ANALYSIS OF HOT SPOTS IN PROTEIN-PROTEIN INTERACTIONS

by

Sabri Bora Erdemli

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computational Science and Engineering

Koc University

July 2005

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Sabri Bora Erdemli

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Metin Türkay, Ph. D. (Advisor)

Özlem Keskin, Ph. D. (Co-Advisor)

Çağatay Başdogan, Ph. D.

Canan Baysal , Ph. D.

Atilla Gürsoy, Ph. D.

Date: _____

To my parents, my brother and
Vinil

ABSTRACT

Proteins generally function through reciprocal cooperation; they bind together into protein complexes and help each other fulfill their functions. Almost every level of cell function is established by protein – protein interactions including formation of structure in cellular organelles, the transport entities across the various biological membranes and signal transduction. The objective of this thesis is to investigate the role of energetically important residues called “hot spots” in protein associations. For this purpose 14 different complexes belonging to 4 different complex types are analyzed. Molecular dynamics simulations are performed over these protein complexes. Hot spots and non hot spot residues in the interfaces are characterized with respect to their flexibility and hydrogen bond forming ability. Interfaces are dominated by hydrophobic residues in almost all types of complexes. Then the second contribution is from residues that have aromatic side chains. Tyr, Ser, Phe, Gly and Thr are more frequent as hot spot residues than the other residues in the interfaces. Hot spots are found to be more buried than the rest of the interface. It is observed that the hot spot residues exhibit less mobility than the rest of the interface residues. The hydrogen bond formation ability of the hot spots are not significantly different from the others. These findings should be very important in understanding protein – protein interaction and development of docking algorithms.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisors, Asst. Prof. Metin Türkay and Asst. Prof. Özlem Keskin, whose expertise, understanding, and patience added considerably to my graduate experience. I would like to thank the other members of my committee, Assoc. Prof. Canan Baysal , Asst. Prof. Atilla Gürsoy and Asist. Prof. Çağatay Başdoğan, for critical reading of this thesis and for their valuable comments. I am grateful to all my friends at Koc university for their continued support to name some my roommate Utkan Öğmen, my officemates Uğur Güner, Ayşe Meriç Ovacık, Volkan Kurt, Erk Subaşı, Murat Danışman and Sefa Nur Demirağ. I would like to thank to my friends from different department but with the same sentiments who are Burak Öztürk , Faik Can Meral and Hüseyin Erdim. My very special friends Cihan Ataş and Barış Bingöl, they were always with me. I would like to thank to Yasin Şenbabaoğlu for his valuable discussions. Above all to my family who were always with me their encouragement, support and care when it was most requirement. Finally, I would like to thank my love, my best friend, Vinil. I could not finish this thesis without her unconditional understanding, patience, endless moral support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ASA	Accessible surface area
MD	Molecular dynamics
NAMD	Not another molecular dynamics
NMR	Nuclear magnetic resonance
PDB	Protein data bank
PME	Particle Mesh Ewald
VMD	Visual Molecular Dynamics
HS	Hot Spot
NHS	Non Hot Spot

Chapter 1

Introduction

There are four main groups of biological polymers that are polysaccharides, lipids, nucleic acids and proteins. Of the four groups of biopolymers, proteins have the most diverse functions. As enzymes and hormones, proteins catalyze and regulate the reactions that occur in the body, as muscles and tendons they provide the body with the means of movement, as hemoglobins they transfer all important oxygen to its most remote corner, as antibodies they provide the body with means of protection against diseases and in combination with other substances in body they provide structural support.

Proteins generally function through reciprocal cooperation; they bind together into protein complexes and help each other fulfill their functions. The biological functions of proteins depend on the physical interaction with other molecules and especially with other proteins. Almost every level of cell function is established by protein – protein interactions including formation of structure in cellular organelles, the transport entities across the various biological membranes and signal transduction. Protein – protein interactions have been the object of intense research for many years because of their importance in development of biological organisms and onset and progress of diseases.

Protein – protein interaction is established at the interfaces. There has been a large number of studies to characterize the protein interfaces. Although some of these studies elucidated that different types of interfaces have similar amino-acid composition [1-3] ,other studies showed differences [4, 5]. Other analysis with a broad range of protein – protein interfaces has shown that there are no general rules for the characterization of protein

interfaces such as planarity, shape and hydrophobicity [4, 6]. Philosophy of interactions is not fully clarified despite intensive research on this subject [7, 8]. In order to explain the characteristic of the interactions, amino acid conservation at protein interfaces, amino acid propensities and interface characteristics such as size and shape are studied recently [8, 9]. Another important subject in recent studies is to focus on important binding sites of the interfaces which play on important role in the docking process. Binding free energy changes in complex formation is a good indicator of binding affinity in protein-protein association. Clackson & Wells [10] found that residues called "hot spots" contribute to the binding free energy with a large extent despite their smaller areas occupied in the in the protein – protein interfaces. There are some studies in the literature that aim to explain the behavior and structure of hot spots in the interfaces. Ma and Nussinov [8] have shown that hot spots predominantly occur in the interfaces. Bogan and Thorn [9] found that tryptophan (TRP), tyrosine (TYR) and arginine (ARG) are the most commonly seen amino acids as hot spots. Moreover, some residues which are energetically less important prevent hot spots from bulk solvent that is especially crucial for electrostatic interaction [9].

There are several experimental and computational methods developed for predicting the hot spots among the interfaces. Clackson and Wells [10], and Wells [11] developed an experimental technique that is alanine scanning mutagenesis followed by the measurement of each mutant's effect on binding. Kortemme and Baker [12] developed a model by examining the contribution of interface residues in binding free energy. They predict hot spots computationally by calculating free energy changes. These calculations include hydrogen bonds, implicit solvent and packing interaction by ignoring changes in backbone conformation or effects on the dynamic interface. Hu et al. used Geometric Hashing algorithm to make structural comparison with a structural non redundant data set of 11 interface families with 97 protein interfaces [13]. They found that conserved interface residues strongly correlated with the conserved interfaces residues. It is postulated that

evolution optimized the protein interfaces according to their functions. This further means that the residues which are critical for binding are evolutionarily conserved [14-16]. It has been shown that alanine scanning mutagenesis data are well correlated with the residue conservation [8,9].

In this study interfaces of 14 different protein complexes are analyzed by using molecular dynamics by focusing on the hot spots across the interface. The protein structures are determined according to their size and type. Four different classes of protein complexes studied: homodimer, enzyme – inhibitor, antibody – antigen and protein – peptide. The computational hot spots are determined according to conservation analysis by using CONSURF [17]. Molecular dynamics simulations were performed using NAMD simulation package. The resulting trajectories are analyzed by using root mean square deviation (rmsd) calculations, accessible surface area (ASA) calculation and hydrogen bond forming ability. Moreover, the interface of the complexes used in the study is investigated considering basic characteristics of the interfaces. It is found that side chain flexibility of computational hot spot residues are less than non hot spot residue across the interface. It has been observed that the difference is significant for all complex types except antibody – antigen by applying statistical analysis. Hydrogen bonding ability does not differ between them except charged residues in the interfaces. ASA of hot spot residues are also less than non hot spot residues in all types of the complexes meaning that hot spots are mainly buried. The characteristics of the interfaces have been different also between different types of the interfaces.

Outline:

This thesis study contains six chapters. In the second chapter, literature survey will be represented which illustrate the recent challenges and advancements in protein related sciences especially concentrated on protein – protein interactions and hot spots is given. Chapter 3 gives information about the biological context of the protein complexes. The concepts of molecular dynamics simulation and the protocol used in this study are illustrated in Chapter 4. Chapter 5 consists of detailed exhibition of results and relevant discussions. The final part of the thesis includes conclusions and appendix that contain tabulated results and detailed explanation of methods.

Chapter 2

Literature Review

This chapter is organized as follows: Section 2.1 is an overview of recent studies on proteins which are related to their structures and function. Section 2.2 elaborates on current researches about protein – protein interactions and protein interfaces. Section 2.3 summarizes the recent progress on energetically important residues in binding called “hot spots”. Section 2.4 concludes the chapter by discussing some experimental and computational method to detect or predict hot spot residues in protein – protein interaction.

2.1 Protein

Proteins have the most diverse functions in biological systems. As enzymes and hormones, proteins catalyze and regulate the reactions that occur in the body, as muscles and tendons they provide the body with the means of movement, as a hemoglobin that transfers all important oxygen to its most remote corner, as antibodies they provide the body with means of protection against diseases and in combination with other substances in body they provide structural support[18].

Proteins are built up of amino acids that are linked by peptide bonds to form a polypeptide chain. All proteins are polymers that are built from 20 different natural amino acids. Natural proteins are unbranched and linear unlike the other chemical polymers and have precise lengths and exact sequences of amino acids.

Proteins have evolved to perform specific functions. The function of the proteins arises according to their three dimensional structure. The particular amino acid sequences, which are linear chains, fold to generate specific three dimensional structures.

The biological functions of proteins depend on the physical interactions with other molecules and especially with other proteins. Every molecule in a cell was first bound by the enzyme that produce it or by the receptor on the cell surface that enabled it to be taken up. To understand the function of a protein, it is important to identify its role in cellular interaction network. Therefore, as a starting point characterization of protein interfaces and interactions supply tremendously essential information.

Figure 2.1 Map of protein- protein interaction in yeast.[19]

2.2 Protein – protein Interaction and Protein Interfaces

Proteins participate in almost every level of cell function, in the structure of cellular organelles, the transport machinery across the various biological membranes and signal transduction to name a few. However, proteins usually can not accomplish their function alone; they bind together with other proteins to fulfill certain biological functions [20]. Because of their importance in development and disease and the process in living organisms, protein – protein interactions have been the object of intense research for many years. In order to understand protein – protein interactions, the protein interfaces and their binding sites should be analyzed in detail. Comprehending the protein – protein interaction extensively will provide great improvement in the conformational drug researches, prediction of protein-protein interaction and constitute protein interaction networks [21]. Protein interaction networks are very complicated as shown in Figure 2.1, dots are the proteins and the lines indicate interactions between proteins.

There are numerous experimental and computational techniques in order to characterize protein interfaces and protein – protein interactions. This is critical in order to understand intermolecular interaction between bio-molecules dominantly in proteins. In early studies the physical and chemical characteristics of protein interfaces are addressed as hydrophobic, planar and round [22]. Further studies extend these researchers. They analyze protein interface according to their size and shape, hydrophobicity [23], amino acid propensity, segmentation and secondary structure, complementarity [24] and existence in nature such as obligatory interfaces and transient ones [6, 25]. The obligate protein – protein interactions the protein can not be found as stable structures alone and they can not also function without partner [26]. The monomers of the transient complexes can be observed independently. An example of protein-protein interface can be seen in Figure 2.2.

