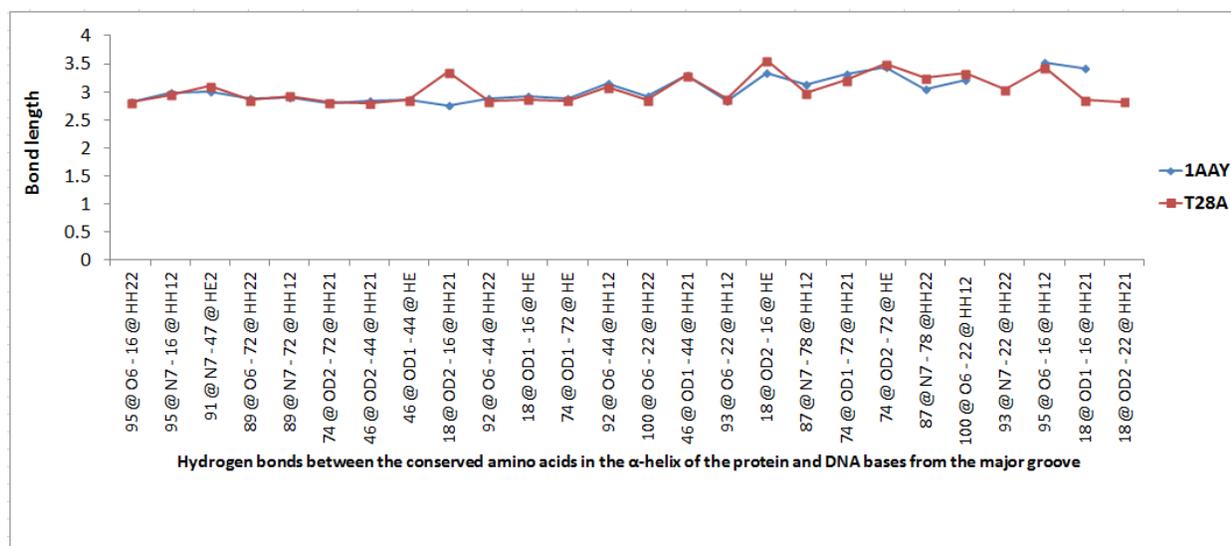


Figure 3.19 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and T28A. Each hydrogen bond is represented as (residue @ atom of the donor - residue @ atom of the acceptor).

3.3.3 Analyzing the binding free energy of G29P:

The second linker residue is Glycine (Gly29). It is located at the end of the α -helix. Gly29 was suggested to contribute toward terminating the α helix⁶, through forming “Gly C-cap” via a hydrogen bond between the CO from Arg25 and backbone amide of Gly29^{6,36}. This small amino acid was mutated to Proline to study the effect of this small sized amino acid on the binding free energy, and to examine the effect of Gly C-cap on stabilizing Zif268-DNA complex. (Figure3.20)

accompanied by a great destabilization of two hydrogen bonds. One of these bonds connects Arg25 and Gly29, and is responsible for C-capping of the α -helix. The hydrogen bonds involving amino acids in the second linker maintained their stability in general. Hydrogen bond analysis for the conserved hydrogen bonds in the α - helices of the protein with DNA bases in the major groove showed that among the highly stable hydrogen bonds (i.e. those with percent occupancy ≥ 80) four hydrogen bonds were destabilized, whereas most of the moderately stable and non-stable hydrogen bonds were stabilized. No significant changes in bond length were observed except for one case, where the length of the bond (100 @ O6 – 22 @ HH22) increased by $\sim .5 \text{ \AA}$. This increase is consistent with the observation that this bond was the most affected bond by destabilization in the G29P mutant. (Figure 3. 21 to 3.26)

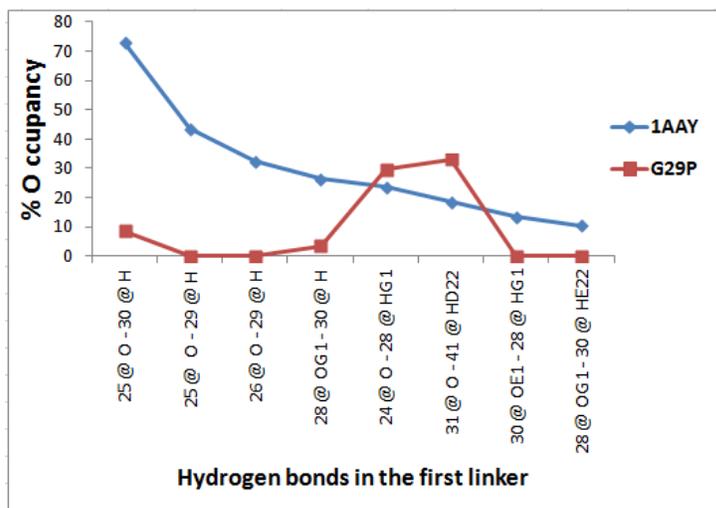


Figure 3.21 Percent occupancy of hydrogen bonds over production runs of 1AAY and G29P. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

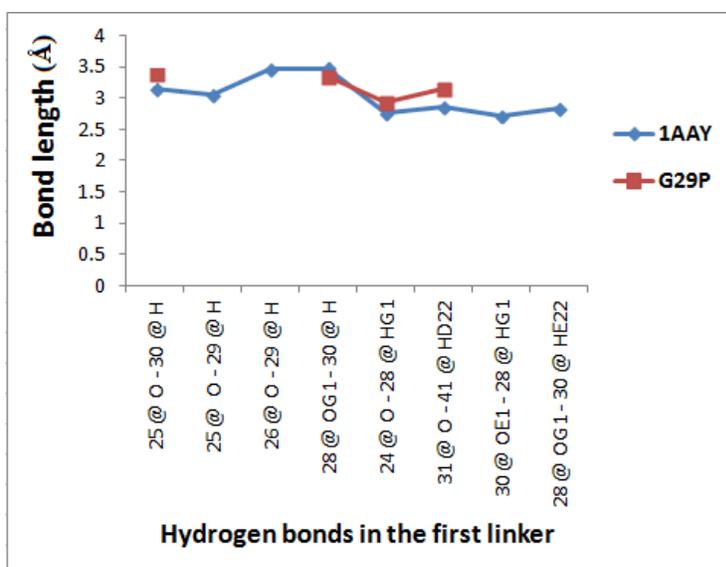


Figure 3.22 Average lengths of hydrogen bonds over production runs of 1AAY and G29P. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

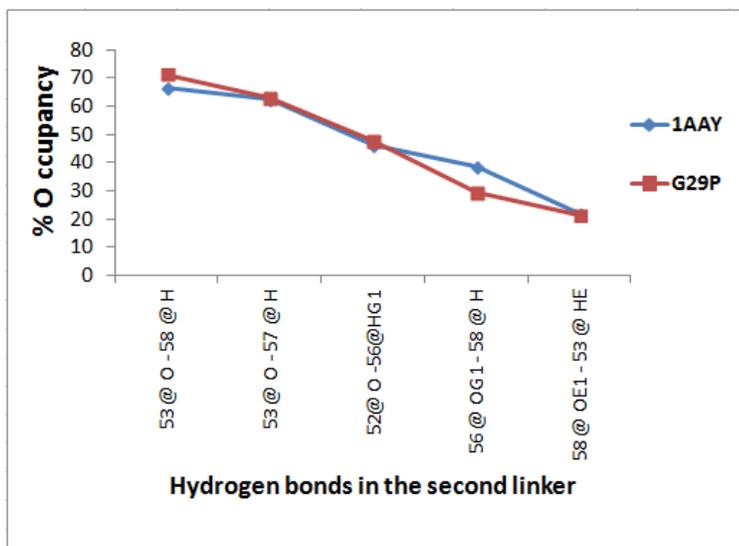


Figure 3.23 Percent occupancy of hydrogen bonds over production runs of 1AAY and G29P. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

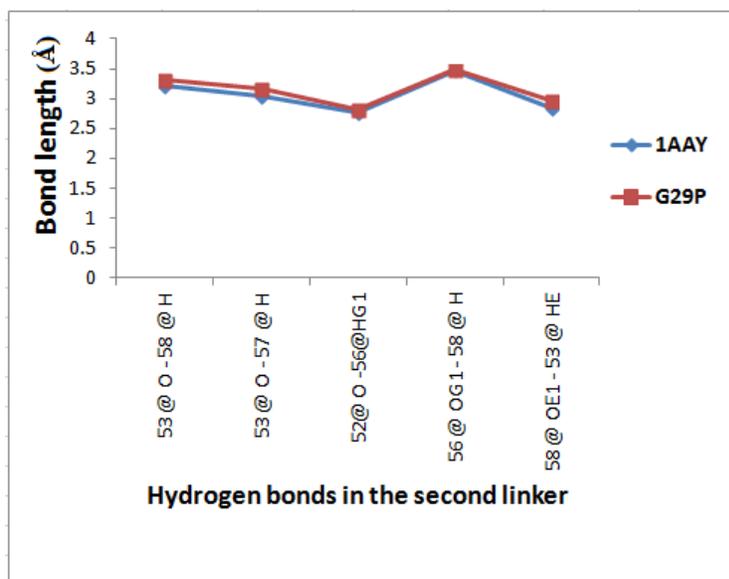


Figure 3.24 Average lengths of hydrogen bonds over production runs of 1AAY and G29P. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

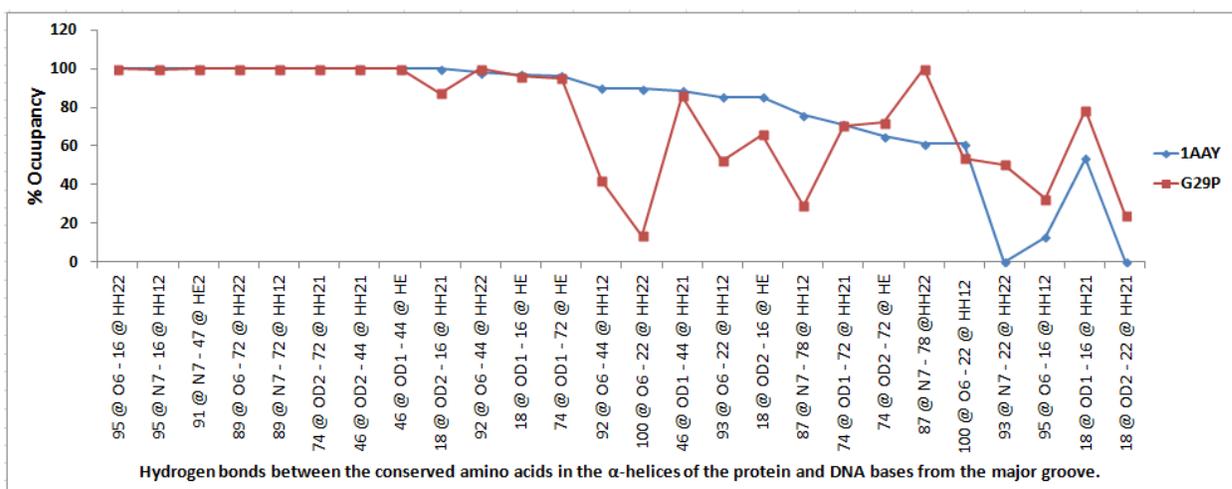


Figure 3.25 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major, over production runs of 1AA and G29P.

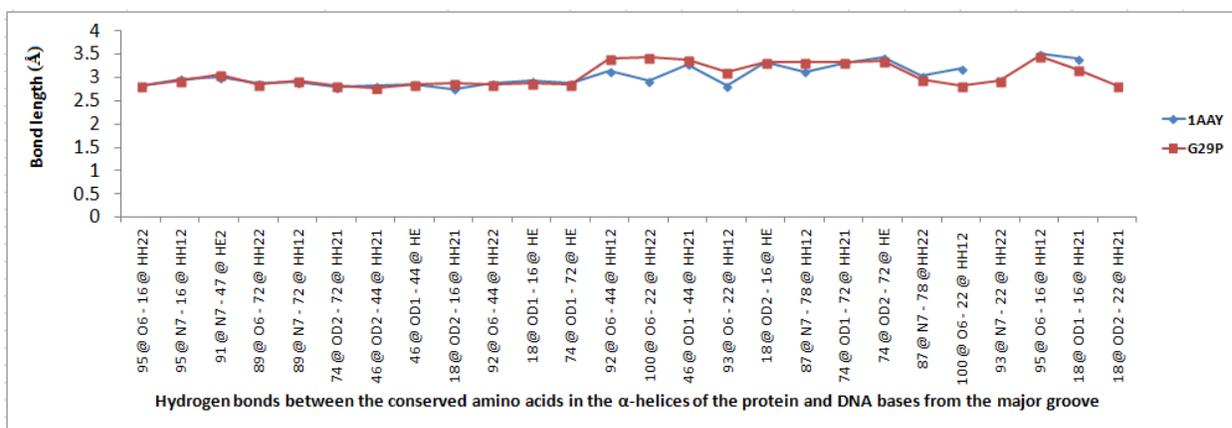


Figure 3.26 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA and G29P.

3.3.4 Analysis of the binding free energy of Q30E-DNA complex:

Glutamine is the third amino acid residue in the linker. It has no important interactions between its side chain and DNA. In addition, no interaction with neighboring amino acid was found⁶. Since the first linker in Zif268 differs from the conserved linker sequence, TGEKP, in having Glutamine (Q) instead of glutamate (E) in this position, the mutant Q30E was created.

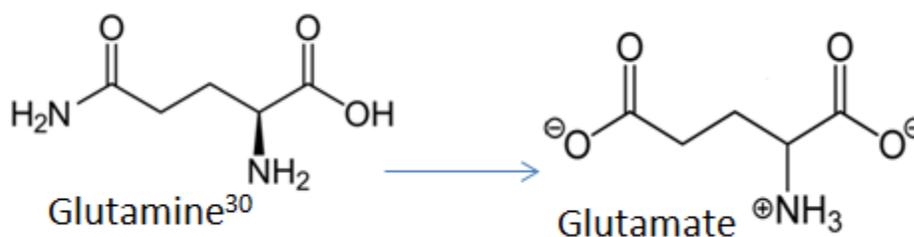


Figure 3.27 Mutating the third amino acid in the first linker (Gln30) to glutamate.

Mutating glutamine to glutamate produced the largest effect on the free energy of binding. Its free energy of binding to target DNA was -204.9 ± 12.14 k which means its -25.5 k cal/mol larger than that of the wild type peptide.

This increase in binding energy may be attributed to the decline in the unfavorable electrostatic contribution to solvation energy which outweighed the

increase in unfavorable electrostatic interactions calculated by molecular dynamics. The increase in favorable van der Waals interactions is also a contribution to this energy.

In the first linker six hydrogen bonds were stabilized. One stable hydrogen bond in 1AAY became unstable in Q30E, and three hydrogen bonds were lost upon mutation. In addition, a new hydrogen bond was formed between one of the carboxylate oxygen's of Glu30 and the backbone amide of the adjacent amino acid Lys31 (Figure 3.28). In the second linker three hydrogen bonds were destabilized and two were stabilized upon mutation. However, no considerable changes on the lengths of these bonds were observed.

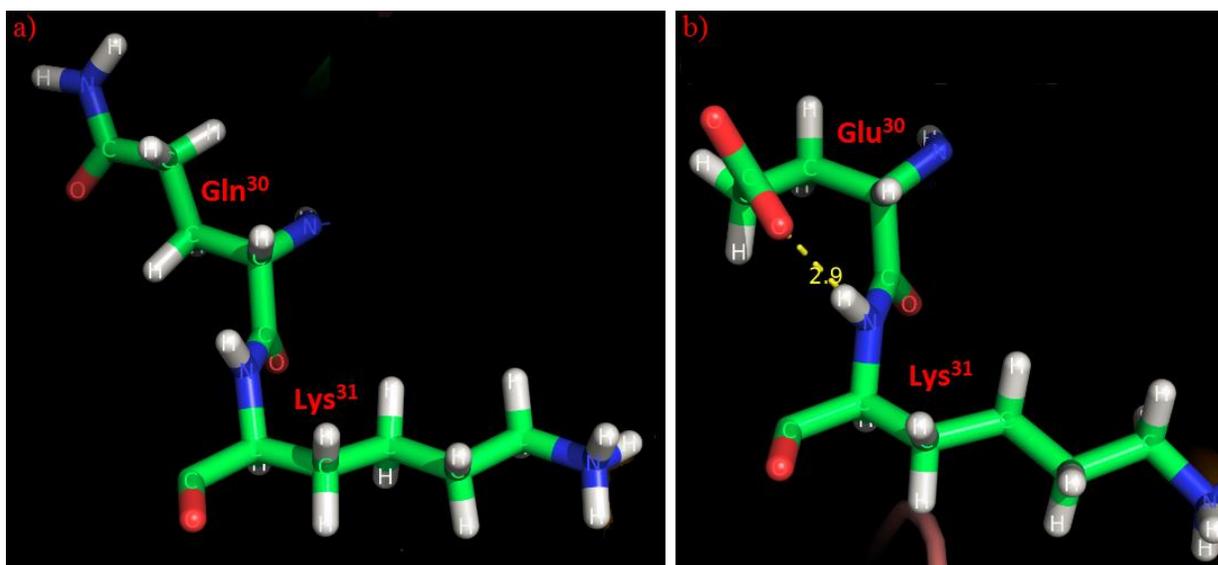


Figure 3.28 The third and fourth amino acids of the first linker in a) Zif268, b) Q30E.

The Hydrogen bonds connecting the conserved amino acids in the α -helices of the Q30E mutant to DNA bases, retained their percent occupancy for the highly stable hydrogen bonds (occupancy ≥ 80), and changed moderately for the rest of hydrogen bonds. (Figure 3.29 to 3.34)

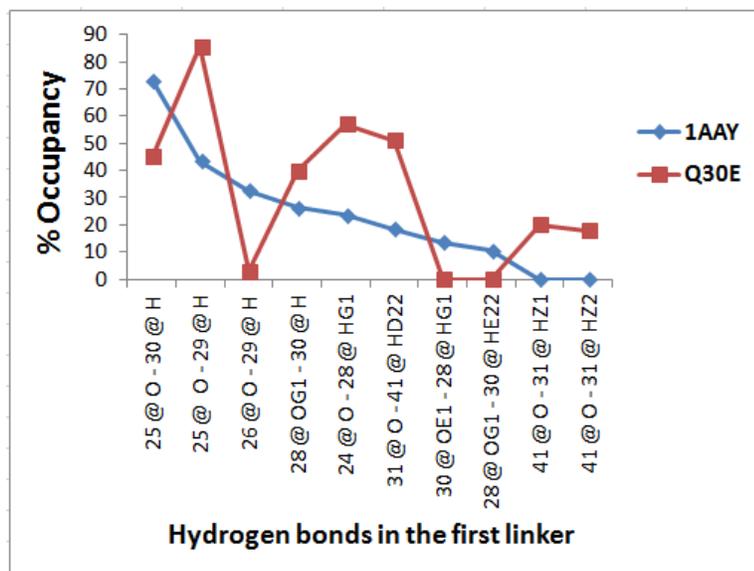
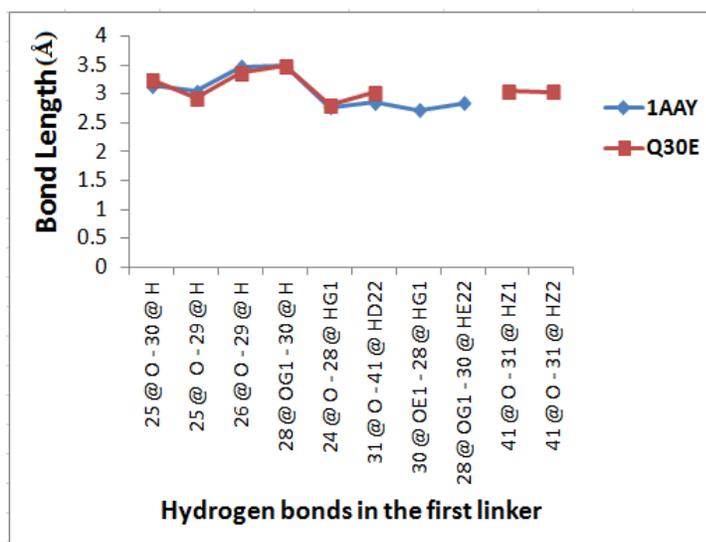


Figure 3.29 Percent occupancy of hydrogen bonds over production runs of 1AAY and Q30E. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.