The size of the interface can be measured directly in absolute dimensions Å or by solvent accessible area calculations (ASA). The shape of the protein interfaces can be determined with a ratio of lengths of the principal axes of the least square planes through the atoms in interfaces. Complementarity can be divided into electrostatic and shape complementarity. The complementarity can be measured with an index called gap index, which is the total gap volume between two interacting proteins per interface ASA. Residue propensity can show residue preferences in the interfaces taking into account the preferences in all protein. The discontinuous nature of protein interfaces can be investigated by dividing the interface into segments.

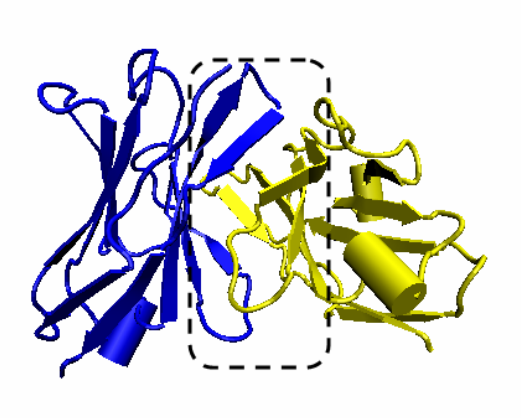


Figure 2.2 Protein interfaces between two chains of Crystal structure of human lamda-6 light chain dimmer PDB code 1CD0.

The ΔASA, ASA change upon complexation, of the homodimers is larger than the hetero complexes [6]. That means, the homodimers are more closely packed than the other types of interfaces. There are some characteristic differences between different types of interfaces; the interfaces are more hydrophobic for obligate complexes mostly homodimers [6, 24]. However, antigen binding sites conversely found to be hydrophilic [27]. Although

there is no correlation between binding free energy and the other characteristic of the interface such as polarity size of the interface overall, there is weak correlation with stability of the complex and size of the interface and hydrophobicity [28, 29]. Transient complexes especially enzyme-inhibitor complexes have polar surface patches and more hydrophilic compared to constitutive (obligatory) interfaces [5]. It is found that although the overall interfaces can be characterized as hydrophobic [30] for all types of complexes. In both transient and obligatory interfaces most of the interfaces are relatively planar and accessible. However, the complexes that are found in only complex state in nature are more hydrophobic [5].

The characteristics of the interfaces regulate the protein-protein interactions: hydrogen bond forming ability of interface residues, electrostatics interactions, Van der Waals interactions and hydrophobicity. Electrostatics and hydrogen bonding is crucial in stability of protein complexes [31, 32]. Besides the distinct characteristics of the protein interfaces, there are also differences in association of complexes in specificity point of view. It has been found that similar proteins can bind together in different ways. Another interesting result is that proteins that have different structures can bind in analogous ways. The unpredictable behavior of the protein interfaces and protein – protein interaction make this area an intense research target.

Molecular dynamical analysis of protein – protein interaction is popular because of the difficulty in protein – protein interactions experiments. Molecular dynamics (MD) is used to analyze only short time scales behavior of the protein – protein interaction because of the computational complexity. MD simulations are performed to analyze many aspects of protein dynamics such as energetic of the protein association [33, 34] and conformation studies on protein complexes [35-37]

2.3 Hot Spots

It has been observed that the contribution of each residue to the binding free energy of association is unevenly distributed across the interface [9, 10]. There are some regions and individual amino acids which contribute to binding free energy dominantly. These residues are called “hot spots”. By definition a hot spot residue contributes free energy of binding by more than 2 kcal/mol despite their small size with respect to rest of the binding site. The characteristics of of hotspots have been discussed in [10, 38, 39].

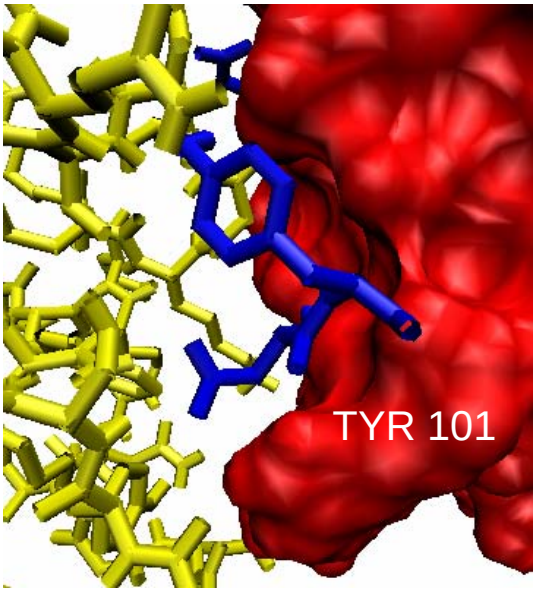


Figure 2.3 The complex of fv fragment of mouse monoclonal antibody d1.3 with hen egg lyzosyme, experimental hot spot TYR 101 extracted from Alanine Scanning Database [39]

The change in free energy of binding is measured experimentally with Alanine Scanning Mutagenesis. Hotspots’ replacement (with alanine) gives a distinct drop in the binding constant (typically tenfold or higher) and destabilizes the bound ensemble relative to the bound one. The compilation of data from alanine scanning database results interesting outcomes Figure 2.4 shows that although free energies of binding are well correlated with

buried surface area for interfaces as a whole, it can be seen that this does not hold at the level of individual side chains. The amino acid preferences of hot spots are represented in Figure 2.5. The distribution of hot spots is random according to the data except Trp, Tyr, and Arg being significantly enriched in hotspots.

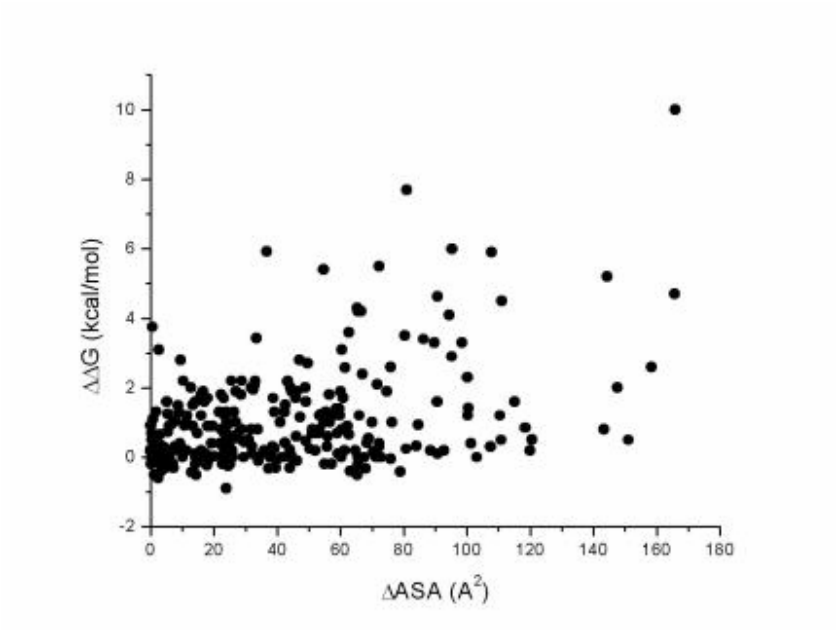


Figure 2.4 the side chain ΔASA on complex formation vs. ΔΔG on mutation to alanine with the same database of proteins used in the above graph [9].

Bogan and Thorn [9] proposed that there are energetically less important residues which surround hot spots. They prevent hot spots from bulk solvent so they increase the importance of electrostatic interaction between hot spots and their interacting partners by decreasing the dielectric constant.

Conservation studies are also important to characterize hot spot residues in the interfaces. It is known that the residues across the protein binding sites are more conserved compared to the rest of the protein surface residues because the interface residues have relatively slow conservation with respect to the rest of the interface [14]. Hence, one can

conclude that conserved residues across the interfaces can act as binding hot spots. It was showed that structurally conserved residues have distinct preferences between interface and rest of the protein [8, 39]. It is found that Phe, Trp and Met shows significant conservation across interfaces. However, there is no conservation on the rest of the surface of the protein.

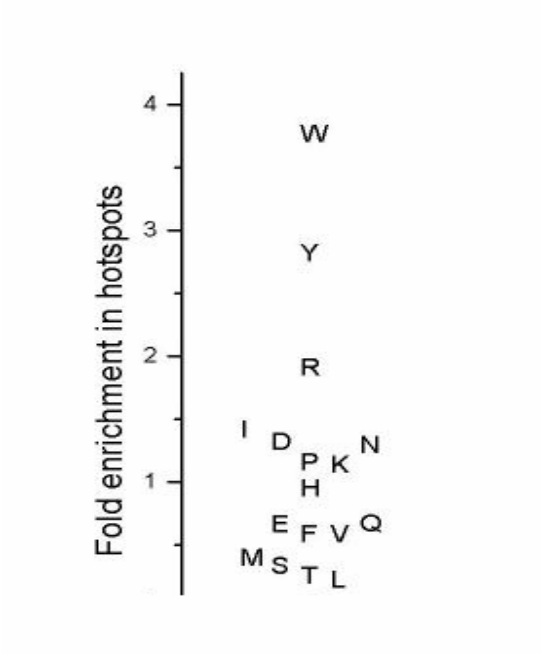


Figure 2.5 *Amino acid preferences in hotspots.* For each amino acid, the frequency of that amino acid in the database as a whole and in the subset with binding energies > 2 kcal/mol was determined [9].

The studies on hot spot predictions and detections have great importance because exploring hot spot across interfaces decrease the large surface area required to block undesired protein-protein interaction.

2.4 Prediction and Detection Methods

There are some experimental and computational methods developed for predicting the hot spots among the interfaces. Clackson and Wells [10] developed an experimental technique that is alanine scanning mutagenesis followed by the measurement of each mutant's effect on binding. A database of hotspots, Alanine Scanning Database (ASEdb) [40], has been compiled as an product of [9]. The ASEdb contains 2919 mutants of which 1953 have structural data for the monomer and 580 have structural data for the dimer.

An alternative computational technique, termed "computational alanine mutagenesis" has been introduced [12], that builds upon previous work in [41]. Kortemme and Baker [4] developed a model by examining the contribution of interface residues in free energy of binding. In this study, 79 % of all interface hot spots can be predicted using the free energy calculations including hydrogen bonds, implicit solvent and packing interaction ignoring changes in backbone conformation or effects on the dynamic interface.