3.30 Average lengths of hydrogen bonds over production runs of 1AAY and Q30E. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

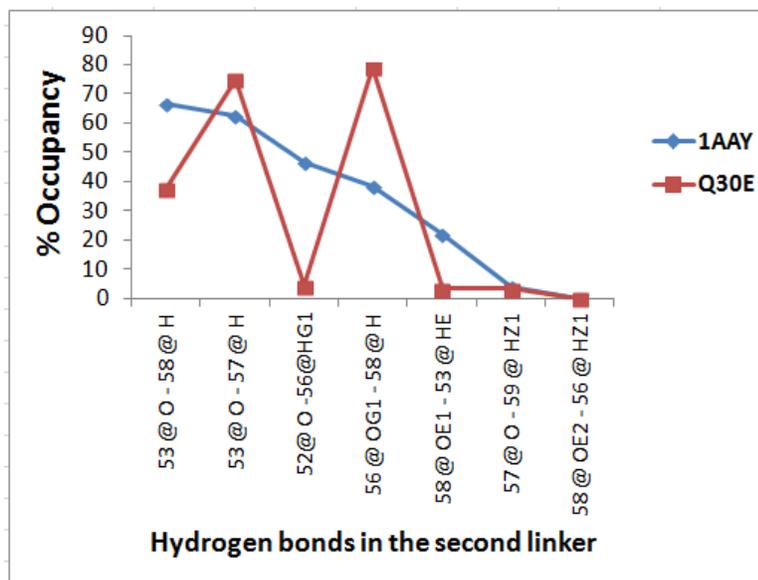


Figure 3.31 Percent occupancy of hydrogen bonds over production runs of 1AAY and Q30E. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

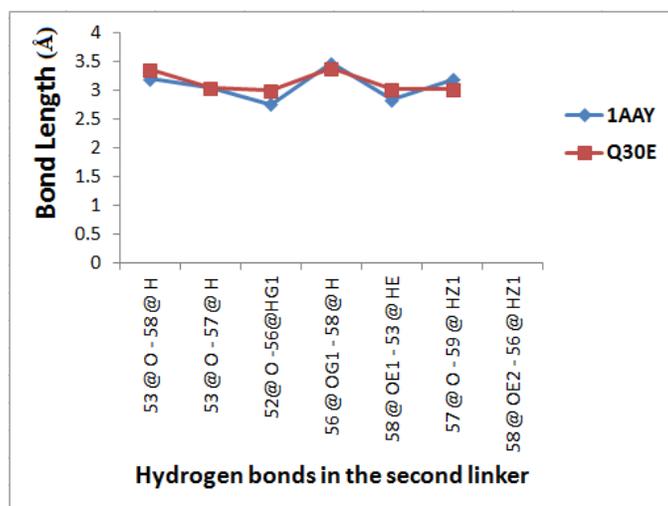


Figure 3.32 Average lengths of hydrogen bonds over production runs of 1AAY and Q30E. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

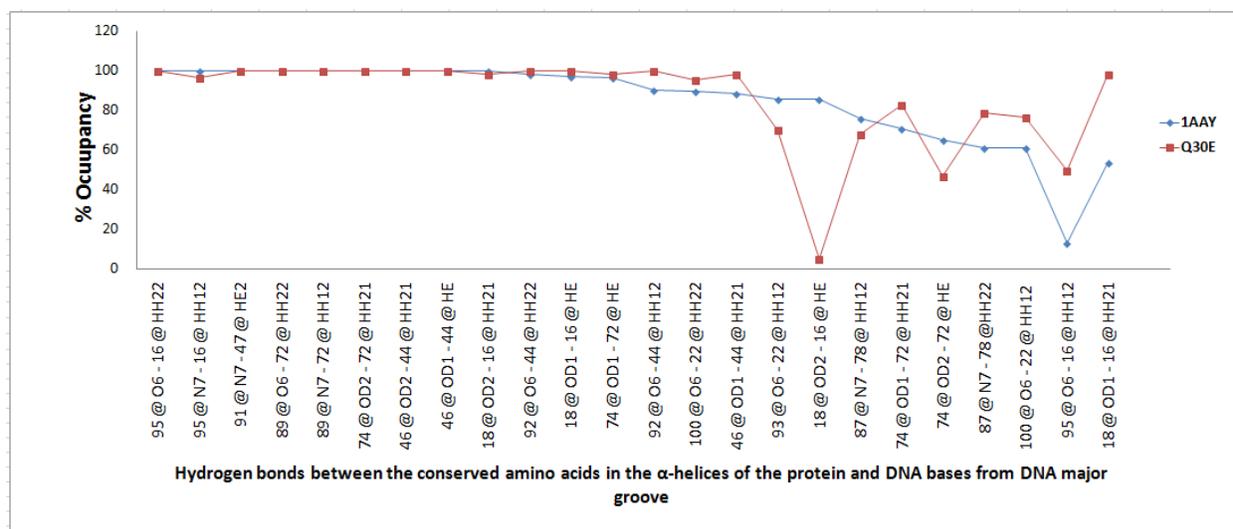


Figure 3.33 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and Q30E.

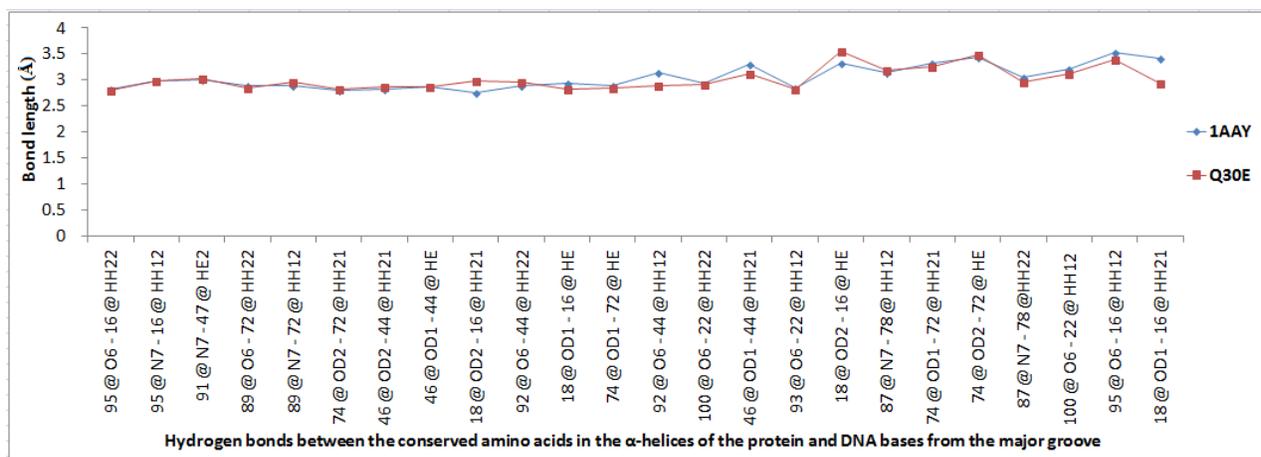


Figure 3.34 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and Q30E.

3.3.5 Analysis of the binding free energy of K31D-DNA complex:

Lysine, the fourth linker residue, was reported to form two water mediated hydrogen bonds with the phosphate group of the thymine base in position 5 in the oligonucleotide²⁰. (Figure 3.35)

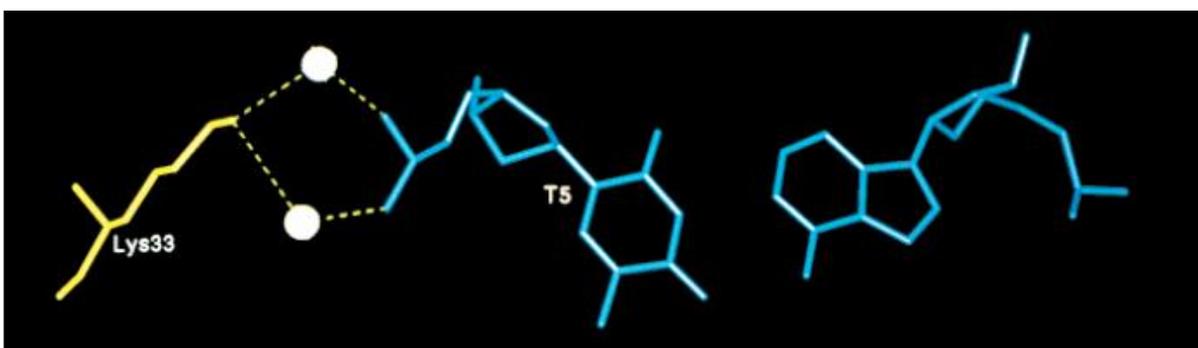


Figure 3.35 Two water-mediated contacts between Lysine33 and the DNA base T5²⁰.

To investigate the effect of this positively charged amino acid on the binding free energy, it was mutated to the negatively charged amino acid Aspartate.



Figure 3.36 Mutating Lysine31 in the first linker to Aspartate.

The binding energy of K31D-DNA complex was found to be -171.4 ± 10.61 kcal/mol. Replacing the positively charged amino acid with a negatively charged one did not seem to have a considerable effect on the value of the binding free energy. However, this mutation caused drastic changes in the three main components of binding free energy. These changes consist of a drop in the unfavorable electrostatic contribution to solvation of about 828 kcal/mol, a reduction in electrostatic interactions of about 822 kcal/mol, and a reduction in the favorable van der Waals interactions of about 12.5 kcal/mol. Eventually all these valuable changes almost canceled each other leaving a net slight effect on the binding free energy.

All changes can be explained as a consequence of two factors. The first is the introduction of the negative charge to this site. The second is losing the water mediated interactions which existed between the positively charged side chain of Arginine and the negatively charged phosphate on Thymine number 5 base²⁰. The introduction of the negatively charged amino acid caused unfavorable electrostatic interaction with the negatively charged phosphate group in DNA backbone. The distances between the oxygen in Aspartate and the two oxygen atoms in the

phosphate group are shown in Figure 3.37.

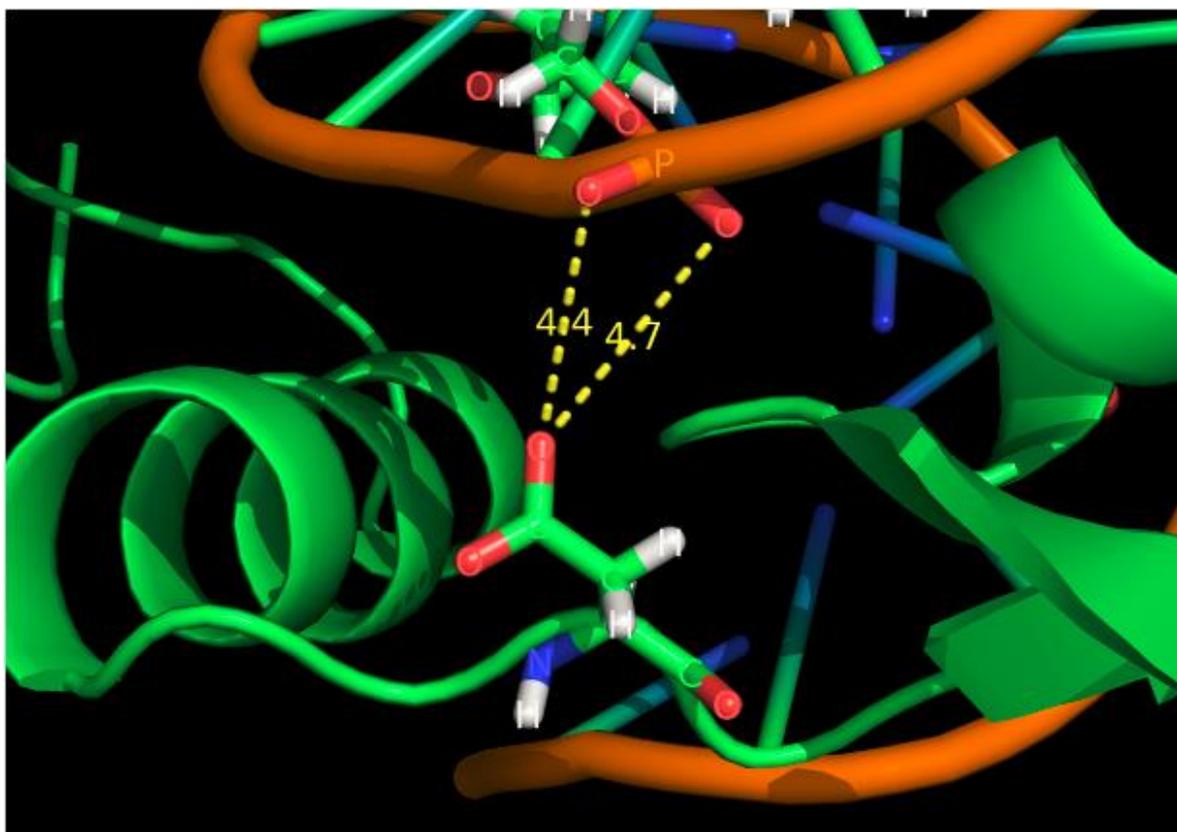


Figure 3.37 The distances in Å between the oxygen in Aspartate and the two oxygen atoms in the phosphate group

Mutating Lysine to Aspartate had an effect on the hydrogen bonds that include amino acids in this linker. Three hydrogen bonds were significantly stabilized, while two unstable hydrogen bonds in 1AAY were lost in K31D, and three remained almost unaffected. In the second linker, one stable hydrogen bond in 1AAY was significantly destabilized in K31D, and a slight stabilization of one

hydrogen bond was observed. Bond lengths involving amino acids from the two linkers were nearly the same.

The hydrogen bonds between conserved amino acids in the α -helices of the protein and DNA bases in the major groove were influenced by this mutation. The highly stable bonds (i.e., those with percent occupancy ≥ 80) retained their status with the exception of four hydrogen bonds; three were destabilized and one was lost. The rest of the hydrogen bonds in this domain changed as follows: Stabilization of four bonds, and destabilization of one bond, and loss of another. Arg22 in 1AAY-DNA complex was connected by four hydrogen bonds with DNA bases. Two of these bonds were with G8 and two with G9. However, the two hydrogen bonds connecting Arg22 with G9 disappeared upon mutation. This was mainly due to a conformational change in Arg22 side chain in K31D-DNA complex.

The average length of hydrogen bonds changed considerably for three bonds which exist between Arg72 and Asp74. Among these three bonds two were shortened. This agrees with the observed stabilization of the two bonds. The one that was elongated is consistent with destabilization of this bond. (Figure 3.38 to 3.43)

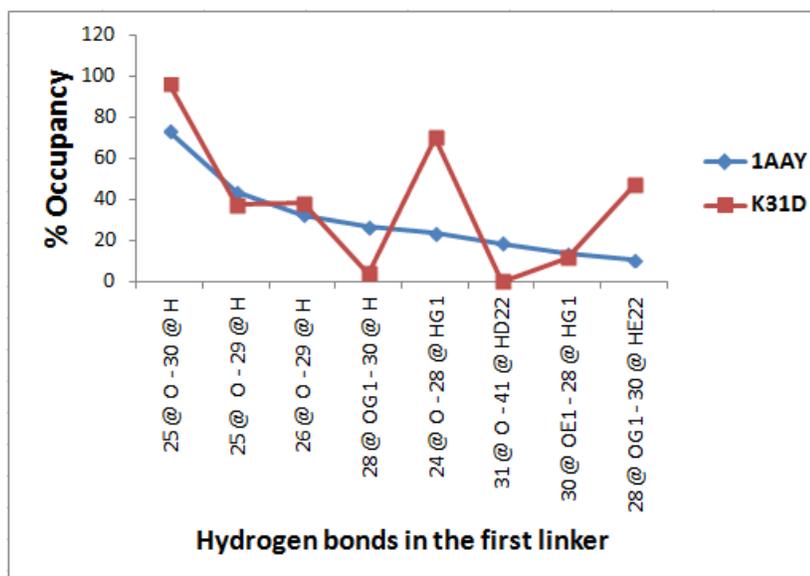


Figure 3.38 Percent occupancy of hydrogen bonds over production runs of 1AAY and K31D. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

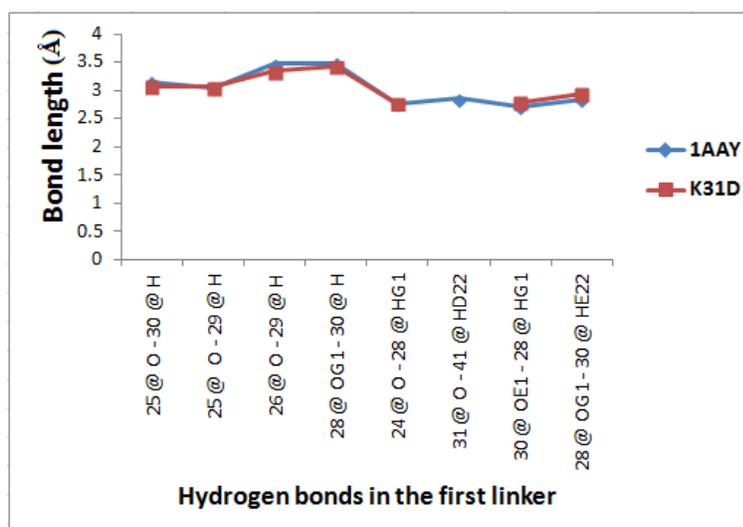


Figure 3.39 Average lengths of hydrogen bonds over production runs of 1AAY and K31D. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

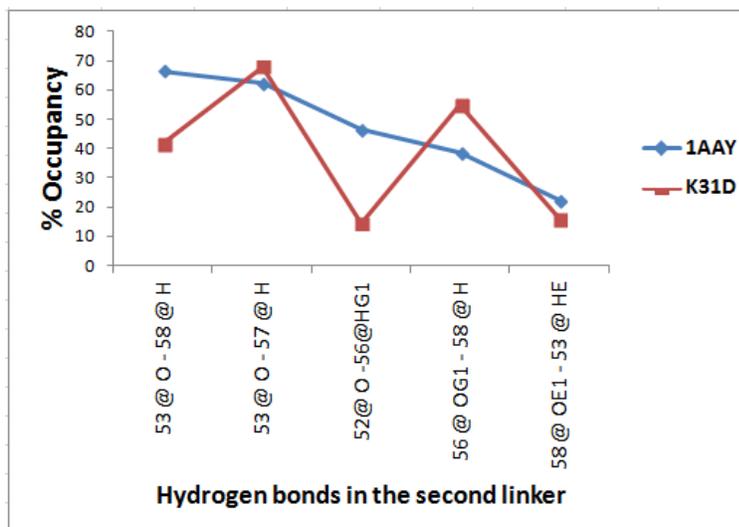


Figure 3.40 Percent occupancy of hydrogen bonds over production runs of 1AAY and K31D. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

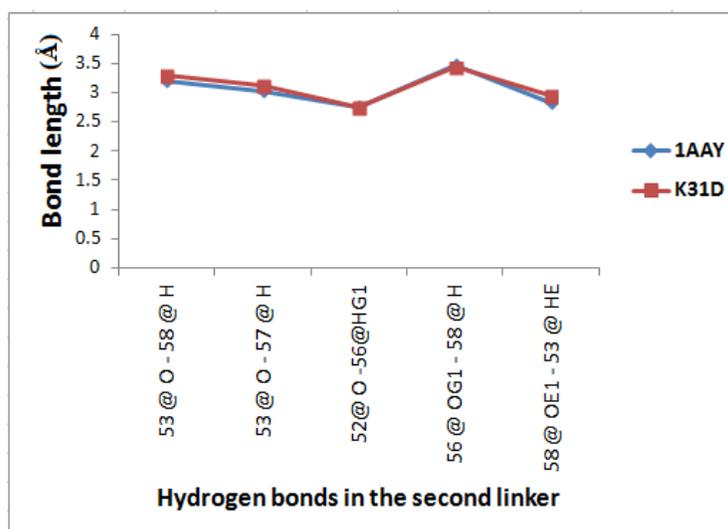


Figure 3.41 Average lengths of hydrogen bonds over production runs of 1AAY and K31D. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

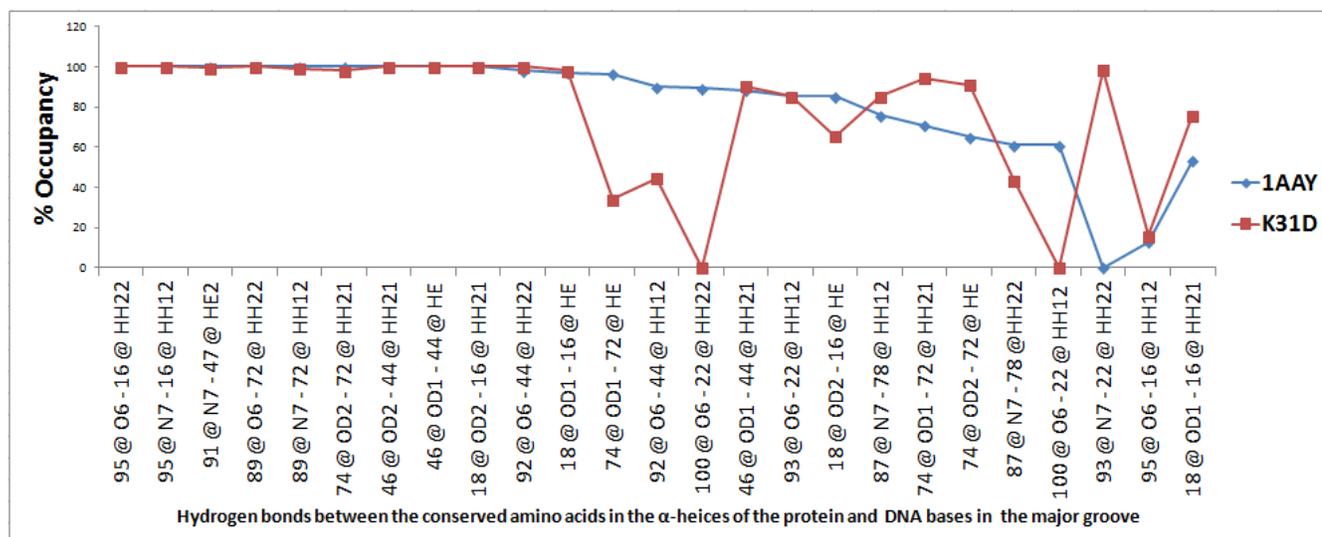


Figure 3.42 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and K31D.

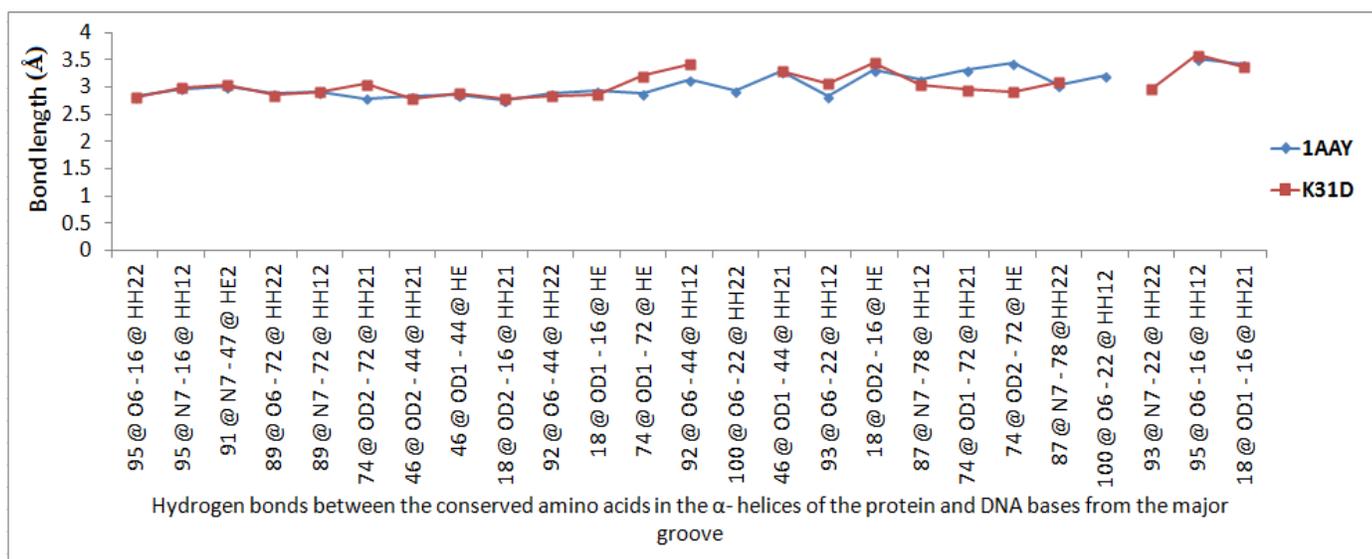


Figure 3.43 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and K31D.

3.3.6 Analysis of the binding free energy of P32G-DNA complex.

The last amino acid in the first linker is Proline, which is a constrained amino acid. No contacts were reported between this Proline and any other part of the complex, suggesting that its conservation may be due to its steric structure^{6,20,28}. To test this hypothesis, Proline was mutated to the small amino acid, Glycine.

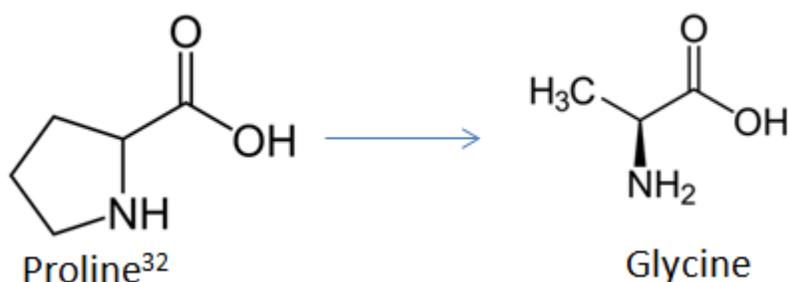


Figure 3.44 Mutating Proline³² in the first linker to Glycine.

The binding energy of P32G-DNA complex was estimated to be -167.97 ± 9.62 kcal/mol. The different energy components did not experience any considerable change. However, the slight decrease in binding energy suggests a potential preference toward having a constrained amino acid in the terminus of the first linker.

Hydrogen bonds in the first linker were subjected to destabilization of three bonds, stabilization of two other bonds, and a loss of three unstable hydrogen

bonds. Similarly, in the second linker one hydrogen bond was destabilized, one stabilized and two hydrogen bonds were lost. Bond lengths did not show any noteworthy changes except for an increase in the length of the hydrogen bond between Ile25 and Gln30 by 0.3 Å corresponding to the calculated decrease in the percent occupancy of this bond of 18%.

The hydrogen bonds between conserved amino acids in the α -helices of the protein and DNA bases in the major groove were affected by this mutation as follows: The highly stable bonds (occupancy $\geq 80\%$) retained their status, with the exception of two bonds that were significantly destabilized. The moderately stable hydrogen bonds showed stabilization of two bonds and destabilization of two other bonds. All of the unstable hydrogen bonds in 1AAY became more stable in P32G. Also a new hydrogen bond between Guanine 93 and Arginine 22 was formed after the mutation. Alteration in lengths of hydrogen bonds was limited to a reduction by 0.3 Å in the hydrogen bond connecting Arg18 and Asp16. This reduction agrees with a 33% increase in percent occupancy of this bond. (Figure 3.45 to 3.50)

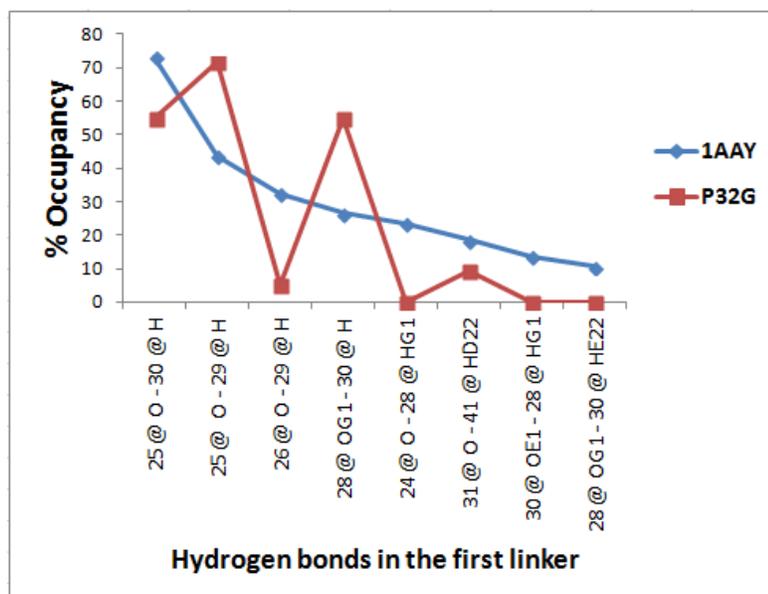


Figure 3.45 Percent occupancy of hydrogen bonds over production runs of 1AAy and P32G. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

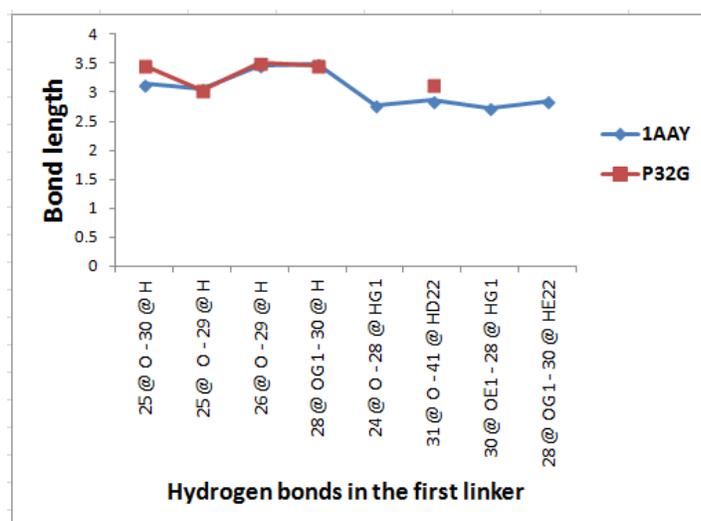


Figure 3.46 Average lengths of hydrogen bonds over production runs of 1AAy and P32G. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

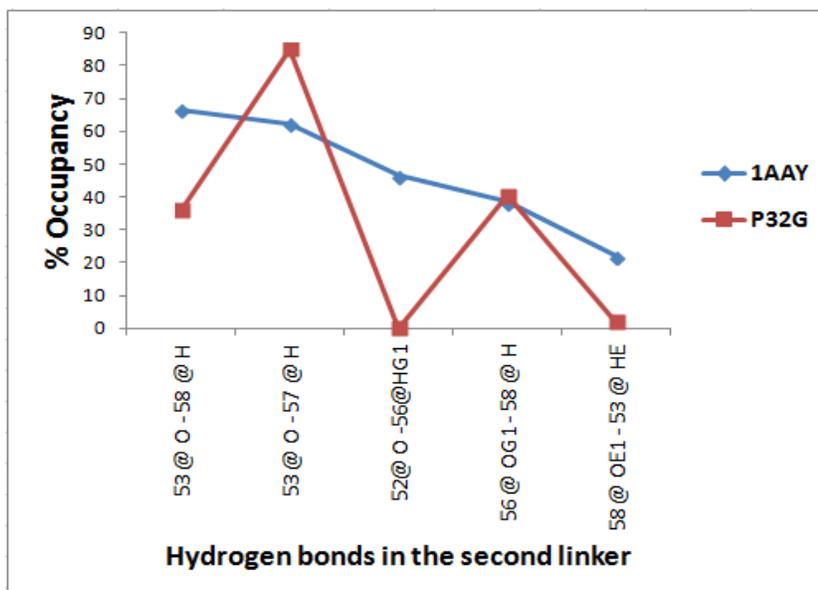


Figure 3.47 Percent occupancy of hydrogen bonds over production runs of 1AAY and P32G. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

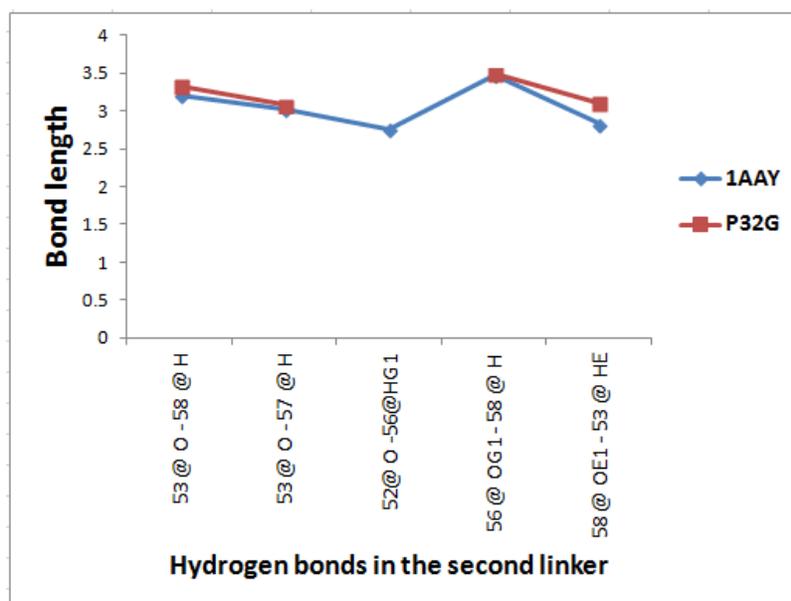


Figure 3.48 Average lengths of hydrogen bonds over production runs of 1AAY and P32G. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

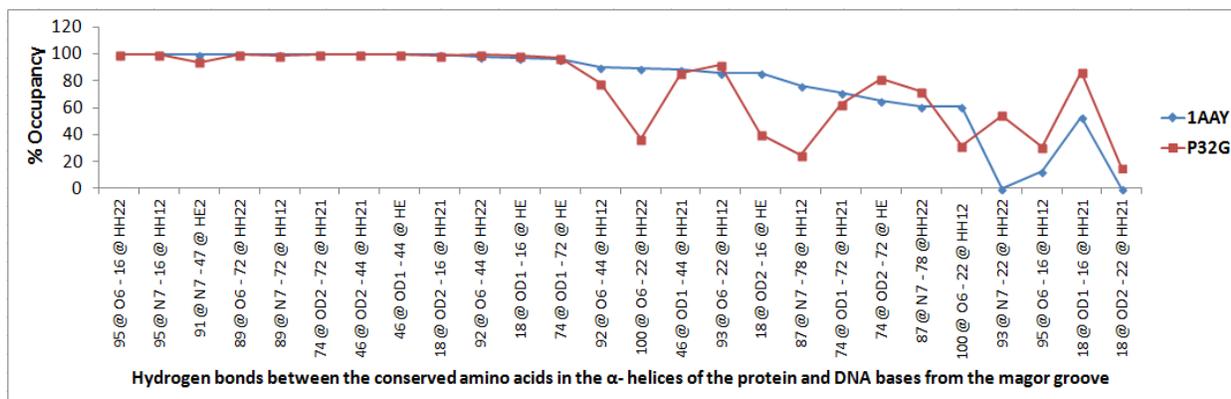


Figure 3.49 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major, over production runs of 1AA Y and P32G.

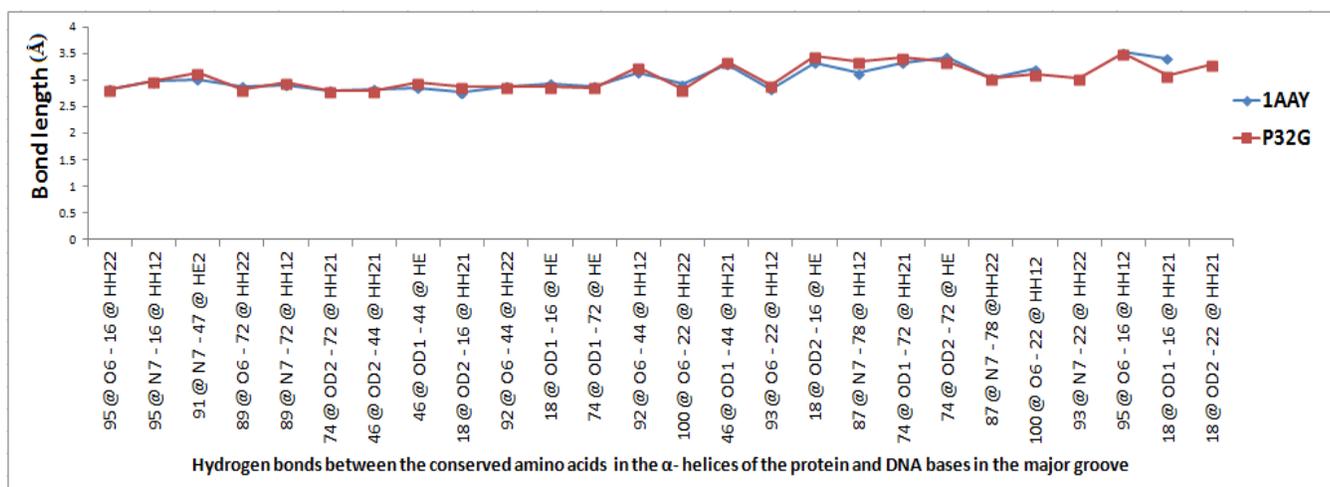


Figure 3.50 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and P32G.

3.3.7 Analysis of the binding free energy of T56Y-DNA complex.

The first amino acid in the second linker Threonine (Thr⁵⁶) was mutated to Tyrosine to test the effect of having a polar amino acid, with larger steric hindrance, on the binding energy.