Another computational method called free energy perturbation calculation [19] with the use of thermodynamic cycles and MM – PBSA (Molecular Mechanics Poisson – Boltzmann Surface Area) [33] to calculate free energy of binding. These methods can be classified as virtual mutagenesis studies. Another study on effect of mutations on the binding free energy change is done by Guerois and Serrano [42] by mutating large number of residues on monomeric proteins and protein – protein complexes. However, the high computational cost of such calculations led to newer approaches which require less computational power.

The conservation models are useful tools in predicting protein – protein interactions [38, 43]. The structural conservation information is used to identify hot spots in protein interfaces [39]. Keskin et al. (2005) used Multiprot [44] (multiple structural alignment

software) with a non redundant data set extracted from PDB database and detect structurally conserved residues across the interfaces and they conclude that structurally conserved residues contribute stability of the interfaces and play important roles in recognition, catalysis and binding. Keskin et al. (2005) found that computational hot spots are not distributed along protein interfaces equally and they are clustered into some region. These outcomes are analogous to the experimental results from alanine scanning database. There are conserved residues around hot regions. They prevent hot spots from bulk solvent so the interaction will be stronger [38].

Chapter 3

Materials

The protein complexes that are studied in this thesis obtained from Brookhaven Protein Data Bank (PDB) [18]. PDB is a data bank including three dimensional structural data of large molecules such as proteins and nucleic acids.

Table 3.1 List of 14 protein complexes used in this study.

Pdb Id	Description	Complex Types	Number of Residues
1vfb	Antibody / Lyzosyme	Antibody/Antigen	352
1fcc	Immunoglobulin/Antigen	Antibody/Antigen	262
1qnz	Anti-Hiv Antibody Complex	Antibody/Antigen	249
1cd0	Immunoglobulin	Homodimer	222
1mr8	Metal Transport	Homodimer	180
1bzd	Binding Protein	Homodimer	254
1cqk	Immune System	Homodimer	202
2sni	Subtilisin / Inhibitor	Enzyme/Inh.	339
1ugh	Glycosylase/Inhibitor	Enzyme/Inh.	305
1sbw	Hydrolase / Inhibitor	Enzyme/Inh	238
2btx	Toxin/Peptide	Protein/peptide	87
1ddh	Histocompatibility/Antigen	Protein/peptide	373
1bjr	Hydrolase/Iron Transport	Protein/peptide	289
1cjq	Ribonuclease / Peptide	Protein/peptide	116

The set of complexes is listed in the Table 3.1. There are four different groups and total 14 protein complexes used in the MD simulations: Homodimer, Enzyme- Inhibitor, Antibody – Antigen and protein – peptide. These are mainly dimers except Antibody –

Antigens complexes and one complex from protein – peptide is 1ddh. The selection of the complexes was performed according to their size, interface size and types in order to obtain variety. Among those criteria, the size of the protein complexes is the most crucial one considering the molecular dynamical study because their size brings computational complexity and time in our analysis.

3.1 Biological Context of Protein Complexes

In this section, the biological contexts of the complexes which are studied are explained briefly by mentioning only their main biological and functional features.

3.1.1 Antibody – Antigen Complexes

Antibodies are immune system-related proteins called immunoglobulin. Each antibody consists of two polypeptides – a heavy chain and a light chain. Moreover, their production is stimulated by bio-molecules called antigens in a living cell. Hence, their interaction has great importance for immune system. There are three antibody – antigen complexes used in this work.

PDB code 1vfb: the complex of fv fragment of mouse monoclonal antibody d1.3 with hen egg lyzosyme [45]

Function: Lyzosyme have primarily a bacteriolytic function; those in tissues and body fluids are associated with the monocyte-macrophage system and enhance the activity of immunoagents.

Resolution: 1.80 Å

Experimental method: X-ray diffraction

Biological unit: The macromolecule is an oligomer of type trimeric

PDB code 1fcc: Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG Immunoglobin. [46]
Function: The complex is used as a molecular biology reagent.
Resolution: 3.50 Å
Experimental method: X-ray diffraction
Biological unit: Macro molecule of type octameric

PDB code 1qnz: Crystal structure of anti-HIV antibody complex with the gp120 v3 peptide [47]
Function: This is an envelope protein that surrounds many types of virus particles.
Experimental method: NMR
Biological unit: The macromolecule is an oligomer of type trimeric

3.1.2 Homodimer Complexes

A homodimer can be defined basically as complex of two identical proteins. There are 4 different crystal structures of homodimers extracted from PDB for this study:

PDB code 1cd0: Crystal structure of human lamda-6 light chain dimer. [48]
Function: This protein is an important factor in the pathogenesis of amyloidosis.
Resolution: 1.90 Å.
Experimental method: X-ray diffraction.
Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 1bzd: Crystal structure of transthyretin [49]
Function: Thyroid hormone-binding protein. Probably transports thyroxine from the bloodstream to the brain.
Resolution: 1.90 Å
Experimental method: X-ray diffraction.
Biological unit: Macromolecule of type tetrameric.

PDB code 1cqk: Crystal structure of the immune system protein of ch3 domain from the mak33 antibody[50]
Function: This protein was crystallized because of it is the simplest model system for studying immunoglobulin folding.
Resolution: 2.20 Å
Experimental method: X-ray diffraction.
Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 1mr8: Crystal structure of the Migration inhibitory factor-related protein 8 from human [51]
Function: This is a metal transport protein which is selectively interacting with the Ca ion.
Resolution: 1.90 Å
Experimental method: X-ray diffraction.
Biological unit: Macromolecule of type dimeric.

3.1.3 Enzyme – Inhibitor Complexes

Enzymes are proteins produced by living organisms and function as biochemical catalysts. Enzyme inhibitors are also proteins that prevents enzyme from functioning properly. Enzymes are important bio-molecules in every level of cell function. Therefore, there are three different enzyme –inhibitor complex included in this work:

PDB code 2sni: Crystal structure of subtilisin novo complex with chymotrypsin inhibitor [52].

Function: Subtilisin is classified as Protease which plays role in hydrolysis of proteins and its inhibitor

Resolution: 2.10 Å

Experimental method: X-ray diffraction.

Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 1sbw: Crystal structure of mung bean inhibitor lysine active fragment complex with bovine beta-trypsin [53]

Function: This enzyme is an hydrolase which means active in hydrolysis reaction.

Resolution: 1.80 Å

Experimental method: X-ray diffraction.

Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 1ugh: Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor [54]

Function: This protein binds specifically and reversibly to the host uracil-DNA glycosylase, preventing removal of uracil residues from PBS2 DNA by the host uracil-excision repair system.

Resolution: 1.90 Å

Experimental method: X-ray diffraction.

Biological unit: The macromolecule is an oligomer of type dimeric.

3.1.4 Protein –Peptide Complexes

The fourth group of interacting complexes is protein – peptide complexes. Peptides are small fragments of proteins and there are numerous interactions between proteins and peptides. The following four complexes are selected to represent this class of interaction:

PDB code 1cjq: crystal structure of a hydrolase which is ribonuclease with a peptide having 15 residues [55].
Function: Endonuclease that catalyzes the cleavage of RNA on the 3' side of pyrimidine nucleotides. Acts on single stranded and double stranded RNA.
Resolution: 3.00 Å
Experimental method: X-ray diffraction.
Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 1ddh: Mhc class i h-2dd heavy chain complexed with beta-2 microglobulin and an immunodominant peptide p18-i10 from the human immunodeficiency virus envelope glycoprotein 120 [56].
Function: Beta-2-microglobulin is the beta-chain of major histocompatibility complex class I molecules. This is a transmembrane protein which functions in the presentation of endogenous antigen to the immune system.
Resolution: 3.10 Å
Experimental method: X-ray diffraction.
Biological unit: The macromolecule is an oligomer of type trimeric.

PDB code 1bjr: Complex formed between proteolytically generated lactoferrin fragment and proteinase [57]

Function: Hydrolyzes keratin at aromatic and hydrophobic residues.

Resolution: 2.44 Å.

Experimental method: X-ray diffraction.

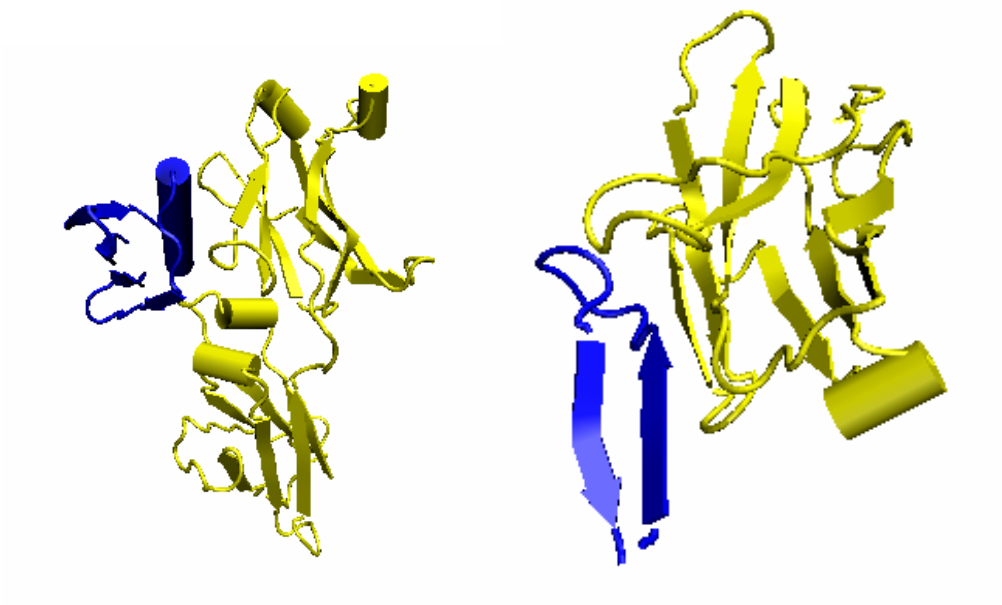
Biological unit: The macromolecule is an oligomer of type dimeric.

PDB code 2btx: NMR Structure Of The Complex Of α-Bungarotoxin With A Library Derived Peptide [58]

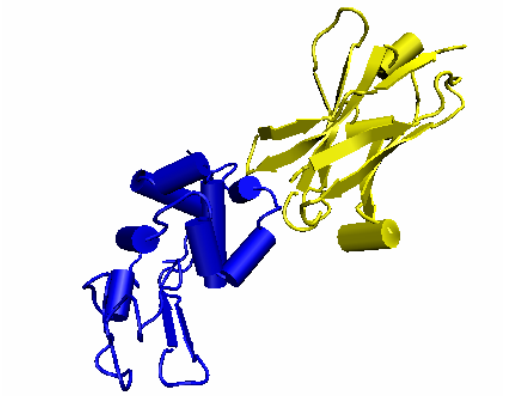
Function: Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. Binds to muscular and neuronal (alpha-7, alpha-8, and alpha-9) nicotinic acetylcholine receptors.