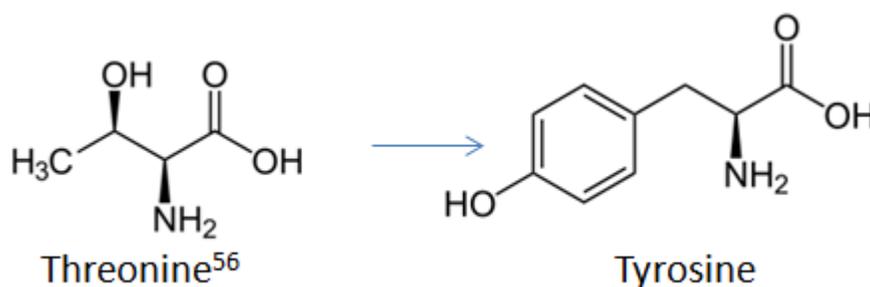


Figure 3.51 Mutating Threonine⁵⁶ in the second linker to Tyrosine.

The binding free energy of T56Y-DNA complex was calculated and gave a value of -158.65 ± 11.88 kcal/mol. Mutating Threonine to Tyrosine had a significant effect on the free energy of binding and produced the mutant with the lowest binding free energy to the target DNA among the ten mutants created in this work. This reduction in binding energy may be explained as being due to the reduction in the favorable electrostatic interactions calculated by molecular dynamics. This lowering in electrostatic energy exceeded the accompanied reduction in the unfavorable electrostatic contribution to solvation energy. These

changes are possibly a consequence of the huge size of the tyrosine side chain, compared to that of Threonine. This is due to the fact that the amino acid is located in the α -helix of finger 2, which subsequently led to loss of the hydrogen bond between Thr⁵⁶ and the back bone amide of the linker third amino acid.

The hydrogen bond analysis of T56Y-DNA complex showed that, four hydrogen bonds were noticeably stabilized, two destabilized, and one hydrogen bond was lost in the first linker. Three hydrogen bonds in the second linker were lost and two weren't significantly affected. However, no measurable changes in hydrogen bond lengths were observed. Among those lost hydrogen bonds, the hydrogen bond connecting Thr56 and Glu58, similar to the bond in the zinc finger protein TFIIIA⁶ was found to be responsible for DNA-induced α -helix capping of the second finger in TFIIIA and was suggested to stabilize the TFIIIA-DNA complex (Figure 3.52). However, the loss of this bond is expected because Thr56 was replaced with Tyrosine. Although the side chain of Tyrosine contains a terminal hydroxyl group, it cannot form a hydrogen bond with Glutamate backbone amide. This is due to the bulkiness of the Tyrosine side chain, which consequently directs this side chain away from the backbone amide of Gln. (Figure 3.53).

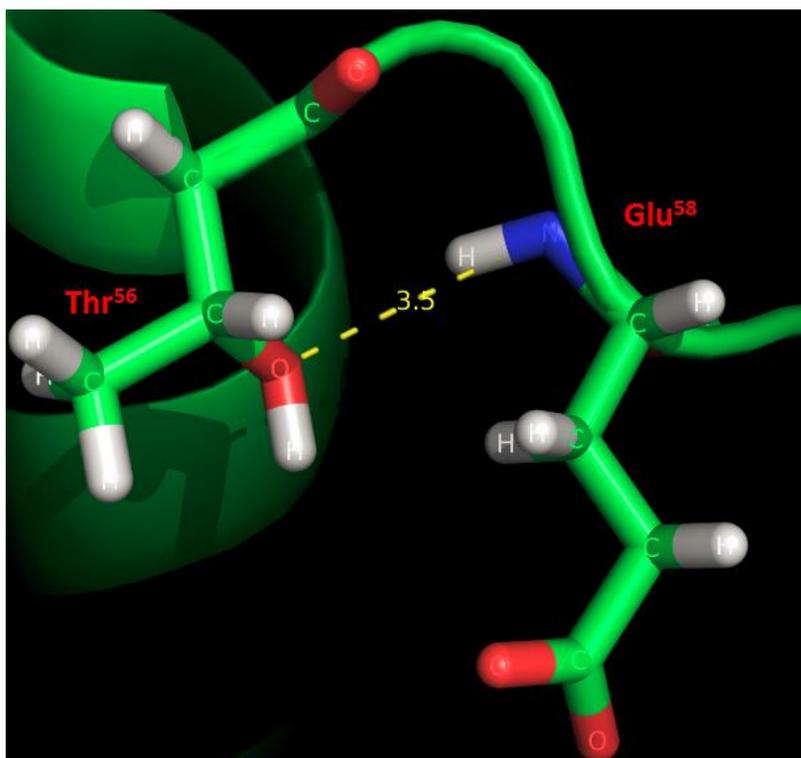


Figure 3.52 Hydrogen bond between Thr56 and Glu58 in 1AAY.

The hydrogen bonds between conserved amino acids in the α -helices of the protein and DNA bases in the major groove were affected by this mutation as follows: Five stable hydrogen bonds in 1AAY (occupancy $\geq 60\%$) were significantly destabilized in T56Y. One hydrogen bond connecting G 9 with Arg22 in 1AAY was lost in T56Y. A hydrogen bond between G at base pair 8 and Arg22 was formed, and one hydrogen bond connecting Asp18 and Arg16 was stabilized to a great extent. This stabilized hydrogen bond in T56Y mutant experienced a 0.63 Å reduction in its length. Among the five destabilized

hydrogen bonds in T56Y, only two had notable increase in their length of approximately 0.4-0.5 Å. (Figure 3.54 to 3.59)

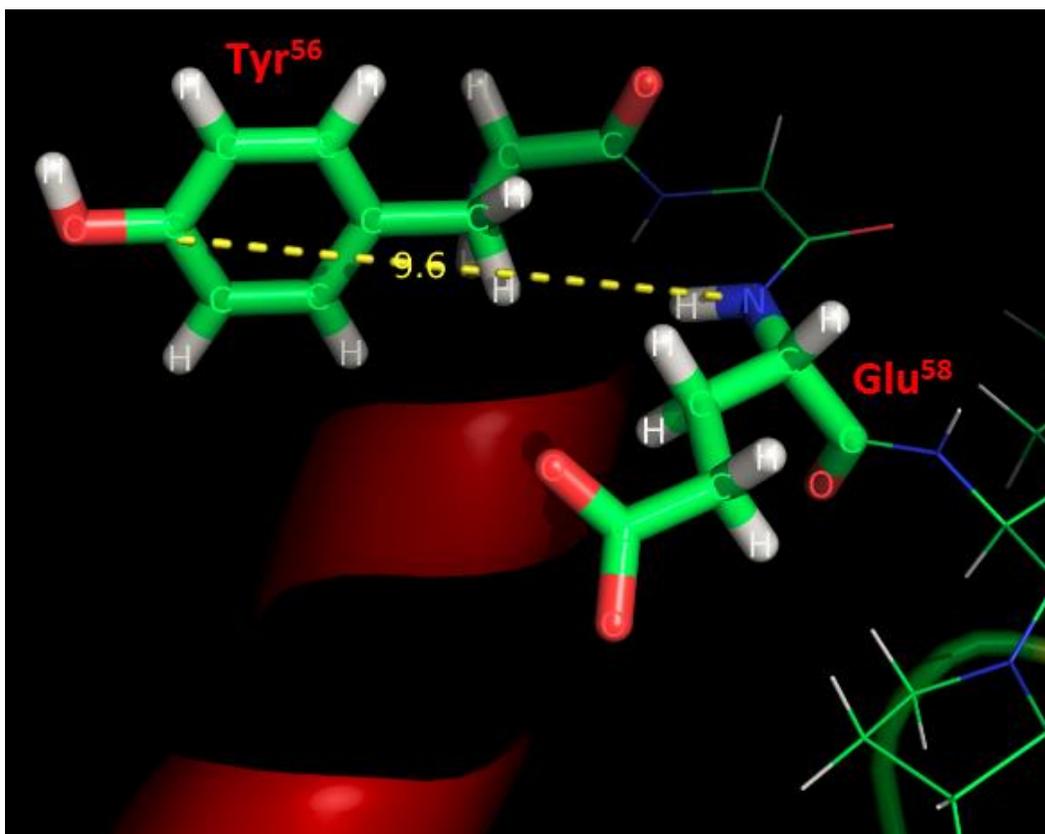


Figure 3.53 The distance between Tyr56 and Glu58 in T56Y mutant.

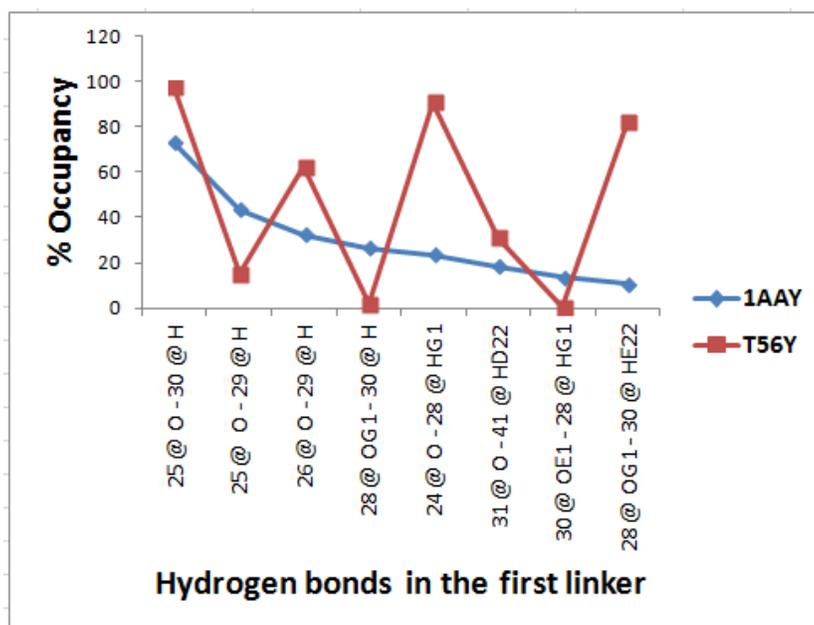


Figure 3.54 Percent occupancy of hydrogen bonds over production runs of 1AAY and T56Y. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

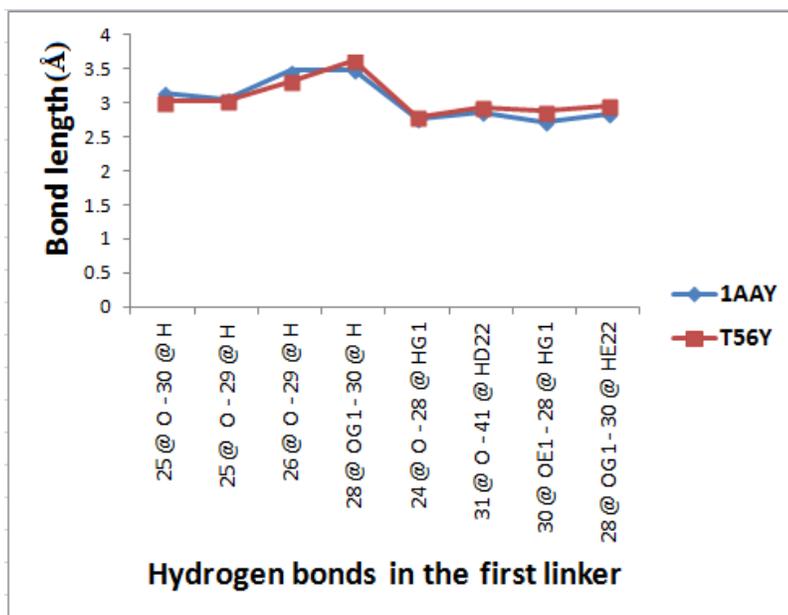


Figure 3.55 Average lengths of hydrogen bonds over production runs of 1AAY and T56Y. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

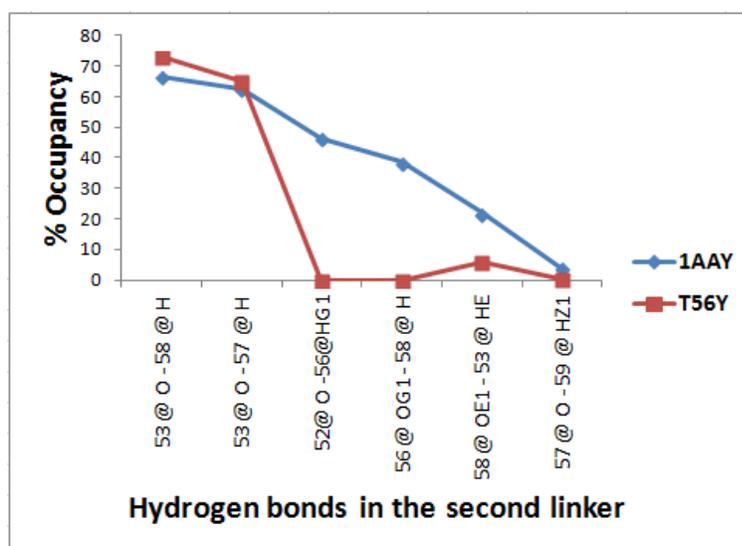


Figure 3.56

Percent

occupancy of hydrogen bonds over production runs of 1AAY and T56Y. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

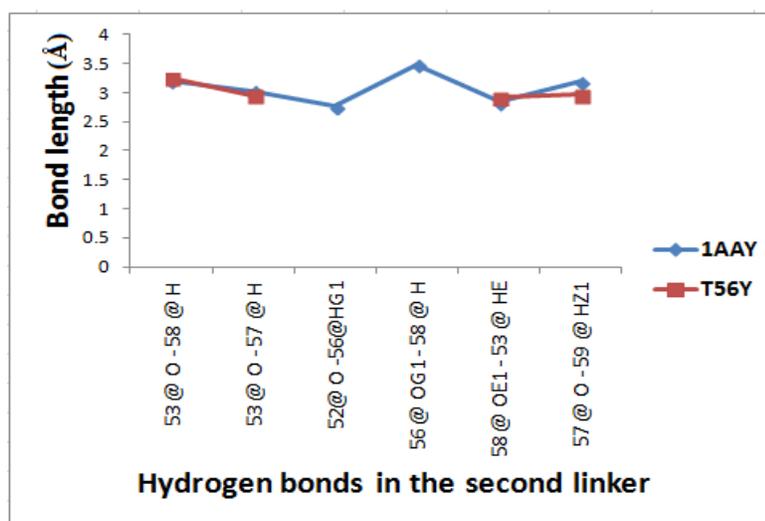


Figure 3.57 Average lengths of hydrogen bonds over production runs of 1AAY and T56Y. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

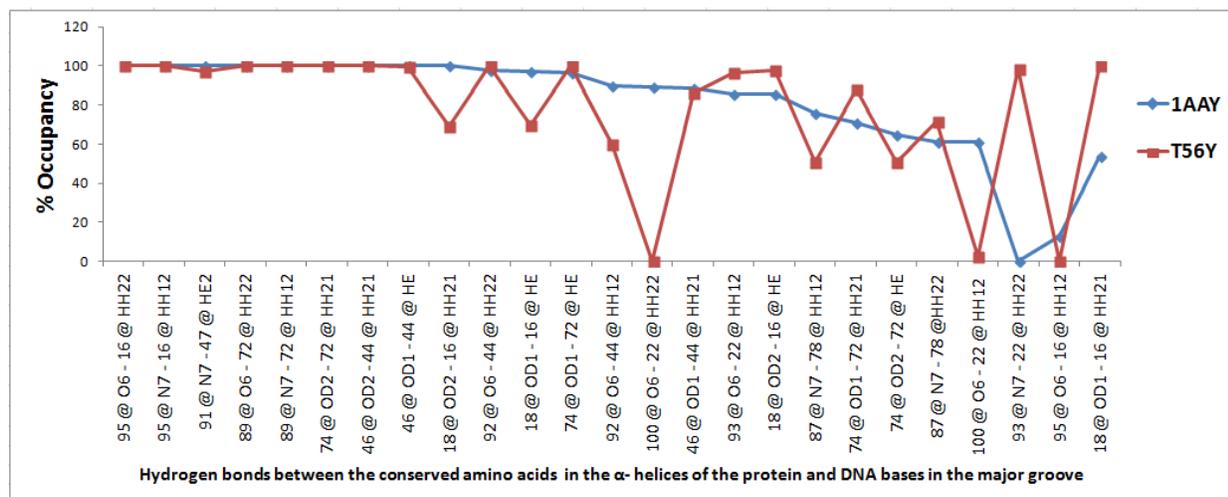


Figure 3.58 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and T56Y.

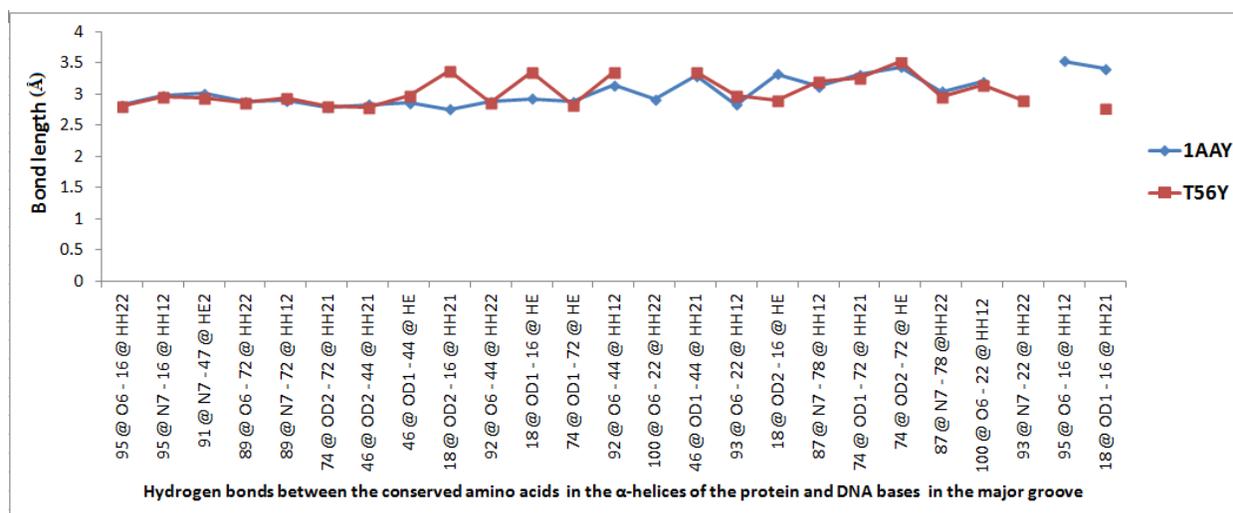


Figure 3.59 Average Lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAY and T56Y.

3.3.8 Analysis of the binding free energy of G57V-DNA complex:

Glycine, the second amino acid in the second linker, was mutated to Valine, which is more hydrophobic and larger than Glycine. Glycine⁵⁷ plays a vital role in forming Gly C-cap via the hydrogen bond between the backbone amide of Gly⁵⁷ and the carbonyl oxygen of Arg⁵³^{6,36}. This mutant is expected to give information about the importance of this C-capping on the binding energy, and to examine the impact of hydrophobicity and steric hindrance in this site on the binding affinity.

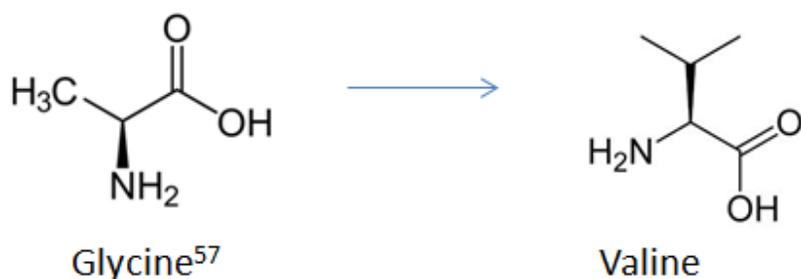


Figure 3.60 Mutating Glycine⁵⁷ in the second linker to Valine.

The binding energy of G57V-DNA complex was calculated, it was found to be -170 ± 9.8 kcal/mol. This mutant gave a value of total binding energy close to that of the wild type binding to DNA. This mutation caused an increase in the unfavorable electrostatic free energy of about 92 kcal/mol. In addition, it caused a reduction in the unfavorable electrostatic contribution to solvation energy of about 84 kcal/mol. One possible explanation of these changes is the increase in

hydrophobicity and steric hindrance as a result of replacing a methyl group with an isopropyl group. Nevertheless, increasing the hydrophobic effect by the isopropyl side chain in addition to the loss of Gly C-cap upon mutation did not have a considerable effect on the binding free energy.

Hydrogen bonds in the first linker were moderately affected by this mutation. Two of them were stabilized and one was destabilized. Among the detectable five hydrogen bonds in the second linker, three were destabilized and two were lost. Nonetheless, no notable changes in hydrogen bond lengths were detected.

The hydrogen bonds between conserved amino acids in the α -helices of the protein and DNA bases in the major groove were affected by this mutation as follows: most highly stable hydrogen bonds in 1AAY (occupancy $\geq 80\%$) maintained their stability in G57V mutant except for three hydrogen bonds. Two of them were destabilized, and one hydrogen bond connecting G9 with Arg22 in 1AAY was lost in G57V. One hydrogen bond between G8 and Arg22 was formed, and one hydrogen bond connecting Asp18 and Arg16 was stabilized to a great extent. Among all the influenced hydrogen bonds in G57V mutant, only the destabilized bond between G7 with Arg44 had a notable increase in its length of 0.38 Å. (Figure 3.61 to 3.66)

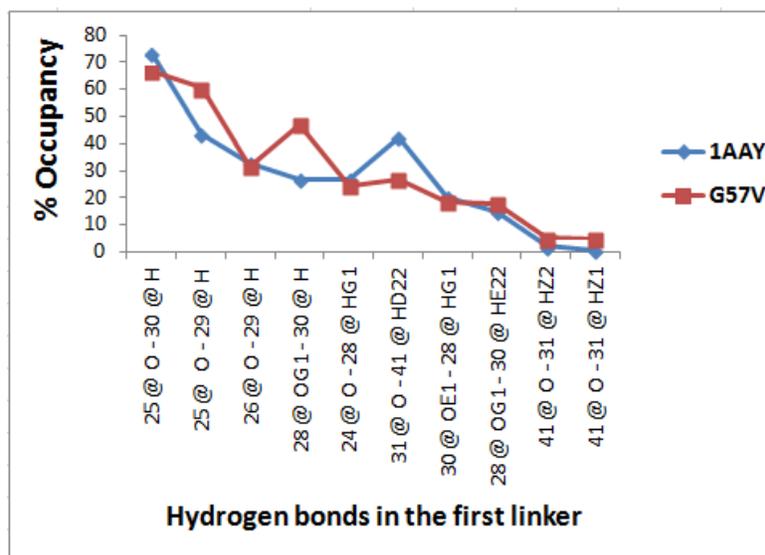


Figure 3.61 Percent occupancy of hydrogen bonds over production runs of 1AAY and G57V. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

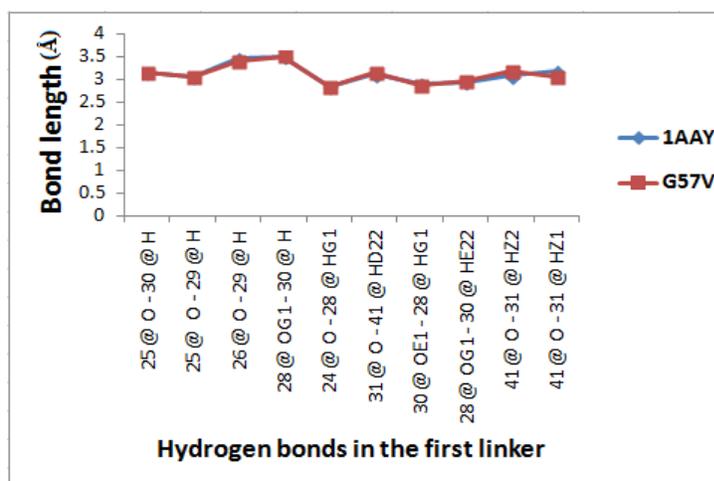


Figure 3.62 Average lengths of hydrogen bonds over production runs of 1AAY and G57V. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

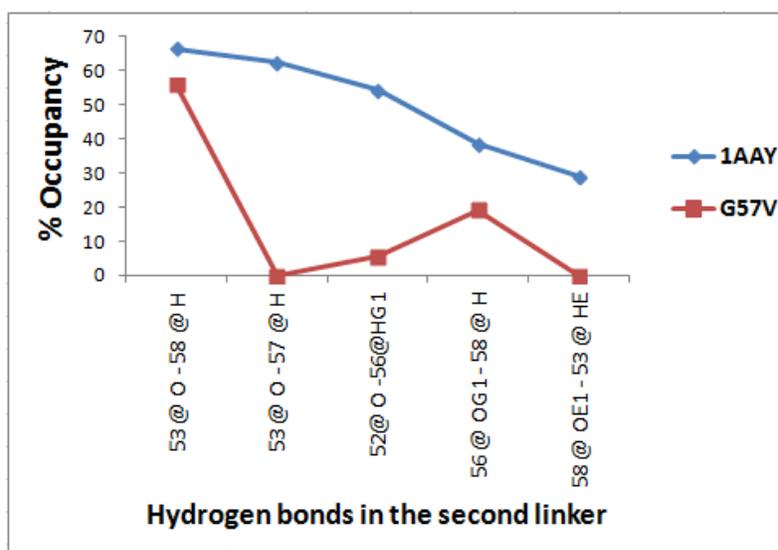


Figure 3.63 Percent occupancy of hydrogen bonds over production runs of 1AAY and G57V. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

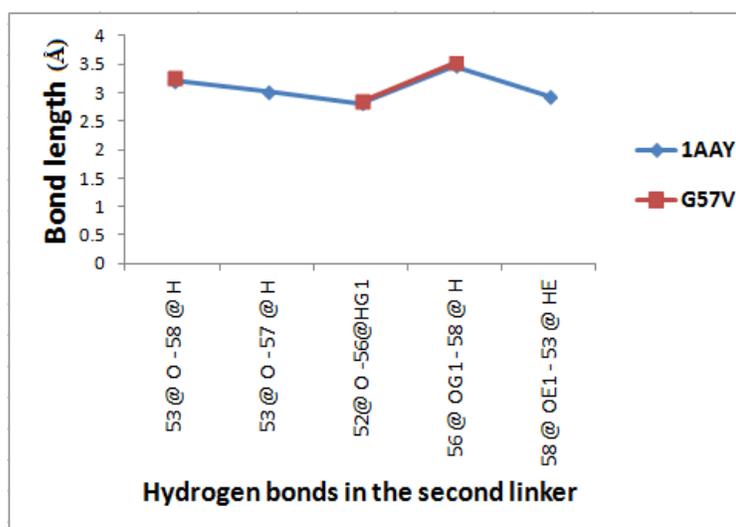


Figure 3.64 Average lengths of hydrogen bonds over production runs of 1AAY and G57V. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

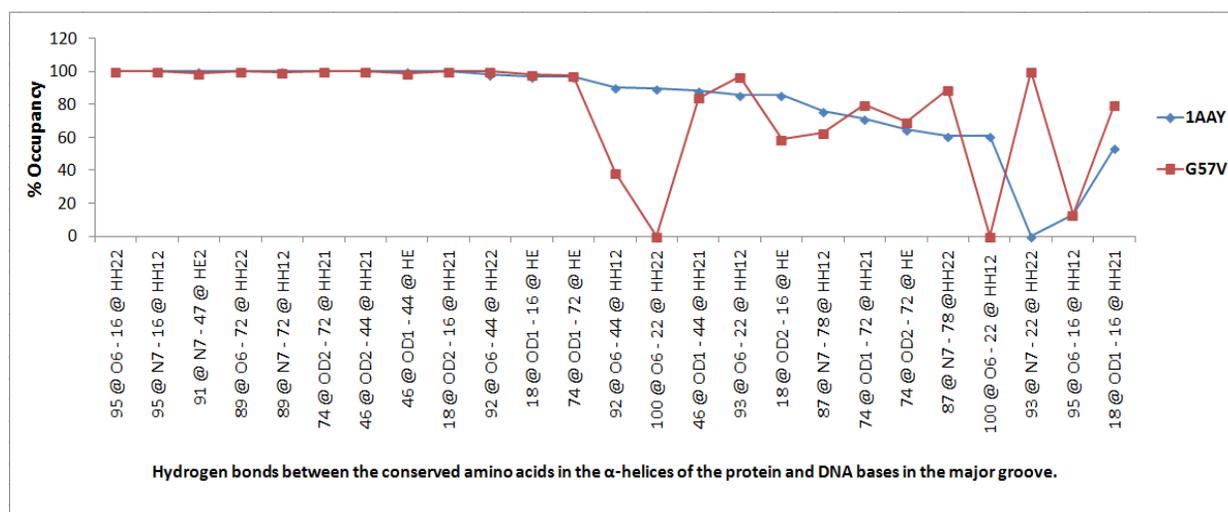


Figure 3.65 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAV and G57V.

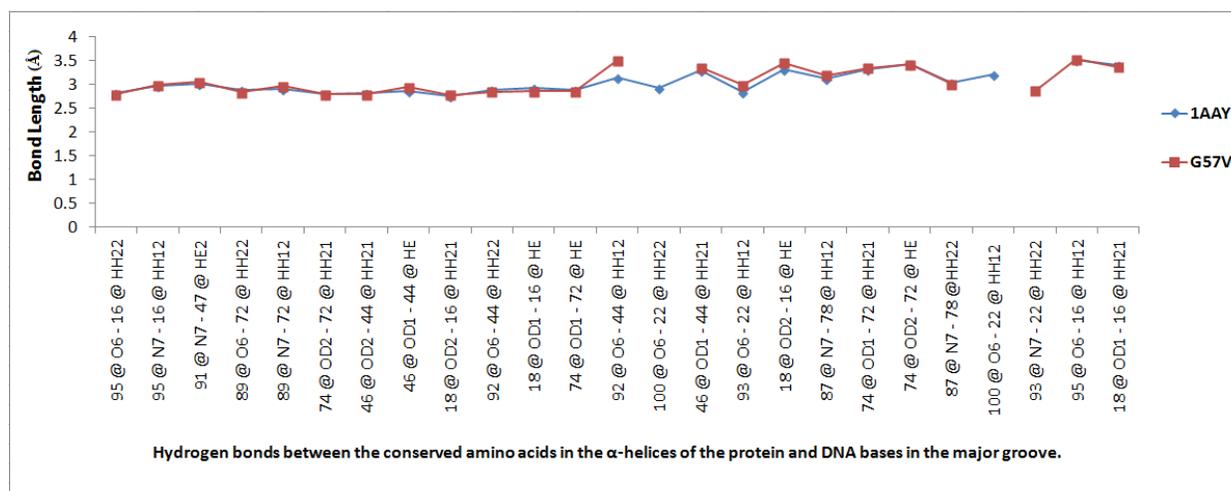


Figure 3.66 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AAV and G57V.

3.3.9 Analysis of the binding free energy of E58Q-DNA complex.

The Glutamate, the third amino acid in this linker, was mutated to Glutamine. This mutation was created to test the possibility of having two identical linkers would be more favorable, and its effect on binding.

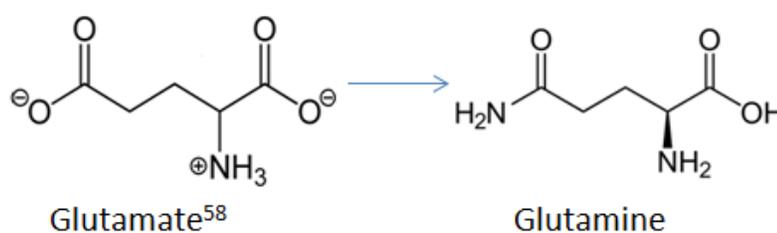


Figure 3.67 Mutating Glutamate⁵⁸ in the second linker to Glutamine.

The binding free energy of E58Q-DNA complex was calculated. The value was found to be -197.55 ± 9.89 kcal/mol. Mutating Glutamate to Glutamine produced a larger binding free energy by namely -18 kcal/mol. This increase in binding free energy may be attributed to the decline in unfavorable electrostatic interactions calculated by molecular dynamics which overcame the increment in the unfavorable electrostatic contribution to solvation free energy, and the decrease in favorable van der Waals interactions. These changes are likely a consequence of replacing the negative charge on Glutamate with the neutral charge on Glutamine.

Hydrogen bond analysis showed that four hydrogen bonds that involve amino

acids in the first linker experienced a considerable increase in their stability, whereas three were destabilized. Among the five hydrogen bonds detected in the second linker, three were notably destabilized, and two remained unaffected. Only one hydrogen bond connecting Asp31 and Asn41 had a considerable change in its bond length of 0.3 Å. This reduction in bond length corresponds to the observed stabilization of this bond in the Q58E mutant. Analysis results of the hydrogen bonds between conserved amino acids in the α -helices of the protein, and DNA bases in the major groove, indicate that the stability of seven highly stable hydrogen bonds in 1AAY were altered in the E58Q mutant, Six of these were notably destabilized, while only one was significantly stabilized. There are moderately stable and unstable hydrogen bonds in 1AAY. Four of these bonds experienced a reduction in percent occupancy, two of which had a reduction in percent occupancy of approximately 55-58 %. In addition to these changes, the bond connecting the two conserved amino acids Arg16 and Asp18 was greatly stabilized.

Six bonds suffered significant changes in their bond lengths of 0.4-0.6 Å. All of these changes agree with the detected change in percent occupancy of these bonds. Formation of a stable hydrogen bond between G8 and Arg44 was detected in the E58Q-DNA complex. (Figure 3.68 to 3.73)

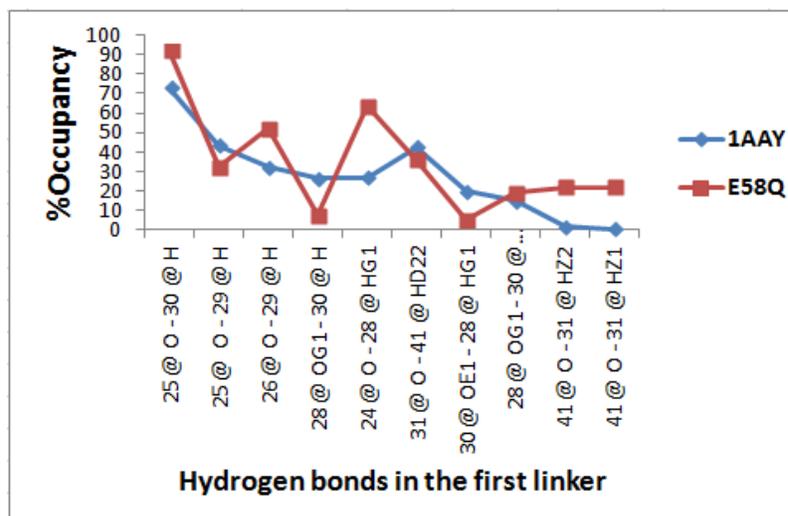


Figure 3.68 Percent occupancy of hydrogen bonds over production runs of 1AAY and E58Q. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

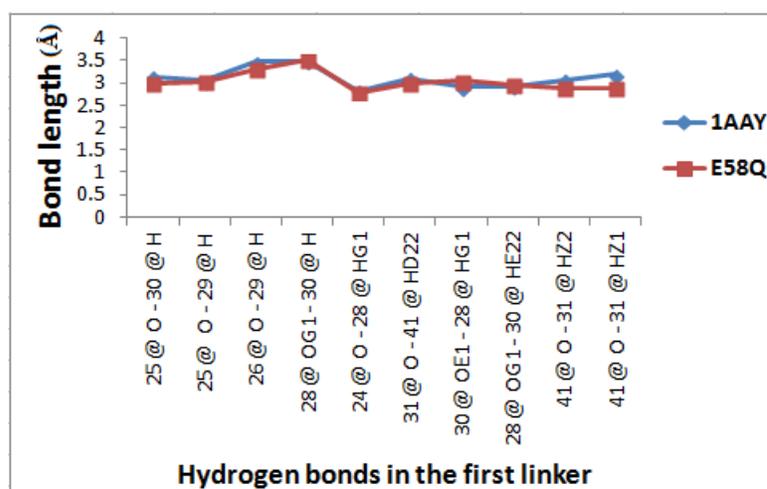


Figure 3.69 Average lengths of hydrogen bonds over production runs of 1AAY and E58Q. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

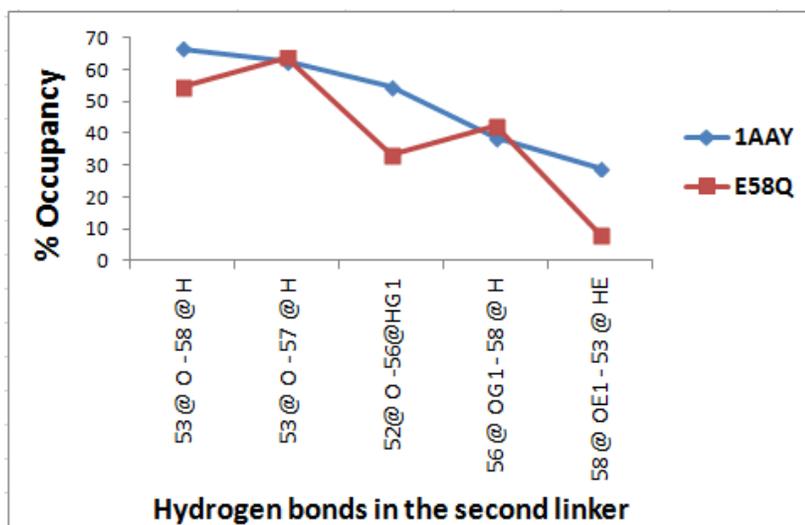


Figure 3.70 Percent occupancy of hydrogen bonds over production runs of 1AAy and E58Q. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

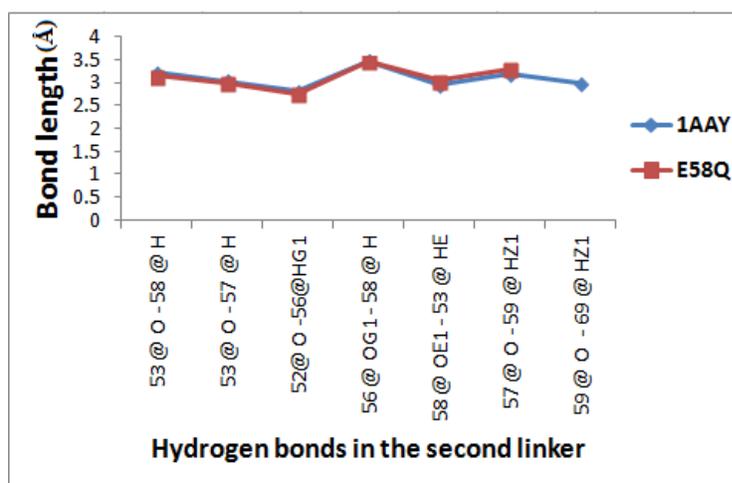


Figure 3.71 Average lengths of hydrogen bonds over production runs of 1AAy and E58Q. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

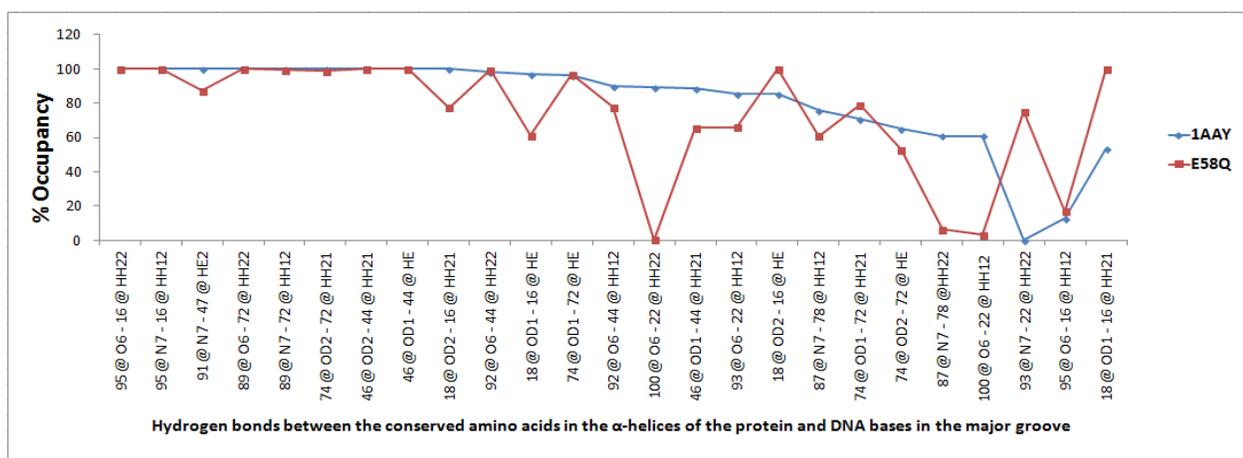


Figure 3.72 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and E58Q.