Experimental method: NMR.

Biological unit : The macromolecule is an oligomer of type dimeric.

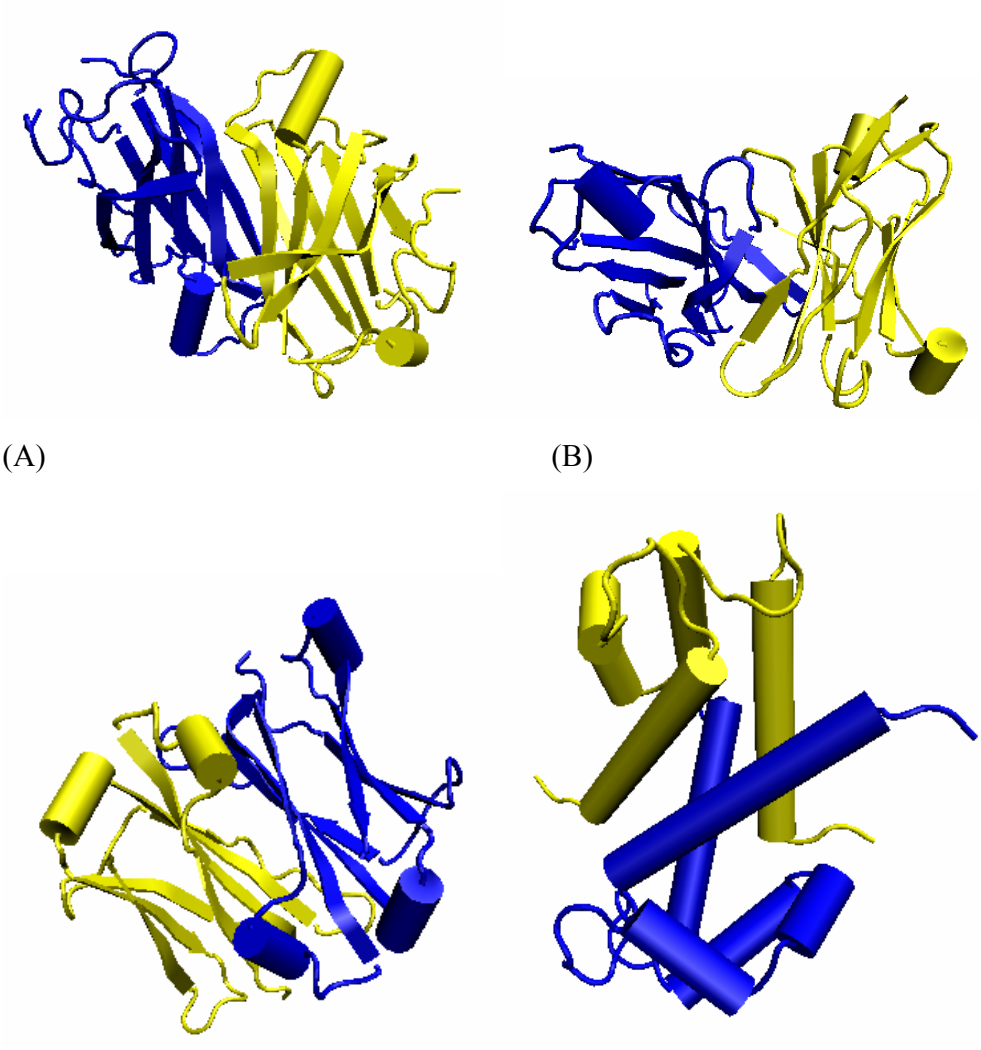


(A) (B)



(C)

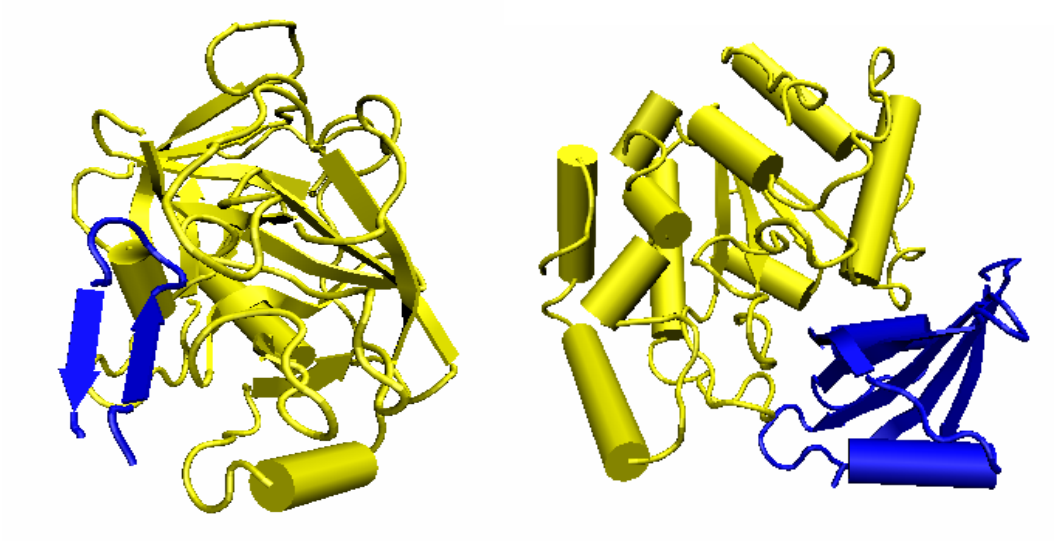
Figure 3.1 Crystal structures of antibody – antigen complexes, figures are drawn by using visual molecular dynamics package (VMD). Only the interacting chains are presented. (A)1fcc (B)1qnz (C)1vfb



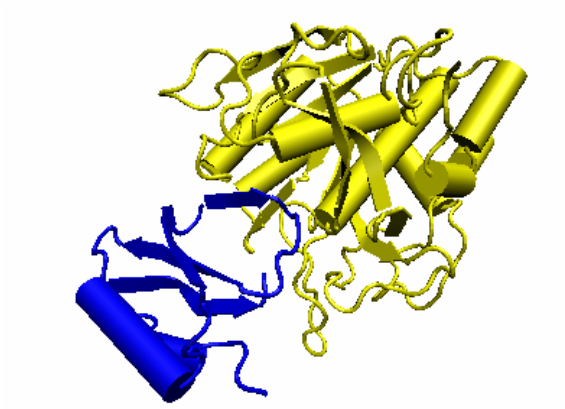
(A) (B)

(C) (D)

Figure 3.2 Crystal structures of homodimer complexes. (A) 1bzd (B) 1cd0 (C) 1cqk (D) 1mr8

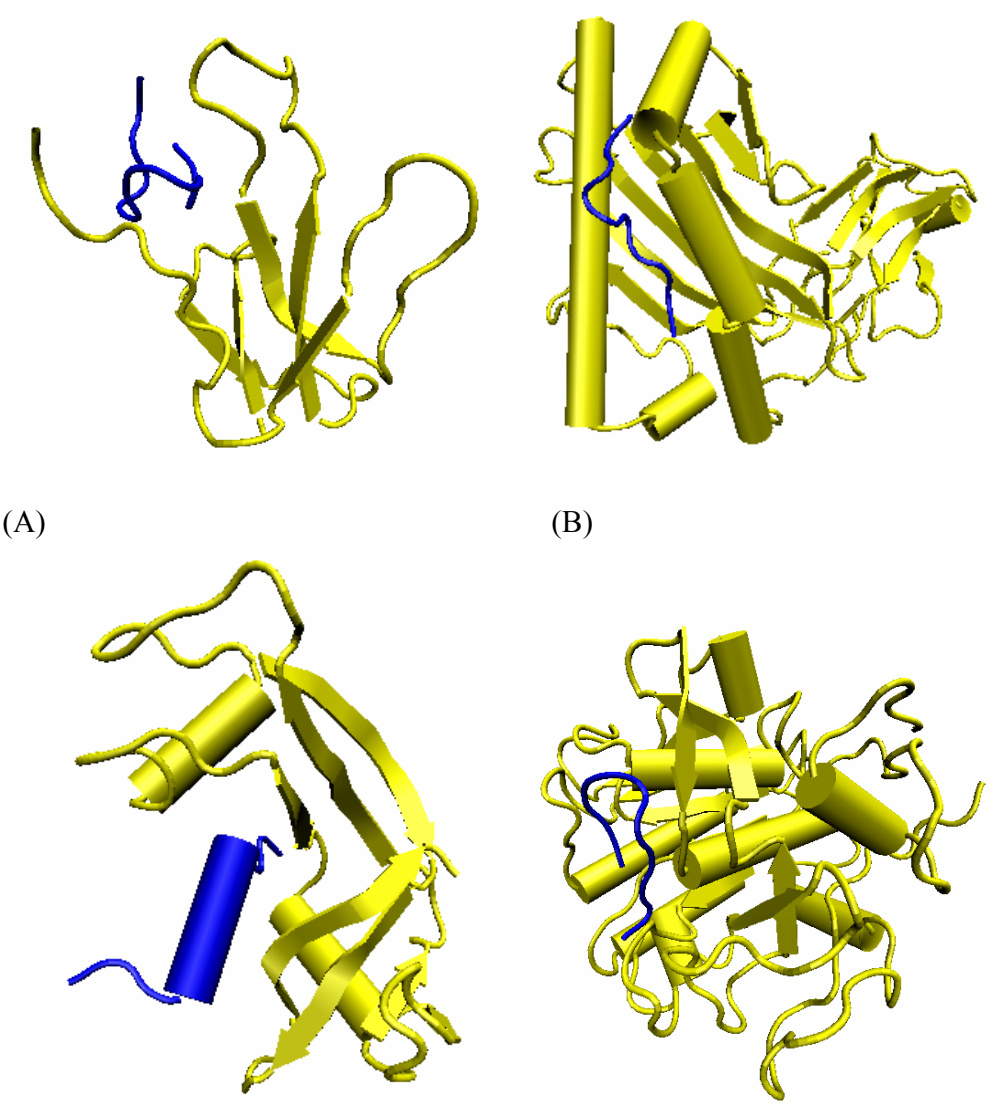


(A) (B)



(C)

Figure 3.3 crystal structures of enzyme - inhibitor complexes (A) 1sbw (B) 1ugh (C) 2sni



(A) (B)

(C) (D)

Figure 3.4 crystal structures of protein – peptide complexes. (A)2btx (B)1ddh (C)1cjq (D)1bjr

Chapter 4

Molecular Dynamics Simulation

4.1 Molecular Dynamics

Molecular dynamics (MD) is an established method of studying the motions of individual atoms or molecules in a system. MD studies the position of particles in the systems with time and produces a trajectory for a system of N interacting particles. Molecular dynamics simulations have been used extensively over the past 25 years for modeling ever large and complex molecules for long periods of time. A number of studies have successfully applied MD simulations to understand the atomic structure and function of proteins in solution, exploring the energetic and conformations of active sites as well as the effects of mutations of the amino acid structure to the function of proteins [59]. Detailed reviews and studies on the principles of molecular modeling are available in [60, 61].