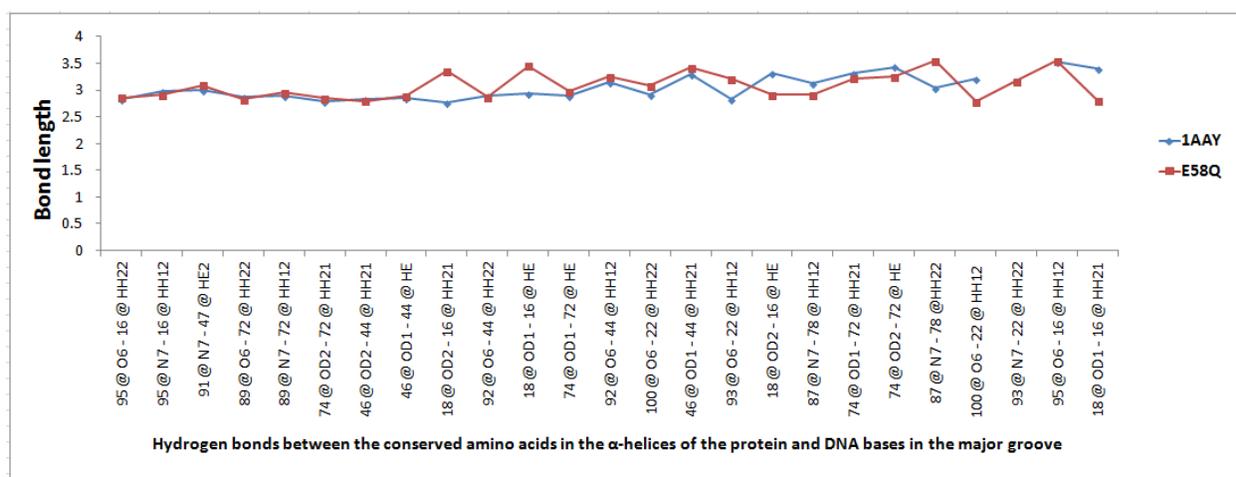


Figure 3.73 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and E58Q.

3.3.10 Analysis of the binding free energy of K59P-DNA complex.

Lysine⁵⁹ forms two water-mediated interactions with the phosphate group of the second base in the oligonucleotide²⁰. Lysine⁵⁹ was mutated to Proline, which is unlikely to perform water mediated contacts, and is more constrained than lysine.

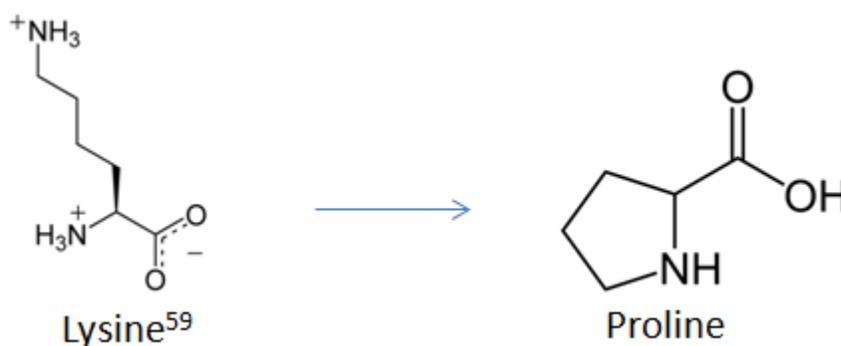


Figure 3.74 Mutating lysine⁵⁹ in the second linker to Proline.

The binding energy of K59P-DNA complex was calculated and found to be -163.56 ± 10.3 kcal/mol. A significant decrease in favorable electrostatic interactions of 398 kcal/mol was compensated by an equivalent reduction in the unfavorable electrostatic contribution to the solvation free energy. An increase in unfavorable van der Waals interactions was also estimated. The net result of these changes was a considerable decrease in the binding free energy. This reduction is probably due to the loss of the two water-mediated hydrogen bonds that Lys⁵⁹ forms with the phosphate group in G2 base, and also due to the constraints

introduced to this site by Proline.

Hydrogen bonds involving the amino acids in the first linker did not suffer considerable changes in K59P mutant, except for destabilization of one hydrogen bond connecting Lys31 in the first linker with Asn41 in finger two. On the other hand, hydrogen bonds involving amino acids in the second linker had more dramatic changes: three hydrogen bonds were significantly destabilized, one hydrogen bond was stabilized and one hydrogen bond was lost. Among these bonds, only one bond connecting Arg53 in the second finger and Glu58 in linker 2, had an increase in bond length. This increase in the bond length agrees with the detected destabilization of this bond in the K59P mutant.

For hydrogen bonds between the conserved amino acids in the α -helices of the protein, and DNA bases in the major groove, two highly stable hydrogen bonds were lost; one connecting G at base pair 10 with Arg22, the other connecting two conserved amino acids, Arg44 and Asp46, located in the second finger. In addition to this, one highly stable hydrogen bond had a reduction in its percent occupancy of ~55%. This change was accompanied by ~0.4 Å increase in its bond length. Extreme stabilization of a crucial bond connecting G8 and Arg22 was also detected. Three other hydrogen bonds between amino acids in the α -helices of

K59P and DNA bases experienced a reduction in their percent occupancy of 72%, 60% and 39% for each bond. (Figure 3.75 to 3.80).

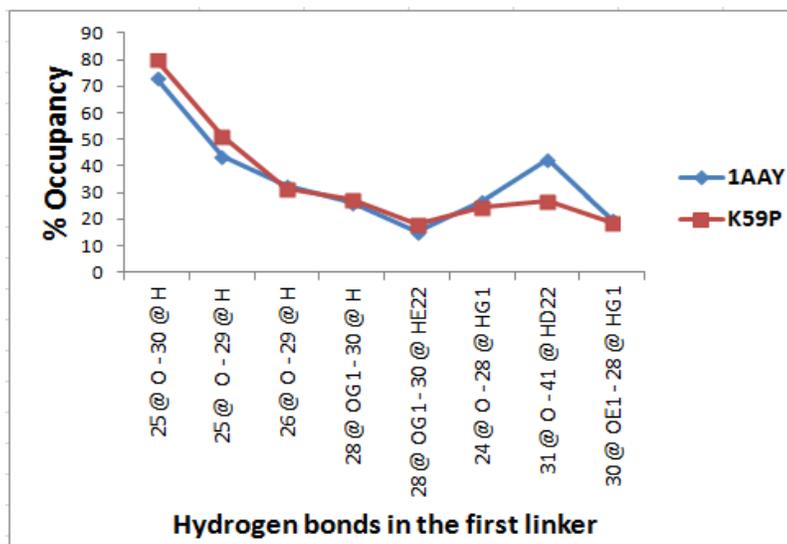


Figure 3.75 Percent occupancy of hydrogen bonds over production runs of 1AA Y and K59P. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

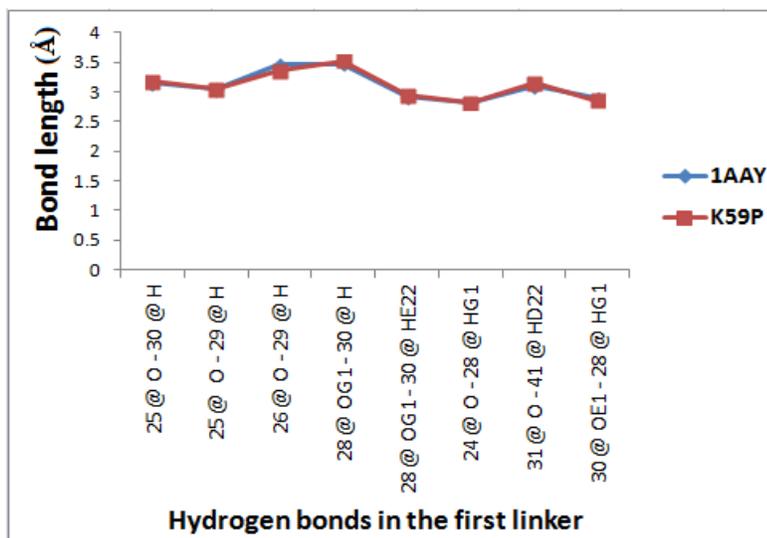


Figure 3.76 Average lengths of hydrogen bonds over production runs of 1AAY and K59P. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered

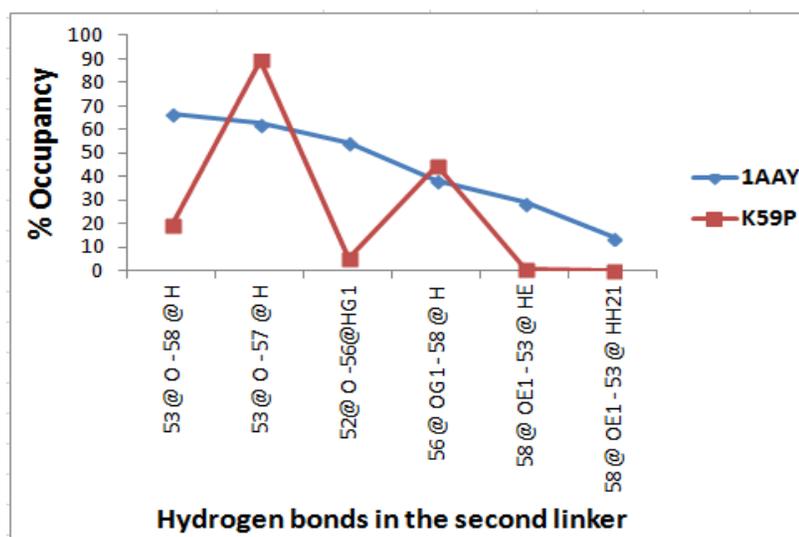


Figure 3.77 Percent occupancy of hydrogen bonds over production runs of 1AAY and K59P. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

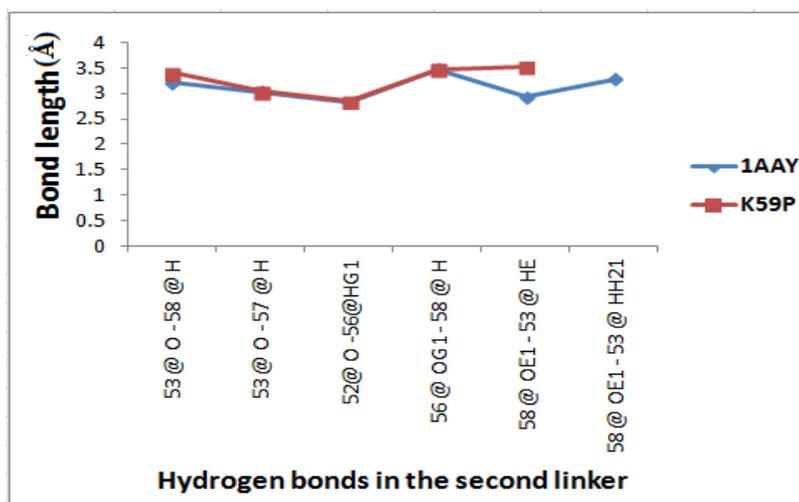


Figure 3.78 Average lengths of hydrogen bonds over production runs of 1AA Y and K59P. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

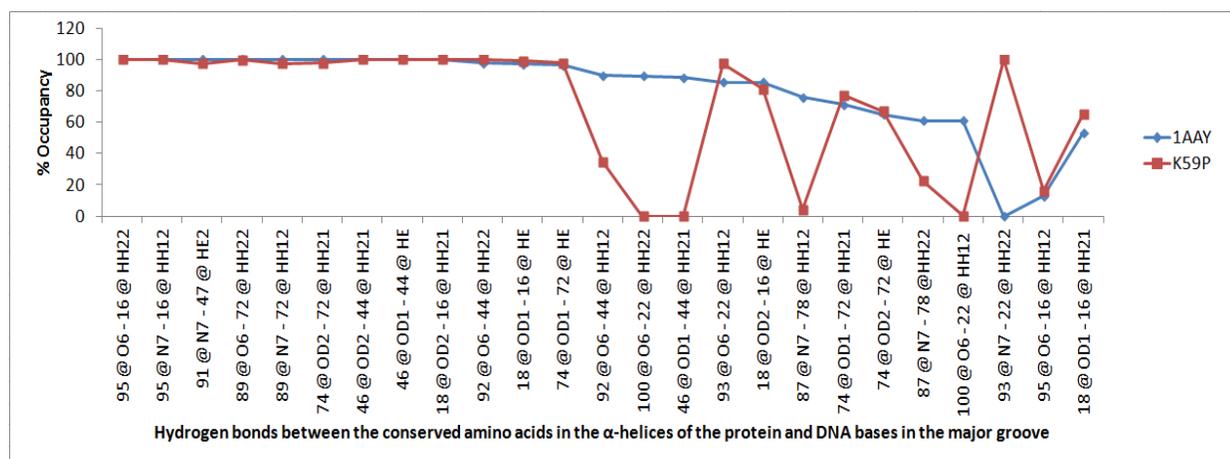


Figure 3.79 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and K59P.

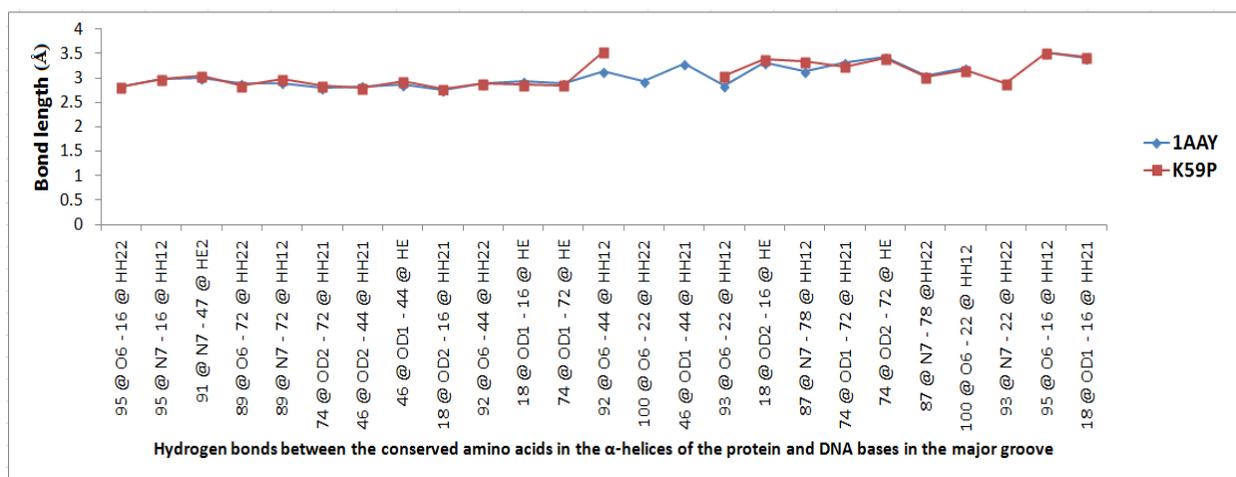


Figure 3.80 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major, over production runs of 1AAY and K59P.

3.3.11 Analysis of the binding free energy of P60A-DNA complex.

The last amino acid in linker two, Pro60, was mutated to Alanine. The latter is a small amino acid, and is expected to provide more conformational freedom to the linker, leading to a more negative binding energy.

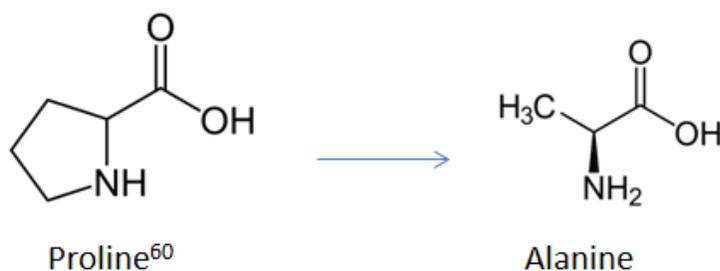


Figure 3.81 Mutating Proline60 in the second linker to Alanine.

The total binding energy for P60A-DNA complex as estimated by MM/GBSA is -207±10 kcal/mol. This value indicates a more favorable energy of binding than that of the wild type protein. The unfavorable electrostatic contributions as calculated by molecular mechanics had a reduction of approximately 37kcal/mol, accompanied by a reduction in the unfavorable Van der Waals interactions by approximately 10 kcal/mol. These calculated changes can be attributed to the small size and greater conformational freedom of Alanine compared to those of Proline.

It is worth mentioning that mutating Proline in the first linker to the smallest amino acid Glycine, did not produce a similar effect. Thus in spite of the high similarity between the two linkers, each linker has its unique effect, by being located between different functional domains, i.e., the fingers.

Hydrogen bond analysis for P60A mutant shows that of the eight analyzed hydrogen bonds in the first linker, two were considerably stabilized. In addition, one unstable bond in 1AAY was further destabilized, and two other unstable hydrogen bonds were lost. One of the stabilized hydrogen bonds in P60A is the hydrogen bond responsible for C-capping through connecting the hydroxyl group oxygen in Thr²⁸ with the hydrogen atom in the backbone amide of Gln³⁰. The percent occupancy of this bond increased by 19%, compared to its occupancy in

1AAY. This means more stabilization for C-capping in the first linker, and subsequently more stabilization for the P60A-DNA complex.