It has become of critical importance to make certain that bio-molecules are modeled in a realistic environment with molecular modeling being applied more and more in life sciences. There are several methods that are improved and developed to provide the required conditions; force fields, periodic boundary conditions, reliable salvation models and computational methods for maintaining physical conditions of simulating system. As such, advances over recent years have led to development of sophisticated methods for incorporating solvents into the model, while simultaneously minimizing the numerical effects of boundary conditions from the edges of a solvent box. The underlying force fields have been developed to better account for solvent polarization effects, and long-range electrostatic non-bonded interactions.

MD is a method for computing equilibrium and kinetic properties of the systems, that obey the laws of classical (Newtonian) physics. In principle, any process with the characteristic timescale $\tau > 10^{-13}$ seconds can be described using classical physics. Therefore, except for the fluctuations in bond lengths and angles, all other motions in bio-molecular systems can be treated as classical. Because MD is based on Newtonian dynamics, MD simulations are deterministic. This means that, in principle, once initial conditions are given, the past or future time evolution can be derived. The acceleration, the second derivative of the distance (r_i''), of a particle with mass m is depend on the force F applied on this particle. The force is determined by differentiating the energy potential energy (E_p). Therefore, there is a connection between negative gradient of the energy function and molecular motion.

$$F = m \times r_i'' = -\frac{\partial E_p}{\partial r_i} \tag{4.1}$$

Where F: is the force acting on a particle, m: mass of the object, r_i'': acceleration, E_p: potential energy and r_i: the displacement.

The implementation of MD algorithm consists of the following steps:

1. Setting parameters required for molecular dynamics, for instance coordinates of the molecule, temperature and pressure etc.
2. Initialization of the velocities by using appropriate distribution function such as Maxwell – Boltzmann velocity distribution depending on the temperature of the molecular dynamics simulation.
3. Minimization of the system with the initial configuration of the system in order to keep the system stable during molecular dynamics simulation.
4. Computation of forces by taking negative gradient of the energy function.
5. Integration of Newton's equation of motion.
6. Repeating step 3 and step 4 until the required simulation time reached.

7. Computing the ensemble averages by using output of the molecular dynamical simulation.

There is a large number of molecular dynamics software available in the market; the most popular simulation packages include AMBER [62], CHARMM [63], GROMACS [64], and NAMD[6]. Throughout this work NAMD is used for molecular dynamics simulation. NAMD is designed for large bio-molecular systems so that it has all of the functionality for simulating proteins and protein – protein interactions. Moreover, it is possible to configure NAMD to be run on high performance computing platform that uses multiple processors.

In the following sections important concepts of molecular dynamics simulations such as force field, salvation, periodic boundary conditions and thermodynamic considerations are discussed. Then the MD simulation procedure that is applied in this thesis is explained.

4.1.1 Force Field

Theoretical studies of biological molecules permit the study of the relationships between structure, function and dynamics at the atomic level. Since many of the problems that one would like to address in biological systems involve many atoms, it is not yet feasible to treat these systems using quantum mechanics. However, the problems become more tractable when empirical potential energy functions that are much less computationally demanding are used. However, numerous approximations are introduced that lead to certain limitations. Force fields offer a compromise energy functions that are derived from classical physics and experimental results. They provide information on what conformations of the molecule are better or worse. Namely the lower the energy value, the better the conformation is.

Current force fields models provide a reasonably good cooperation between accuracy and computational efficiency. They are often calibrated to experimental results and quantum

mechanical calculations of small model compounds. Their ability to reproduce physical properties measurable by experiment is tested; these properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy and inelastic neutron scattering and thermodynamic data. There are many constants, or parameters, involved in the energy functions. Finding these parameters is a major part of developing an accurate and computationally favorable energy function. This is an area of continuing research and many groups are working to derive functional forms and parameters for potential energy functions of general applicability to biological molecules. Among the most commonly used potential energy functions are the AMBER [65], CHARMM [63], GROMOS [66]and OPLS/AMBER [67] force fields. The continuing development of force fields remains an intense area of research with implications for both fundamental researches as well as for applied research in the life sciences.

CHARMM [62] (Chemistry at Harvard Macromolecular Mechanics) refers to both a program for macromolecule dynamics and mechanics and the energy function developed for use in the program. CHARMM is a popular force field used mainly for the study of macromolecules. It is an all-atom force field that considers all atoms as opposed to only considering polar hydrogens. In recent version, the parameters were created using experimental data and supplemented with ab initio results. CHARMM potential function is in the following form (Eq 4.2):

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_o)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_x (1 + \cos(nx - \delta)) +$$
$$\sum_{impropers} K_{imp} (\phi - \phi_0)^2 + \sum_{nonbond} \varepsilon \left[\left(\frac{R_{\min_{ij}}}{r} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \qquad (4.2)$$

In the Eq (4.2), the first five terms, except second one Urey Bradley 1,3 nonbonded interaction, is used to represent bound interaction in the system. However the last two terms are for non bonded interaction, namely Lennard Jones and electrostatic interactions. The first term in Eq. (4.2) represents the potential due to length variations of th bond connecting two atom, which is usually modeled as a simple harmonic spring. K_b, is the force constant describing stiffness of the bond b_i and b_0 are the instantaneous and equilibrium lengths of the bond respectively. The third term is to represent bond – angle potentials. As similar to the first term, it is modeled as simple harmonic spring where K_θ is the spring constant and θ and, θ_0 is the instantaneous and equilibrium angle value, respectively. Fourth term is the dihedral (torsional) potential which plays a crucial role in secondary structure of proteins. $K_{x,}$ is barrier height and n is the integer which determines the periodicity of the potential. Last term in bonded interaction part is the improper potential part. Improper dihedral angles determine the correct geometry of atoms. K_{imp} represents the stiffness of the potential and the Φ and Φ_0 are the instantaneous and equilibrium values of improper angle.

There are three different types non bonded interaction includes Urey Bradley, van der Waals and electrostatic interaction [62]. The first term in non bonded interaction is Urey Bradley potential which defines the potential between first and third atom defining an angle. K_{UB} is the equilibrium constant given by parameter statements and S and S_0 are the distance between these two atoms in normal state and equilibrium state respectively. The common form of van der Waals potential is given by Lennard –Jones in the sixth term in Eq. (4.2), where ε is the Lennard Jones well depth $Rmin_{ij}$ is the location of the Lennard Jones energy minimum and R is the radius of the atom. Last terms express the coulombic interaction between two charged particles in which q_i and q_j are the atoms charge ε_1 is the effective dielectric constant, r_{ij} is the distance between atoms.

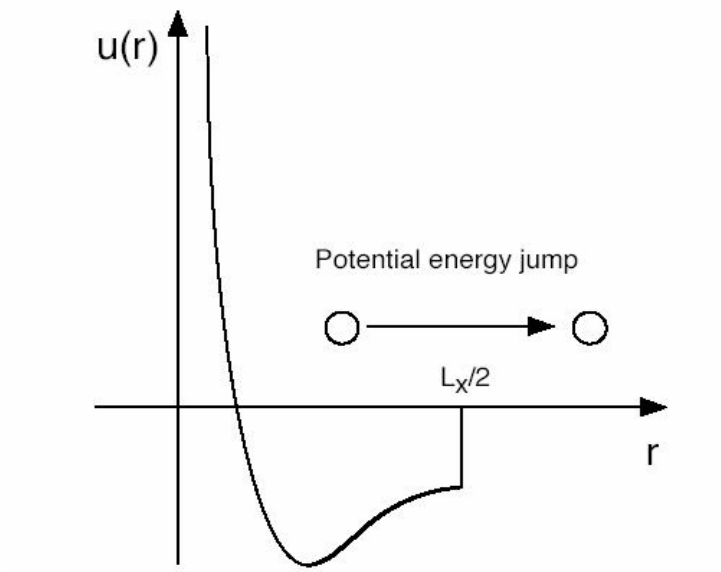


Figure 4.1 potential function vs. distance, $L_x/2$ is the cut off distance

Shifting function is an important concept if the non bonded interactions are concerned. In a simulation using periodic boundary condition, minimal image convention should be seriously tajen in tı account. The cut off distance should be half of the periodic box in order to prevent system from interacting with itself. However, this truncation causes discontinuity in potential energy function, so the potential function can not be differentiated. Therefore, a cut off length $r_c < L_x/2$ is introduced and the potential energy functions are modified as follows:

$$U'(r) = \begin{cases} U(r) - U(r_c) - (r - r_c)\left.\frac{du}{dr}\right|_{r=r_c} & r < r_c \\ 0 & r >= r_c \end{cases} \quad (4.3)$$

For the Lennard-Jones potential it is conventionally taken as $r_c = 2.5\sigma$, the illustration of the shifting function of Lennard-Jones is in Figure 4.2.