The second linker had a single noticeable change on its hydrogen bonds: a considerable stabilization of one hydrogen bond connecting Arg⁵³ in finger 2 and Glu⁵⁸ in the second linker. In spite of these changes in hydrogen bonds' stability, no considerable changes in bond lengths were observed.

Analysis of the crucial hydrogen bonds connecting the conserved amino acids in the α -helices of the protein, and DNA bases in the major groove, indicate the following changes: Among the highly stable hydrogen bonds, two hydrogen bonds connecting Arg⁴⁴ and G at base pair 8 were extremely destabilized, and two other hydrogen bonds were lost. The moderately stable and unstable hydrogen bonds in 1AAY witnessed different changes in P60A, where three bonds were destabilized, five hydrogen bonds were stabilized and two new bonds were formed. The bond connecting Arg²² and G at base pair 9 had a reduction in its bond length of 0.35 Å corresponding to the observed increase in percent occupancy of this bond. The hydrogen bond connecting Arg⁴⁴ and Asp⁴⁶ had an increase of 0.33 Å corresponding to the observed destabilization of this bond. (Figure 3.82 to 3.87)

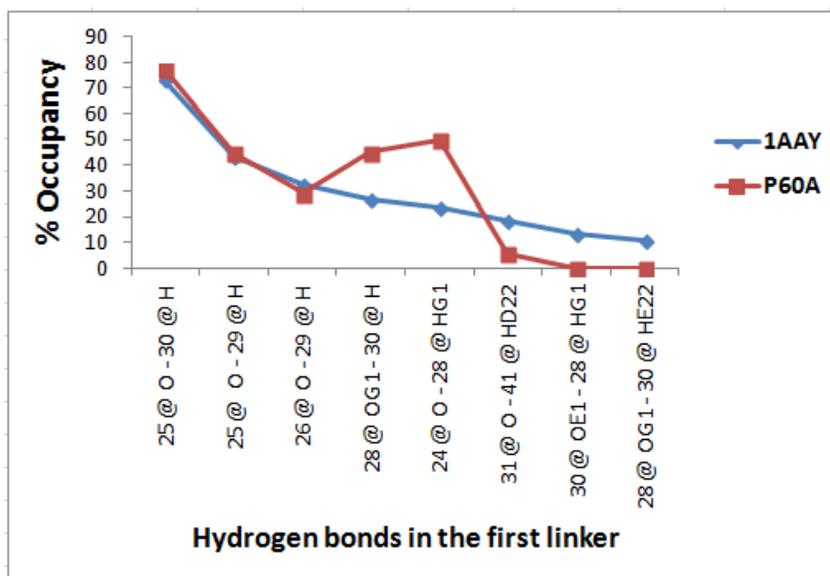


Figure 3.82 Percent occupancy of hydrogen bonds over production runs of 1AAY and P60A. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

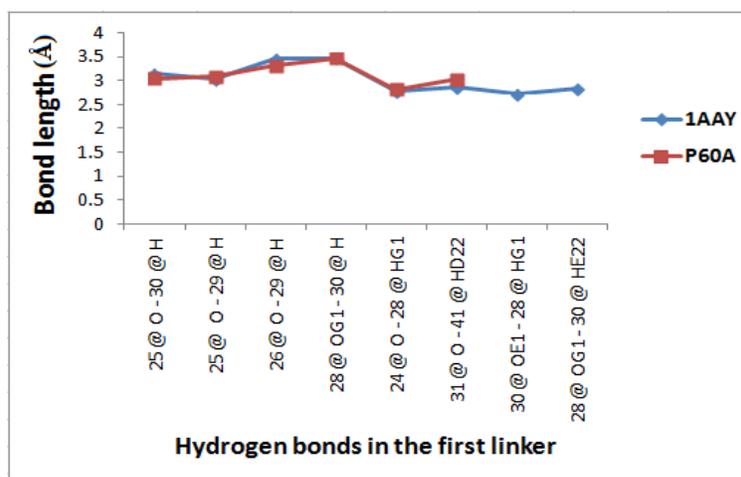


Figure 3.83 Average lengths of hydrogen bonds over production runs of 1AAY and P60A. Only hydrogen interactions between amino acids in linker 1 and other amino acids are considered.

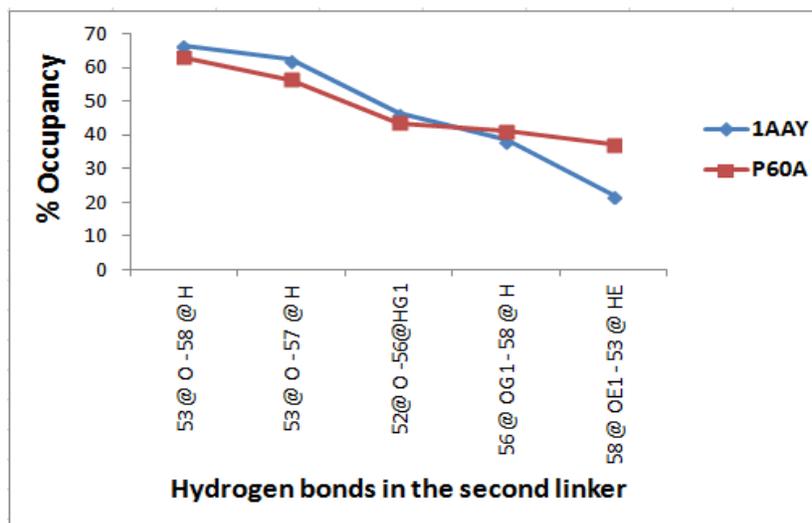


Figure 3.84 Percent occupancy of hydrogen bonds over production runs of 1AAY and P60A. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered.

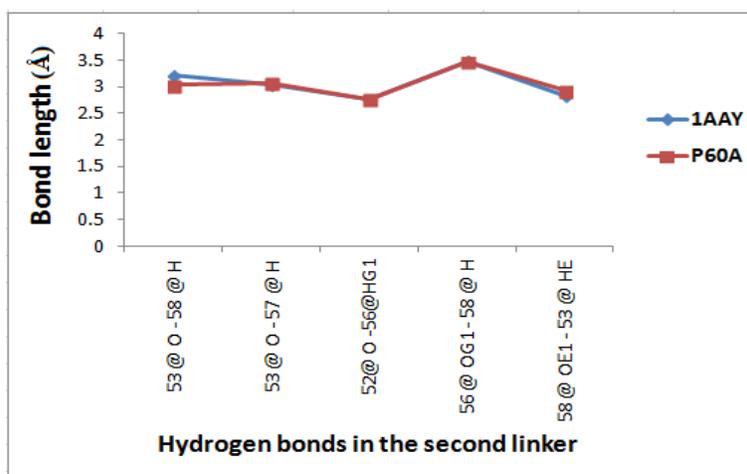


Figure 3.85 Average lengths of hydrogen bonds over production runs of 1AAY and P60A. Only hydrogen interactions between amino acids in linker 2 and other amino acids are considered

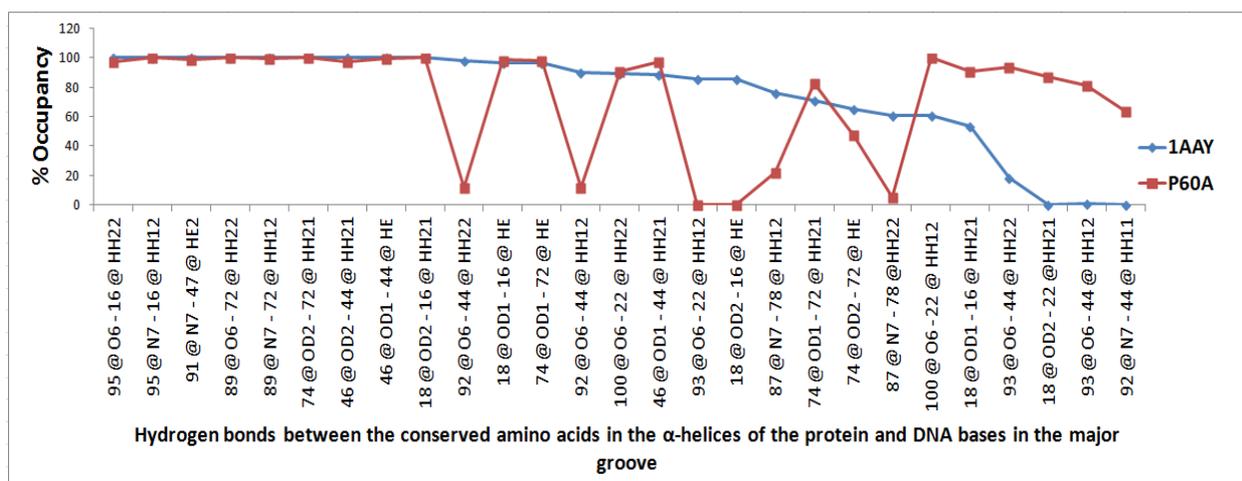


Figure 3.86 Percent occupancy of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and P60A.

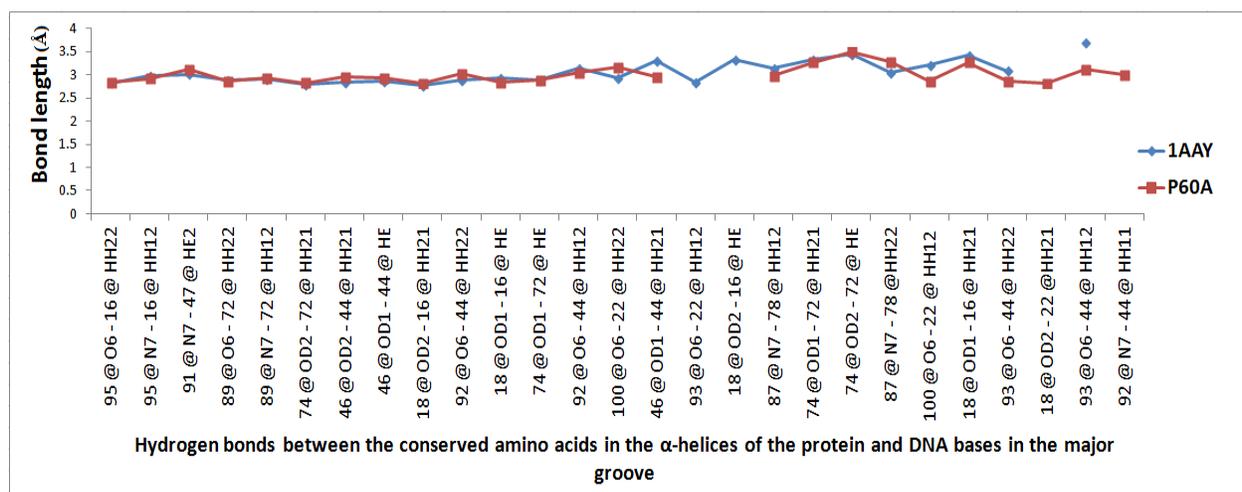


Figure 3.87 Average lengths of the crucial hydrogen bonds between, the conserved amino acids in the α -helices of the protein and bases in DNA major groove, over production runs of 1AA Y and P60A.

3.4 Mutagenesis studies of Zif268-DNA complex.

The role played by the highly conserved TGEKP Krüppel-type linker is not well understood. Mutagenesis studies including calculating either the dissociation constant or the binding affinity of mutants of the TGEKP linker and some of its variants in several transcription factors which contain zinc finger protein, such as ADR1, SP1 and TFIIIA, confirmed the importance of Krüppel-type linkers in stabilizing TFIIA-DNA complex³⁶.

Before comparing our computational results with relevant experimental mutagenesis studies, there are several things which should be considered. There are two known ways to calculate binding free energy. The first is by following a single trajectory approach. The molecular dynamic simulations are only carried out for the complex. This is then, followed by extracting snapshots for the complex, the unbound protein, and the unbound DNA from the trajectory of the complex. This approach saves computational time considerably⁷¹. On the other hand, it was reported to produce overstabilized values of binding energy⁷². This may be explained by the fact that it neglects any conformational changes that might take place in both DNA and protein systems upon binding^{20,36,72}. When the snapshots of the free DNA and protein are extracted from the trajectory of the complex, it gives

strained conformations that are different from their actual conformation in the free state⁷¹.

The second way is by following multiple trajectories. The molecular dynamic simulations are carried out for the complex, unbound protein and unbound DNA. This is followed by the extraction of snapshots for each one of these three systems from its respective trajectory. This approach is more accurate than the first one, but it is time consuming⁷¹. It approximately needs double the time needed to perform the single trajectory approach.

In our work the single trajectory approach was adopted to lower the computational cost. However this might have led to overstabilized values for the calculated binding free energies, compared to the experimentally calculated binding free energy of Zif268-DNA complex. It should be pointed out that both the Zif268 and DNA experienced conformational change due to binding^{20,36,72}.

Another key point here is to differentiate between two related terms which are the relative binding free energy and the absolute binding free energy, with the latter being equivalent to the binding free energy calculated experimentally. However, calculating the absolute binding free energy using MM/GBSA is usually avoided for two reasons. First, the absolute binding free energy can be calculated

by subtracting the unfavorable conformational entropy from the relative binding free energy. This is known to consume a considerable computational time and result in a large error⁵³.

The second reason is that MM/GBSA method was reported to produce overestimated binding free energy values due to overestabilization of the ionic interaction⁷³. It is clearly established in the literature that MM/GBSA method is usually most beneficial in calculating relative binding energy⁷⁴⁻⁷⁶, where any approximation and overestimation will be introduced in equal values, and would cancel out when comparing these values together.

Having these key points clarified, we can now proceed toward other experimental mutagenesis analysis that resembles our work in some parts and compare the results in a more flexible way.

The experimental value for binding free energy is -13.409 kcal/mol⁷². The absolute binding free energy for Zif268-DNA complex was calculated to get an estimate of overestabilization introduced to the computational result. The entropy contribution for this complex is -94.6 ± 15.5 kcal/mol. Thus the absolute binding free energy as calculated by MM/GBSA equals -84.79 ± 32.39 kcal/mol.

In one experimental study, a mutational analysis of zif268 was performed⁷⁷.

Point mutants of this protein were prepared *in vitro*, then the binding affinity of thirty one point mutants of zif268 to one of its reported binding site (5' -GCG GGG GCG- 3') was examined using gel shift assay. The binding phenotype of these complexes was reported. Among the thirty one point mutants, nine of them occurred in the linkers and are summarized in Figure 3.32⁷⁷.

Finger 1	Linker 1	Finger 2	Linker 2	Finger 3	Binding phenotype
	T G Q K P		T G E K P		
	T R Q K P		T G E K P		++
	T G Q N P		T G E K P		-
	T G Q K P		I G E K P		+
	T G Q K P		A G E K P		±
	T G Q K P		T D E K P		++
	T G Q K P		T G G K P		±
	T G Q K P		T G E N P		-
	T G Q K P		T G E K T		++
	T G Q K P		T G E K A		++

++ means ≥ 80% of wild-type binding
 + means ~ 25- 40% of wild-type binding affinity
 ± means < ~ 10% of wild-type binding affinity
 - refers to no detectable binding.

Figure 3.32 Amino acid sequence of the two canonical linkers in Zif268 and nine of its point mutants. Amino acids are represented by their one letter abbreviation. Amino acids in red show the resultant amino acid in each mutant. The binding phenotype of each protein to one of Zif268 characterized binding site (5-GCG GGG GCG-3) is shown on the right column. Data from Wilson, T. E. et al. (1992)⁷⁷.

As can be seen from the figure above four point mutants reduced the binding affinity only slightly, these are⁷⁷:

- (i) mutating the smallest amino acid, Glycine, in the first linker to a positive amino acid, Arginine,.

- (ii) mutating Glycine in the second linker to a negative amino acid, Aspartate,.
- (iii) mutating Proline in the second linker to Alanine.
- (iv) mutating Proline in the second linker to Threonine.

In our study mutating Glycine to Proline caused a considerable reduction in binding free energy, whereas mutating Glycine in the second linker reduced the binding free energy slightly. Mutating Proline in the second linker to Alanine caused a considerable increase in binding energy in contradiction with the experimental study in which a slight decrease in binding affinity was observed⁷⁸.

In the experimental study mutating Threonine in the second linker to Isoleucine reduced the binding affinity to less than the half of its original value⁷⁷. In our study mutating the same amino acid to tyrosine caused a significant decrease in the binding free energy.

In the experimental study some mutants showed no detectable binding affinity compared with the wild type, these are⁷⁷:

- (i) mutating Lysine to Asparagine in the first and second linkers, separately.
- (ii) mutating Threonine in the second linker to Alanine.

(iii) mutating Glutamate in the second linker to Glycine.

In our work mutating Lysine in the first linker to a negative amino acid, Aspartate, caused only a slight reduction in binding free energy. Mutating Lysine in the second linker to a non polar and constrained amino acid Proline, caused a significant decrease in binding free energy, while mutating Threonine in the second linker to Tyrosine caused also a significant decrease in binding free energy. However mutating glutamate to Glutamine caused an increase in binding free energy.

Both binding sites (5' -GCG TGG GCG- 3') and (5' -GCG GGG GCG- 3') were reported as consensus high-affinity binding sites for Zif268, with a small difference in their binding affinity with Zif268 in favor of the former one^{6,77}

⁷⁹, which was used in this study.

Given that binding affinity differs according the identity and position of the mutated amino acid, in addition to the identity of the resultant amino acid, the results of this experimental work cannot be directly compared to our results. Even so some general conclusion may be extracted in the view of these studies.

Chapter 4

CONCLUSION

The binding free energy for Zif268-DNA with its optimal DNA binding site (5' A GCG TGG GCG T 3') was calculated by employing Molecular Mechanics/ Generalized Born Solvent Area method (MM/GBSA) in AMBER 9. It was shown that the estimated free energy of binding gave an overstabilized value compared to the value calculated experimentally. However, this overstabilization should not affect the comparative analysis in this study.

Ten amino acids in the linkers region of Zif268 were mutated separately, producing the following point mutants: T28A, G29P, Q30E, K31D, P32G, T56Y, G57V, E58Q, K59P, P60A. The free energy of binding for these ten mutants was calculated using MM/GBSA method. It was found that four mutants (T28A, Q30E, E58Q and P60A) resulted in considerably higher binding energy than the wild type protein Zif268. One mutant (T56Y) produced considerably lower binding energy, whereas

five mutants (G29P, Q30E, P32G, E58Q and K59P) did not show significant changes in binding energy.

The binding energy of Q30E and E58Q mutants to optimal DNA site were higher than that of Zif268, suggesting that each linker has its unique effect on binding energy. This suggestion is also supported by the difference in the effect of mutating Proline into a small amino acid in both linkers.

Hydrogen bond analysis revealed changes in stability and bond lengths of hydrogen bonds in the zinc finger protein-DNA complexes upon mutation. These changes included hydrogen bonds all over the zinc finger protein-DNA complexes. This finding suggests that mutating an amino acid in the linkers region affects the binding energy by both affecting a wide range of hydrogen bonds, and probably other electrostatic interactions in the zinc finger-DNA complex.

Having four mutants with higher binding energy and five mutants with no considerable change in binding energy indicates that the reason of conservation of the consensus linker sequence (TGEKP) could be due to the importance of this linker in biological roles other than specific binding to DNA.

This work can give a base for using molecular dynamics simulations in estimating the binding of zinc fingers to DNA bases and comparing the results with experiments to help in building more zinc finger proteins for medical purposes.

APPENDICES

Appendix A: Zinc finger protein files

File 1: metals.prep

```

0 0 3
-----calcium-----
CAL
CAL INT 0
CHANGE OMIT DU  BEG
0.0
1 DUMM  DU  M  0  -1  -2  0.000    0.000    0.000    0.000
2 DUMM  DU  M  1   0  -1  1.0000   0.0000   0.0000   0.000
3 DUMM  DU  M  2   1   0  1.0000   90.0000  0.0000   0.000
4 KA    KA  M  29.683 127.561 -18.450  2.00

DONE
-----manganese-----
MNG
MNG INT 0
CHANGE OMIT DU  BEG
0.0
1 DUMM  DU  M  0  -1  -2  0.000    0.000    0.000    0.000
2 DUMM  DU  M  1   0  -1  1.0000   0.0000   0.0000   0.000
3 DUMM  DU  M  2   1   0  1.0000   90.0000  0.0000   0.000
4 MN    MN  M  29.683 127.561 -18.450  2.00

DONE
-----zinc-----
ZNA
ZNA INT 0
CHANGE OMIT DU  BEG
0.0
1 DUMM  DU  M  0  -1  -2  0.000    0.000    0.000    0.000
2 DUMM  DU  M  1   0  -1  1.0000   0.0000   0.0000   0.000
3 DUMM  DU  M  2   1   0  1.0000   90.0000  0.0000   0.000
4 ZN    ZN  M  29.683 127.561 -18.450  2.00
DONE
STOP

```

File 2: Frcmod.ZNA

```
got these parameters from parm99.dat,  
This frcmod file is for ZNA (zinc 2+  
ion).
```

```
MASS  
ZN 65.4
```

```
NONB  
ZN 1.10 0.0125
```

File 3: ZNA.lib

```

!!index array str
"ZNA"
!entry.ZNA.unit.atoms table str name str type int typex int resx int flags int seq int
  elmnt dbl chg
  "ZN" "ZNA" 0 1 131075 1 30 2.000000
!entry.ZNA.unit.atomsperinfo table str pname str ptype int ptypex int pelmnt dbl pchg
  "ZN" "ZNA" 0 -1 0.0
!entry.ZNA.unit.boundbox array dbl
-1.000000
  0.0
  0.0
  0.0
  0.0
!entry.ZNA.unit.childsequence single int
10
!entry.ZNA.unit.connect array int
0
0
!entry.ZNA.unit.hierarchy table str
  abovetype int abovex str belowtype int belowx
  "U" 0 "R" 1
  "R" 1 "A" 1
!entry.ZNA.unit.name single str
""
!entry.ZNA.unit.positions table dbl x dbl y dbl z
1.759000 0.0 9.000000
!entry.ZNA.unit.residueconnect table int c1x int c2x int c3x int c4x int c5x int c6x
0 0 0 0 0
!entry.ZNA.unit.residues table str name int seq int childseq int startatomx str restype
  int imagingx
  "1." 9 2 1 "?" 0
!entry.ZNA.unit.residuesPdbSequenceNumber array int
1
!entry.ZNA.unit.solventcap array dbl
-1.000000
  0.0
  0.0
  0.0
  0.0
!entry.ZNA.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0

```

File 4: frmod.zf

```

# modifications to force field for zinc finger
proteins

MASS
ZN 65.36

BOND
NA-ZN 70.000 2.05000 #kludge by JRS
ZN-S 70.000 2.10000 #kludge by JRS
ZN-SH 70.000 2.90000 #for pcy
CT-SH 222.000 1.81000 #met(aa)

ANGLE
ZN-NA-CW 50.000 126.700 #JRS estimate
ZN-NA-CR 50.000 126.700 #JRS estimate
ZN-NA-CP 50.000 126.700 #JRS estimate
ZN-NA-CC 50.000 126.700 #JRS estimate
ZN-SH-CT 50.000 120.000 #JRS estimate
ZN-S -CT 50.000 120.000 #JRS estimate
ZN-S -C2 50.000 120.000 #JRS estimate
ZN-S -C3 50.000 120.000 #JRS estimate
NA-ZN-NA 10.000 110.000 #dac estimate
NA-ZN-SH 10.000 110.000 #dac estimate
NA-ZN-S 10.000 110.000 #dac estimate
SH-ZN-S 10.000 110.000 #dac estimate
ZN-SH-CT 50.000 120.000 #JRS estimate
CT-CT-SH 50.000 114.700 #met(aa)
HC-CT-SH 35.000 109.500
HI-CT-SH 35.000 109.500
CT-SH-CT 62.000 98.900 #MET(OL)
HS-SH-ZN 32.500 132.500
SH-ZN-SH 15.700 115.700 #
HS-SH-ZN 32.400 132.400 #
H-NA-ZN 10.000 110.000 #
H-NA-ZN 10.000 110.000 #(MET(OL))

DIHE
X -NA-ZN-X 1 0.000 180.000 3.000
X -ZN-SH-X 1 0.000 180.000 3.000
X -ZN-S -X 1 0.000 180.000 3.000
X -CT-SH-X 3 1.000 0.000 3.000

NONBON
ZN 1.10 0.0125

```

Appendix B: Input files for simulation.**File 1:** min.in

```
1AAV
&cntrl
  imin   = 1,
  maxcyc = 1000,
  ncyc   = 500,
  ntb    = 1,
  ntr    = 1,
  cut    = 10.0
/
Hold the ZNA fixed
50.0
RES 108 110
END
END
```

File 2: heat.in

```
laay-heat
&cntrl
  imin=0,irest=0,ntx=1,
  nstlim=25000,dt=0.002,
  ntc=2,ntf=2,
  cut=10.0, ntb=1,
  ntpr=500, ntwx=500,
  ntt=3, gamma_ln=2.0,
  tempi=0.0, temp0=300.0,
  ntr=1, restraintmask=':1-115',
  restraint_wt=2.0,
  nmropt=1
/
&wt TYPE='TEMP0', istep1=0, istep2=25000,
  value1=0.1, value2=300.0, /
&wt TYPE='END' /
Keep ZNA fixed with weak restraints
50.0
RES 108 110
END
END
```

File 3: density.in

```
laay-density
&cntrl
  imin=0,irest=1,ntx=5,
  nstlim=25000,dt=0.002,
  ntc=2,ntf=2,
  cut=8.0, ntb=2, ntp=1, taup=1.0,
  ntp=500, ntwx=500,
  ntt=3, gamma_ln=2.0,
  temp0=300.0,
  ntr=1, restraintmask=':1-61',
  restraint_wt=2.0,
/
Keep ZNA fixed with weak restraints
50.0
RES 108 110
END
END
```

File 4: equil.in

```
1AAY-equil
&cntrl
  imin=0,irest=1,ntx=5,
  nstlim=250000,dt=0.002,
  ntc=2,ntf=2,
  cut=10.0, ntb=2, ntp=1, taup=2.0,
  ntp=1000, ntwx=1000,
  ntt=3, gamma_ln=2.0,
  temp0=300.0,
/
Keep ZNA fixed with weak restraints
50.0
RES 108 110
END
END
```

File 5: prod.in

```
1AAZ-prod
&cntrl
  imin=0, irest=1, ntx=5,
  nstlim=250000, dt=0.002,
  ntc=2, ntf=2,
  cut=10.0, ntb=2, ntp=1, taup=2.0,
  ntpr=5000, ntwx=5000,
  ntt=3, gamma_ln=2.0,
  temp0=300.0,
/
Keep ZNA fixed with weak restraints
50.0
RES 108 110
END
END
```

File 6: calculate-entropy.in

```

#
# Input parameters for mm_pbsa.pl
# This example just generates snapshots from a trajectory file
#
# Holger Gohlke
# 08.01.2002
#
#####
@GENERAL
#
# General parameters
# 0: means NO; >0: means YES
#
# mm_pbsa allows to calculate (absolute) free energies for one molecular
# species or a free energy difference according to:
#
# Receptor + Ligand = Complex,
# DeltaG = G(Complex) - G(Receptor) - G(Ligand) .
#
# PREFIX - To the prefix, "{_com, _rec, _lig}.crd.Number" is added during
# generation of snapshots as well as during mm_pbsa calculations.
# PATH - Specifies the location where to store or get snapshots.
#
# COMPLEX - Set to 1 if free energy difference is calculated.
# RECEPTOR - Set to 1 if either (absolute) free energy or free energy
# difference are calculated.
# LIGAND - Set to 1 if free energy difference is calculated.
#
# COMPT - parmtop file for the complex (not necessary for option GC).
# RECPT - parmtop file for the receptor (not necessary for option GC).
# LIGPT - parmtop file for the ligand (not necessary for option GC).
#
# GC - Snapshots are generated from trajectories (see below).
# AS - Residues are mutated during generation of snapshots from trajectories.
# DC - Decompose the free energies into individual contributions
# (only works with MM and GB).
#
# MM - Calculation of gas phase energies using sander.
# GB - Calculation of desolvation free energies using the GB models in sander
# (see below).
# PB - Calculation of desolvation free energies using delphi (see below).
# Calculation of nonpolar solvation free energies according to
# the NPOPT option in pbsa (see below).
# MS - Calculation of nonpolar contributions to desolvation using molsurf
# (see below).
# If MS == 0 and GB == 1, nonpolar contributions are calculated with the
# LCPO method in sander.
# If MS == 0 and PB == 1, nonpolar contributions are calculated according

```

```
#
PREFIX          prod
PATH            ./traj1AAY
#
COMPLEX        1
RECEPTOR    1
LIGAND        2
#
COMPT         ./1AAY-dry.top
RECPT        ./1AAY-rec.top
LIGPT        ./1AAY-lig.top
#
GC            0
AS            0
DC            0
#
MM            0
GB            0
PB            0
MS            0
#
NM            1
#
#####
@NM
#
# Parameters for sander/nmode calculation (this section is only relevant if NM = 1 above)
#
#   The following parameters are passed to sander (for minimization) and nmode
#   (for entropy calculation using gasphase statistical mechanics).
#   For further details see documentation.
#
#   DIELC - (Distance-dependent) dielectric constant
#   MAXCYC - Maximum number of cycles of minimization.
#   DRMS - Convergence criterion for the energy gradient.
#
DIELC          4
MAXCYC        1000
DRMS           0.1
#
```

References

- (1) Clark, D. P.; Pazdernik, N. J. *Molecular Biology*; Elsevier Science, 2013.
- (2) Ganguly, A.; Rajdev, P.; Williams, S. M.; Chatterji, D. *The Journal of Physical Chemistry B* **2011**, *116*, 621.
- (3) Bujnicki, J. *Prediction of Protein Structures, Functions, and Interactions*; John Wiley & Sons, 2008
- (4) Harrison, S. C. *Nature* **1991**, *353*, 715.
- (5) Miller, J.; McLachlan, A. D.; Klug, A. *The EMBO journal* **1985**, *4*, 1609.
- (6) Pavletich, N.; Pabo, C. *Science* **1991**, *252*, 809.
- (7) Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*; University Science Books, 1994.
- (8) Krishna, S. S.; Majumdar, I.; Grishin, N. V. *Nucleic Acids Research* **2003**, *31*, 532.
- (9) Laity, J. H.; Lee, B. M.; Wright, P. E. *Current Opinion in Structural Biology* **2001**, *11*, 39.
- (10) Roat-Malone, R. M. *Bioinorganic Chemistry: A Short Course*; Wiley, 2007.
- (11) Jacobs, G. H. *The EMBO journal* **1992**, *11*, 4507.
- (12) Pellegrino, G. R.; Berg, J. M. *Proceedings of the National Academy of Sciences* **1991**, *88*, 671.
- (13) Nardelli, J.; Gibson, T.; Charnay, P. *Nucleic Acids Research* **1992**, *20*, 4137.
- (14) Miller, J. C.; Pabo, C. O. *Journal of Molecular Biology* **2001**, *313*, 309.
- (15) Wolfe, S. A.; Grant, R. A.; Elrod-Erickson, M.; Pabo, C. O. *Structure* **2001**, *9*, 717.
- (16) Wolfe, S. A.; Greisman, H. A.; Ramm, E. I.; Pabo, C. O. *Journal of Molecular Biology* **1999**, *285*, 1917.
- (17) Benos, P. V.; Lapedes, A. S.; Stormo, G. D. *Journal of Molecular Biology* **2002**, *323*, 701.
- (18) Marco, E.; García-Nieto, R.; Gago, F. *Journal of Molecular Biology* **2003**, *328*, 9.
- (19) Locasale, J. W.; Napoli, A. A.; Chen, S.; Berman, H. M.; Lawson, C. L. *Journal of Molecular Biology* **2009**, *386*, 1054.
- (20) Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. *Structure* **1996**, *4*, 1171.
- (21) Feinauer, C. J.; Hofmann, A.; Goldt, S.; Liu, L.; Máté, G.; Heermann, D. W. In *Advances in Protein Chemistry and Structural Biology*; Rossen, D., Ed.; Academic Press: 2013; Vol. Volume 90, p 67.
- (22) Klug, A. *Quarterly Reviews of Biophysics* **2010**, *43*, 1.

- (23) Wolfe, S. A.; Ramm, E. I.; Pabo, C. O. *Structure* **2000**, *8*, 739.
- (24) Peisach, E.; Pabo, C. O. *Journal of Molecular Biology* **2003**, *330*, 1.
- (25) Rebar, E. J.; Greisman, H. A.; Pabo, C. O. In *Methods in Enzymology*; John, N. A., Ed.; Academic Press: 1996; Vol. Volume 267, p 129.
- (26) Morozov, A. V.; Havranek, J. J.; Baker, D.; Siggia, E. D. *Nucleic Acids Research* **2005**, *33*, 5781.
- (27) Condit, C. M.; Railsback, L. B. *Journal of biomedical discovery and collaboration* **2007**, *2*, 5.
- (28) Choo, Y.; Klug, A. *Nucleic Acids Research* **1993**, *21*, 3341.
- (29) Nagaoka, M.; Nomura, W.; Shiraishi, Y.; Sugiura, Y. *Biochemical and Biophysical Research Communications* **2001**, *282*, 1001.
- (30) Kim, J.-S.; Pabo, C. O. *Proceedings of the National Academy of Sciences* **1998**, *95*, 2812.
- (31) Anand, P.; Schug, A.; Wenzel, W. *FEBS Letters* **2013**, *587*, 3231.
- (32) Ryan, R. F.; Darby, M. K. *Nucleic Acids Research* **1998**, *26*, 703.
- (33) Mark, P. F.; Deborah, S. W.; Ishwar, R.; David, A. C.; Joel, M. G.; Peter, E. W. *Nature Structural & Molecular Biology* **1997**, *4*, 605.
- (34) Laity, J. H.; Dyson, H. J.; Wright, P. E. *Proceedings of the National Academy of Sciences* **2000**, *97*, 11932.
- (35) Laity, J. H.; Chung, J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2000**, *39*, 5341.
- (36) Laity, J. H.; Dyson, H. J.; Wright, P. E. *J Mol Biol* **2000**, *295*, 719.
- (37) Liu, Q.; Segal, D. J.; Ghiara, J. B.; Barbas, C. F. *Proceedings of the National Academy of Sciences* **1997**, *94*, 5525.
- (38) Nekludova, L.; Pabo, C. O. *Proceedings of the National Academy of Sciences* **1994**, *91*, 6948.
- (39) Urnov, F. D.; Miller, J. C.; Lee, Y.-L.; Beausejour, C. M.; Rock, J. M.; Augustus, S.; Jamieson, A. C.; Porteus, M. H.; Gregory, P. D.; Holmes, M. C. *Nature* **2005**, *435*, 646.
- (40) Fogolari, F.; Brigo, A.; Molinari, H. *Biophysical Journal* **2003**, *85*, 159.
- (41) Ajay; Murcko, M. A. *Journal of Medicinal Chemistry* **1995**, *38*, 4953.
- (42) Hou, T.; Chen, K.; McLaughlin, W. A.; Lu, B.; Wang, W. *PLoS computational biology* **2006**, *2*.
- (43) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 5221.
- (44) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *The Journal of Physical Chemistry A* **1997**, *101*, 3005.
- (45) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *Journal of Chemical Theory and Computation* **2005**, *2*, 128.
- (46) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 3986.
- (47) Naïm, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. *Journal of*

Chemical Information and Modeling **2007**, 47, 122.

(48) Swanson, J. M. J.; Henschman, R. H.; McCammon, J. A. *Biophysical Journal* **2004**, 86, 67.

(49) Beierlein, Frank R.; Kneale, G. G.; Clark, T. *Biophysical Journal*, 101, 1130.

(50) ALMLÖF, M., Uppsala University, 2007.

(51) Lu, N.; Kofke, D. A. *The Journal of Chemical Physics* **2001**, 115, 6866.

(52) Lu, N.; Kofke, D. A. *The Journal of Chemical Physics* **2001**, 114, 7303.

(53) Hayes, J. M.; Archontis, G. *Molecular Dynamics - Studies of Synthetic and Biological Macromolecules*, 2012.

(54) Homeyer, N.; Gohlke, H. *Molecular Informatics* **2012**, 31, 114.

(55) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier Science, 2001.

(56) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, K. M. M., D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S.; Hayik, A. R., G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H.; Gohlke, L. Y., C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C.; Schafmeister, W. S. R., and P.A. Kollman *AMBER 9* University of California, San Francisco, 2006.

(57) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Computer Physics Communications* **1995**, 91, 1.

(58) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, 3, 198.

(59) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *WIREs Comput Mol Sci* **2012**, 1.

(60) Ponder, J.; Case, D. In *Protein Simulations*; Elsevier: 2003; Vol. 66, p 27.

(61) Granadino-Roldán, J. M.; Obiol-Pardo, C.; Pinto, M.; Garzón, A.; Rubio-Martínez, J. *Journal of Molecular Graphics and Modelling* **2014**, 50, 142.

(62) Honig, B.; Nicholls, A. *Science* **1995**, 268, 1144.

(63) Warwicker, J.; Watson, H. C. *Journal of Molecular Biology* **1982**, 157, 671.

(64) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006**, 65, 712.

(65) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press, 1997.

(66) DeLano, W. L. 2002.

(67) P.Allen, M. *NIC Series* **2014**, 23, 1.

(68) Nurisso, A.; Daina, A.; Walker, R. In *Homology Modeling*; Orry, A. J. W., Abagyan, R., Eds.; Humana Press: 2012; Vol. 857, p 137.

(69) Hamed, M. Y. *Journal of Computer-Aided Molecular Design*, Manuscript submitted for publication.

(70) Tsui, V.; Radhakrishnan, I.; Wright, P. E.; Case, D. A. *Journal of*

Molecular Biology **2000**, 302, 1101.

(71) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2012**, 8, 3314.

(72) Lee, J.; Kim, J.-S.; Seok, C. **2010**, 114, 7662.

(73) Campbell, Z. T.; Baldwin, T. O.; Miyashita, O. *Biophysical Journal*, 99, 4012.

(74) Hayes, J. M.; Archontis, G. *MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies*; INTECH Open Access Publisher, 2012.

(75) Hou, T.; Wang, J.; Li, Y.; Wang, W. *Journal of Chemical Information and Modeling* **2011**, 51, 69.

(76) Singh, N.; Warshel, A. *Proteins: Structure, Function, and Bioinformatics* **2010**, 78, 1705.

(77) Wilson, T. E.; Day, M. L.; Pexton, T.; Padgett, K. A.; Johnston, M.; Milbrandt, J. *Journal of Biological Chemistry* **1992**, 267, 3718.

(78) Swirnoff, A. H.; Milbrandt, J. *Molecular and cellular biology* **1995**.

(79) Christy, B.; Nathans, D. *Proceedings of the National Academy of Sciences* **1989**, 86, 8737.