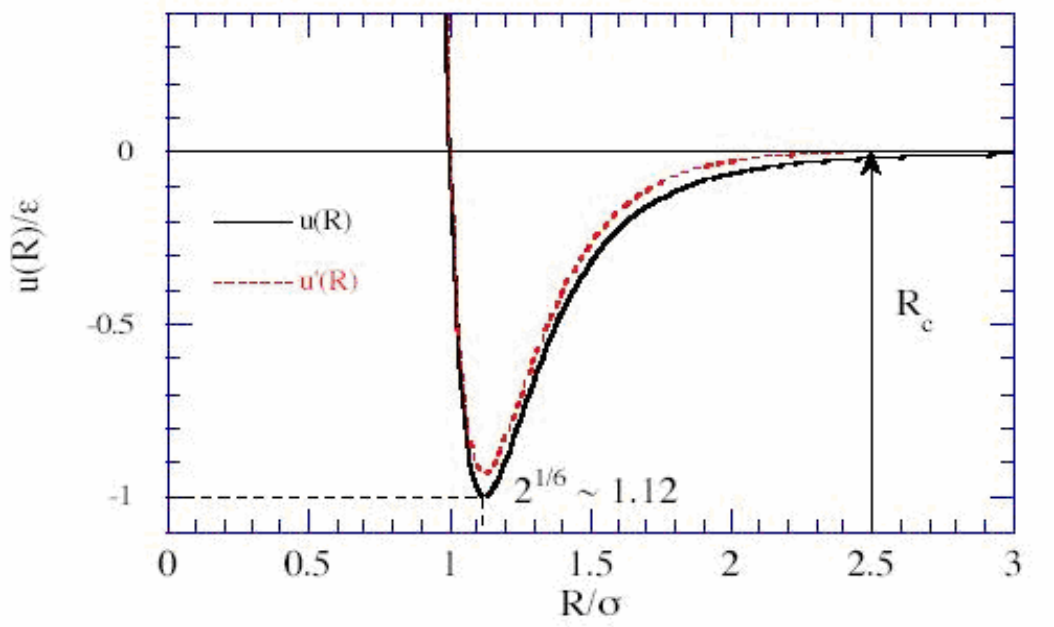


Figure 4.2 Truncated Lennard-Jones Potential

4.1.2 Minimization of the system

Before starting a molecular dynamics the complete system should be minimized in order to eliminate as many poor contact as possible. Minimization of the system is barely means energy minimization. There are number of minima in the energy landscape of simulating complex, especially for bio-molecules. The goal of the energy minimization is to find the local energy minimum. Although the observed local minimum is much higher than the global minimum, it is still a good starting point for initial conditions. Initial conditions of the molecular dynamics system composed of three different quantities, mass, coordinates and velocity. In MD simulations with proteins, the crystal structures in PDB database can be used. The initial velocities of the atoms set according to Maxwell –Boltzmann given in Eq. (4.3) velocity distribution which depends on the temperature of the system.

$$P(v_{i,\alpha}) = \left(\frac{m}{2\pi k_b T} \right)^{1/2} e^{\frac{m v_{i,\alpha}^2}{2 k_b T}} \tag{4.4}$$

4.1.3 Integration Algorithms

Molecular dynamics simulations characterize the physical system by means of potential energy functions. In each time step force is calculated as a negative gradient of the potential energy in Eq (4.1) in order to find the position and velocity of each atom in the next simulation step. The numerical solution of these systems requires finite difference integrators. The most common class of integrators are Verlet integrator [68]. Verlet integrator calculates velocity and coordinates by expanding governing equations according to Taylor expansion. The following two equations simply show the Verlet calculation scheme:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2} a(t) \Delta t^2 \tag{4.5}$$

$$v(t + \Delta t) = v(t) + \frac{1}{2} \Delta t (a(t) + a(t + \Delta t)) \tag{4.6}$$

Where r represents displacement, v is velocity, a is acceleration, t is the time and Δt is the time step.

These calculations carried out according to the finite different schemes as mentioned before. Hence, there is always an approximation for system during calculations. Therefore a good integrator should fulfill the following requirements [68]:

- It should approximate the trajectory very well
- It should be stable and conserves the energy even there is perturbations it should not lose its stability
- It should be robust in the sense that it allows long time step simulations.

4.1.4 Periodic Boundary Conditions

It is necessary to determine the conditions imposed at the simulation boundaries, as with any simulation conducted within a restricted volume. This is particularly true in the presence of implicit solvent, where the simulation volume may need to be kept quite limited for computational performance purposes. In such conditions, most of the solvent molecules will be at or near to the volume boundary. Inflicting reflective boundary conditions (i.e. creating solid walls that molecules "bounce" off) tends to lead to large regions of immobile water molecules, providing a poor illustration of the biological environment.

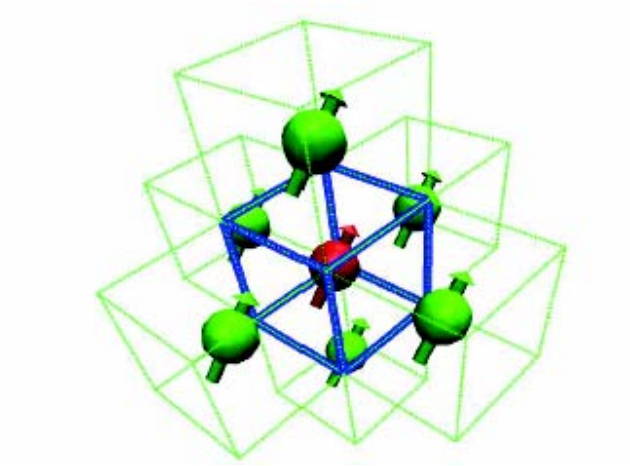


Figure 4.3: A graphical representation of periodic boundary conditions.

An alternative approach is to apply periodic boundary conditions in which molecules leaving the system volume at one boundary immediately enter (maintaining a constant momentum) at the opposite boundary. This effectively creates an infinite array of repeating systems side by side allowing water molecules to exist in a bulk state in Figure 4.3. The primary simulation box is shown together with its nearest-neighbor images. The application of periodic boundary conditions is requested primary care about the location of the system that one would like to simulate. To avoid a solute interacting with its own image it is apparent that the cutoff of all non bonded interactions must be smaller than the distance

between the edge of the solute and the nearest edge of its nearest image. This kind of representation increase computational complexity but it enables more accurate representations of the biological environment.

4.1.5 Solvation and Water Models

MD simulations should include the effects of solvent on the system in order to accurately represent the biological environment in which many systems under study. An accurate representation be done explicitly with the insertion of large quantities of solvent molecules, or implicitly by modifying the simulation to account in an estimated way for the effects of the solvent on the system. Such solvent effects include the damping of electrostatic interactions, especially in the case of water (due to its high dielectric constant), as well as frictional drag and hydrogen bonding effects.

Implicit models can account for the first two of these effects, but are unable to account for hydrogen bonding that is often important driving force for solute conformational change. Hence, as computing capability has steadily increased, so explicit solvent models have become the standard for biological simulations, with water understandably being the most extensively used solvent. A variety of models exist to represent water; these include the extensively used models are TIP3P, TIP4P, PPC and TIP5P. Figure 4.4 shows the various configurations these models adopt. For a brief overview see [69];for a detailed review of the field see [70].

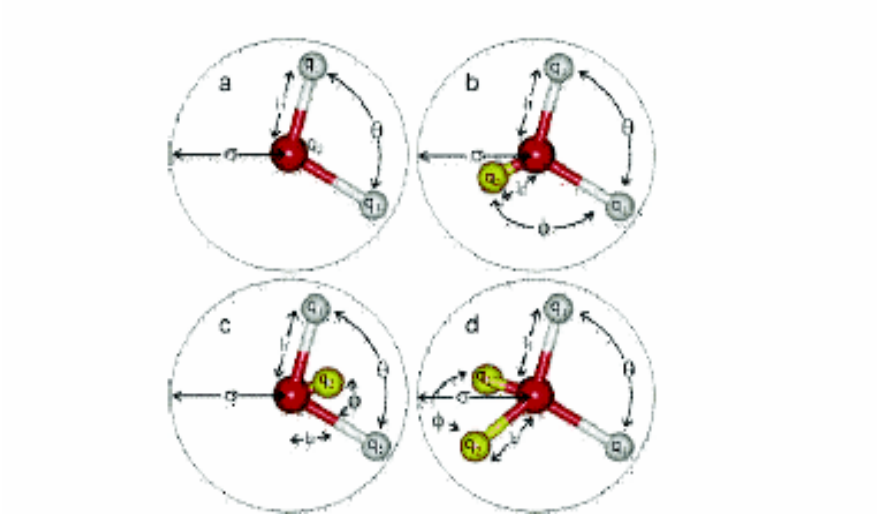


Figure 4.4: A graphical representation of selected explicit water model**.** Models differ in the quantity, positions and magnitudes of the partial charges used to represent the electrostatic and dipolar effects of water. (a) TIP3P (b) PPC (c) TIP4P (d) TIP5P. Image extracted from [69] where numeric data is available

4.1.6 Thermodynamic Consideration

Molecular dynamics simulations can be executed in a variety of thermodynamic ensemble. Common thermodynamic ensembles include the microcannonical (NVE) ensemble in which the number of particles, system volume and total system energy are held constant, as well as the NVT and NPT ensembles, both of which maintain a constant temperature, one at constant volume and one at constant pressure, respectively. NVT and NPT ensemble is widely used in simulation of bio-molecular systems. Considering the computational convenience, the canonical ensemble (NVT) has been preferred in Molecular Dynamics. The temperature was kept constant by using Langevin Dynamics scheme, which is already implemented in NAMD. The Langevin equation is a stochastic differential equation in which two force terms have been added to the second law of Newton to

approximate the effects of neglected degrees of freedom. This consists of adding a random force and subtracting a friction force from each atom during each integration step.

4.2 Standard Protocol Applied in this Study

In this study, MD simulations are performed in order to incorporate sequence information with the structural and dynamical information of protein complexes. The MD simulations were carried out with the molecular dynamics simulation package NAMD [71]with CHARMM [63]force field parameters (PARAM27). First, the protein complex was solvated with the solvate plug-in of VMD [72](visual molecular dynamics) package. This plug-in uses TIP-3 water molecules in order to solvate the protein. Particle Mesh Ewald (PME) [73]method is applied to the simulation in order to decrease memory requirement and make the simulation computationally more efficient. Ewald summation was introduced as a technique to sum the long-range interactions between particles and all their infinite periodic images efficiently. Particle Mesh Ewald (PME) is an efficient full electrostatics method for use with periodic boundary conditions. PME divides the potential energy into Ewald's standard direct and reciprocal sums and uses the conventional Gaussian charge distributions. PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus substantially reducing memory requirements. This method is readily implemented in NAMD package. Particle-mesh Ewald (PME) summation requires the system to be electrically neutral. In order to neutralize the system, we used VMD auto-ionize plug-in that adds sodium and chlorine ions to the system. This feature of VMD auto-ionize plug-in distributes ions randomly but the user can define minimum distances between ions and molecules as well as between ions. In the MD simulation, NVT ensemble and periodic boundary conditions with a rectangular box are applied. The temperature of the simulations was kept constant at 300 K by using Langevin Dynamics. Initial velocities were generated randomly from a Maxwell distribution at 300 K in accordance with the masses assigned to

the atoms. The time step was 2 fs. Simulations were performed on the independently crystallized structures by using as initial condition the unbound x-ray structure of the protein complexes Trajectories were sampled at 40-ps intervals. Initial equilibration was done for 10,000 steps, followed by at least 6-ns production runs. The simulations were carried out in a Linux-based cluster from Racksaver cluster and each node has two 3.06 GHz Intel Pentium Xeon processors and Beowulf Cluster with nodes having Intel Pentium 4 2.4 GHz processor.

Chapter 5

Results and Discussions

5.1 Extraction of Interface Residues

In this thesis 14 different protein complexes are studied from composed of 4 main complex types: 4 homodimer, 4 protein – peptide, 3 antibody – antigen and 3 enzyme inhibitor complexes. By means of their accessible surface area (ASA) calculations, done by NACCESS, the interface residues on the crystal structure of the complexes are defined. A probe of given radius is rolled around the surface of the molecule, and the path traced out by its centre is the accessible surface. In general, the probe has the same radius as water (1.4 Angstroms) and hence the surface described is often referred to as the solvent accessible surface. The details of the calculation can be found in Appendix A-2. The number of residues in the interface between complexes differs in numbers and in accessible surface areas. The number of residues in the interfaces, i.e. interface sizes, vary between 25 and 73 and the total accessible surface are as of the interfaces are between 800 $Å^2$ and 2597 $Å^2$ as listed in Table 1. The interface ASAs, the third column of the table, are calculated as the cumulative sum of the ASA of interface residues in the complex form which is exactly the same for hot spot ASA calculations. There is a linear correlation with a correlation constant 0.92 between interface size and the accessible surface area. It shows that the larger the number of residues in the interface, the greater the ASA of the interface is (Fig 5.1).

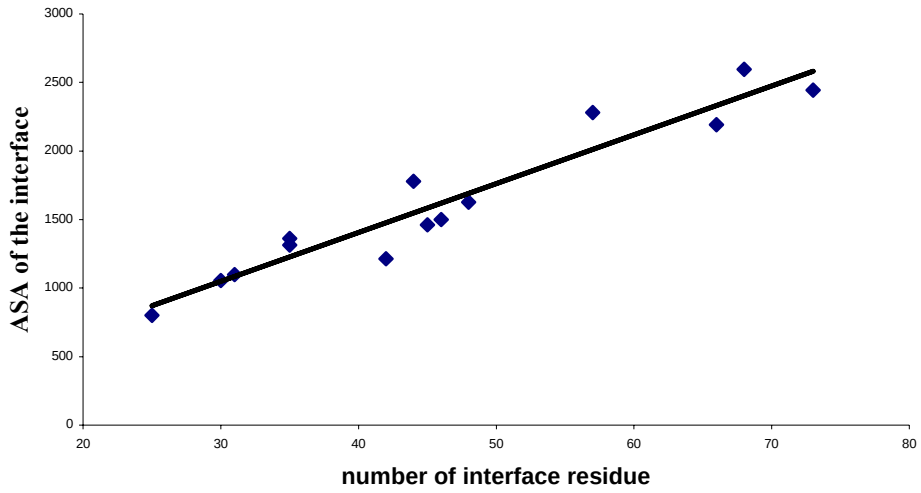


Figure 5.1 Number of interface residue vs. ASA of the interface of the protein complexes

5.2 The residue preferences in the interface and the characteristics of the interface residues

Table 5.1 lists the percentages of the different characteristics of amino acids in interfaces for different types of interfaces. Further hydrophobic (A,P,L,I,M,V), charged (D,E,R,K), polar (N,Q,S,T), and aromatic residue (W,Y,F,H) percentages are given in the table. Clearly, interfaces are dominated by hydrophobic residues in all three cases. Next, it is mostly aromatic residue contribution.

Table 5.1 The interface characteristics of different types of complexes, the values are determined according to following criterias

INTERFACE TYPE	INTERFACE
HOMODIMER	charged : 15.27 aromatic : 23.64 Polar : 30.54 hydrophobic : 30.54
ENZYME/INHIBITOR	charged : 15.44 aromatic : 18.38 polar : 32.35 hydrophobic : 33.82
PROTEIN/PEPTIDE	charged : 24.68 aromatic : 18.98 polar : 26.58 hydrophobic : 29.74
ANTIBODY/ANTIGEN	charged : 25.31 aromatic : 21.51 polar : 30.37 hydrophobic : 22.78

The hot spot data from experimental alanine scanning database were compiled and it was found that Trp, Arg, Tyr, Asp, Pro and His were enriched in the interfaces as hot spots [9]. Hu et. al. made a systematic structural analysis over 97 protein – protein interfaces and they showed that polar residues His, Asn, Gln, Thr and Ser and partially polar residue Phe has superiority in the interfaces [13]. In our dataset we found that Tyr, Ser, Phe, Gly and Thr existed more frequently than the other residues in the interfaces. It is consistent with the other observation done over larger interfaces and experimental results. Moreover, we know that polar residues preferentially conserved in the protein – protein interfaces. It is obvious to have Phe, which is partially polar, Ser and Thr more frequently in the interfaces. It is also remarkable that Tyr is also supplemented in our data set which is strongly correlated with the experimental data.

Table 5.2 Hot spot residue preferences in each type of protein complexes.

Complex Type	Hot Spot Preferences
Homodimer	Tyr, Ser and Phe
Enzyme / Inhibitor	Gly and Ser
Protein / Peptide	Arg and Thr
Antibody / Antigen	Gly and Leu
All Complexes	Tyr, Ser, Phe, Gly and Thr

5.3 Identification of hot spots

The hot spot residues are identified by considering residue conservation during evolution. Evolutionarily conserved residues are critical for the function and stability of the complexes. Ma et al. [8] have found that binding sites are well conserved structurally during evolution with respect to the remainder of the protein surface. Hu et al. [13] and Ma et al. [38]have shown that alanine scanning mutagenesis data are well correlated with the residue conservation. Similarly sequence alignment is used in extracting the hot spot residues. CONSURF server is used to identify hot spot residues in this study. It maps the evolutionarily conserved regions on the surface of proteins of known structure by using phylogenetic tree information. According to this approach, there are total 153 computational hot spot residues predicted among the 617 interfaces residues. The fifth column of Table 5.3 shows the number of hot spot residues in the interfaces.

Table 5.3 the list of PDB ID of protein complexes and their corresponding information about interfaces and hot spots located in those interfaces. The first column is the PDB ID of the protein complexes, the second is the total number of interface residues in two chains of complex, and the third column is the cumulative total of the ASAs of interface residues in complex form. Fourth column is the total number of hot spot residues identified by CONSURF in corresponding complexes. Last column is the sum of the ASAs of hot spot residues in complex form.

Pdb Id	Interface Size	Interface ASA(Å^2)	Number Of Hot Spots	Hotspot ASA(Å^2)
1vfb	21	791.16	2	90.46
1fcc	35	1114.90	3	35.98
1qnz	31	1436.17	2	26.76
1cd0	46	1736.40	12	333.49
1mr8	68	2026.64	14	211.93
1bzd	44	1210.43	22	419.39
1cqk	57	2096.21	15	333.24
2sni	48	1253.63	19	254.64
1ugh	66	2091.12	20	209.10
1sbw	45	1287.23	10	204.33
2btx	30	2011.34	4	363.89
1ddh	49	1509.21	7	145.58
1bjr	42	888.89	17	146.23
1cjq	35	1343.24	6	120.23

Table 5.4 the table of the number of computational hot spots predicted in their corresponding interfaces. The second column is the general description of complexes extracted from PDB. Third column represents the type of each complex that belongs to. Fourth column is total number of residues in protein complex and last column represents the number of hot spot residues.

Pdb Id	Description	Complex Types	Complex Size	Number of Hot Spots
1vfb	Antibody / Lysozyme	Antibody/Antigen	352	2
1fcc	Immunuglobulin/antigen	Antibody/Antigen	262	3
1qnz	Anti-Hiv Antibody	Antibody/Antigen	249	2
1cd0	Immunoglobulin	Homodimer	222	12
1mr8	Metal Transport	Homodimer	180	14
1bzd	Binding Protein	Homodimer	254	22
1cqk	Immune System	Homodimer	202	15
2sni	Subtilisin / Inhibitor	Enzyme/Inh.	339	24
1ugh	Glycosylase/Inhibitor	Enzyme/Inh.	305	20
1sbw	Hydrolase / Inhibitor	Enzyme/Inh	238	10
2btx	Toxin/Peptide	Protein/peptide	87	4
1ddh	Histocompatıbility/Antigen	Protein/peptide	373	7
1bjr	Hydrolase/Iron Transport	Protein/peptide	289	17
1cjq	Ribonuclease / Peptide	Protein/peptide	116	6

The list of hot spot residues in individual complexes is given in Appendix part A-3. In this table, the first column shows the residue numbers and the type of the residues and the “tick” sign in the second column indicates that the residue is predicted as hot spot residue.

5.4 Hot spots are buried and tightly packed

Since hot spots contribute dominantly to protein–protein interactions, the question arises as to whether their number increases with the interface size, or is limited [38]. Figure 5.2(A) indicates a relation between the number of hot spots and the number of interface residues in the interfaces. There is a relationship between the number of hot spots and the interface size by looking at our data and the correlation constant is 0.91 as a result of linear regression if the three complexes are not considered: 1bzd, 2sni and 1bjr. These complexes have more conserved residues than the other complexes. Hence, it was expected that conserved residues would appear at a much higher frequency at the interface compared to non conserved residues.

When we analyze the accessible surface areas rather than the number of hot spots, we observe more buried hot spots. Figure 5.2(B) provides the cumulative sum of the ASAs of hot spots versus the ASAs of the interfaces they belong to. The calculations are detailed in Appendix A-2. The hot spot ASAs are obtained from the complex form of the interfaces. As the interfaces get larger, the ASA of the total hot spots do not increase linearly. The graph suggests that the hot spots are more buried, since the ASA does not increase linearly with the interface for large interfaces.

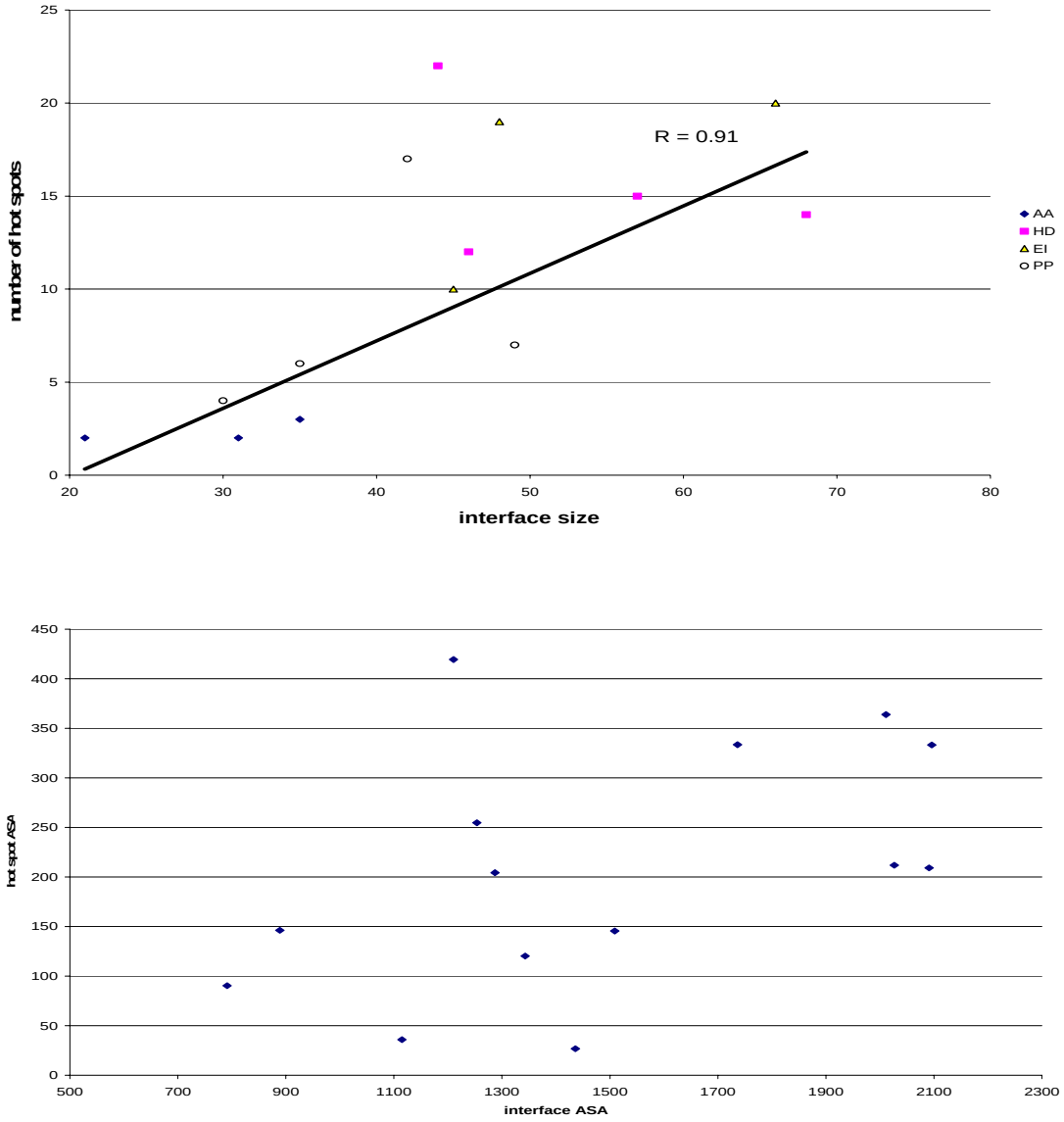


Figure 5.2 (A) Correlation between the numbers of interface residues in the interfaces with the number of hot spots. (B) Correlation between the accessible surface areas (ASA) of interfaces with the accessible surface areas of the hot spot residues.

Table 5.5 ASA calculations for over all complexes and individual complex types. ASA values are from the complex form. ASA values are given as per resiudue.

Complex type	Hot spot ASA (A^2)	Non Hot Spot ASA(A^2)
All	21.9	32.7
Homodimer	26.1	32.2
Enzyme /Inhibitor	10.9	35.18
Antibody / Antigen	14.8	28.9
Protein / Peptide	28.06	36.89
P -value	0.022	

The average accessible surface area of hot spot and non hot spot residues are represented for all complex types and overall. The hot spot residues in all cases have less ASA than the non spot residues. In order to understand the significance of the difference between hot spot residues and non hot spot residues paired t-test is applied on the above table. The p-value (0.022) was found well under the α – value (0.05). Therefore, the difference is statistically significant for all types of interfaces.

5.5 Molecular dynamics results

The molecular dynamics (MD) simulations were performed for the 14 different protein complexes, which are composed of 11 dimers and 3 trimers, and for their interacting monomers separately. In the MD simulations, NVT ensemble and periodic boundary conditions with a rectangular box are applied. The temperature of the simulations was kept constant at 300 K by using Langevin Dynamics. The time step is 2 fs and trajectories are sampled at 40-ps intervals. Initial minimization is done for 10,000 steps, followed by 6-ns simulation runs.

The outputs of the MD simulations are analyzed between 2ns and 6 ns period. Although the simulations have long equilibration period, there might be instability at the beginning of the simulations. Hence, the first 2 ns part was taken out because of instabilities

at the beginning of the simulation and having no equilibration part in the simulation procedure In Figure 5.3. (A), the difference in the fluctuations of total energy of complex during simulation can be seen. The black line indicates the 2ns. The corresponding rmsd graph of the C α atoms of the 1CD0 can be seen as Figure 5.3. (B). It can be observed that there is an increase in rmsd up to the point close to the 2 ns line, which correspond also 50^{th} time frame. After this point, the rmsd of the complex fluctuate around imaginary line which is plotted by dashed line on the graph.

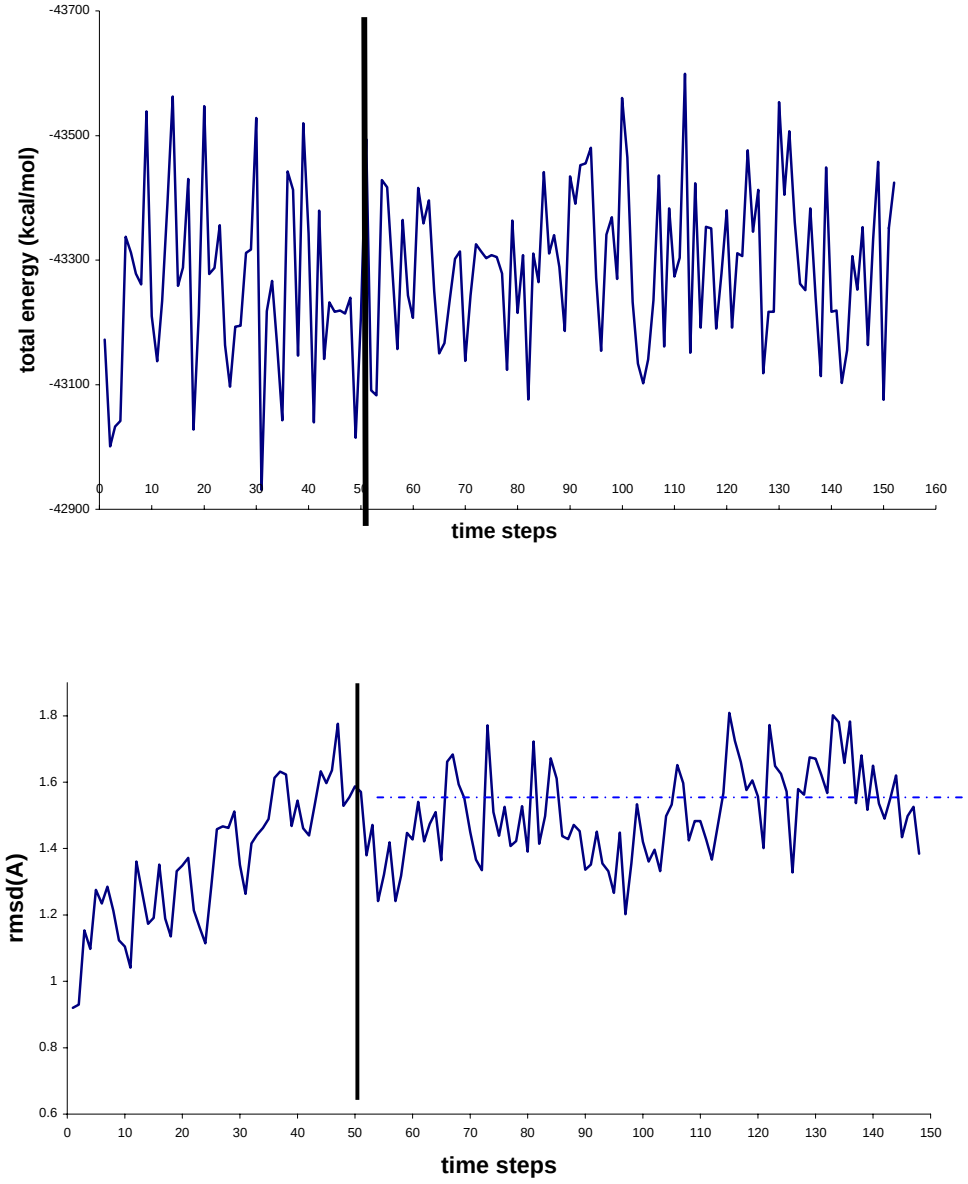


Figure 5.3 (A) total energy vs time frames of 1CD0 in complex simulation (B) rmsd of carbon α atoms of complex during simulation.

Trajectories are analyzed in order to observe interface residues' flexibilities and hydrogen bonding forming ability. Table 5.8 lists the overall averages of rmsd and H-bonding between 2ns and 6ns simulations for the complexes. These values are extracted from each snapshot from each MD trajectory. The root mean square deviation (rmsd) can measure the similarity in three-dimensional structure of protein complexes. The side chain flexibility is observed by means of rmsd calculation during the trajectory, see Appendix A-3. The side chain rmsd is calculated after the backbone of the interface residues is aligned in order to avoid systematic errors [35].

Figure 5.4 (B) shows the side chain dynamics of a hot spot residue (SER 42) in MD simulations of complex 1cd0 in both complex and monomeric simulation. The blue line corresponds to the simulations for the complex, and red one is for the monomeric simulation. As expected, the residue exhibits higher fluctuations during the trajectory of the monomeric chain. In the figure 5.4 –A there is side chain rmsd graph of a non hot spot residue (PRO 44) in 1cd0. There is a difference between monomeric and dimeric case but the difference is not the same during all simulation time. Almost 40 time frames which correspond to the 1.5 ns, non hot spot residue are in different conformations in both simulations. However, they are in different conformation in 80 time frames which is two times higher than the non hot spot case. In this study, although it is observed that hot spot residues in the interfaces are less mobile than the non hot spot residues, it is not always the case. The non hot spot residue (PRO 44) in Figure 5.4-B is appeared to be less mobile than the hot spot residue (SER 42). In all the interfaces, charged residues are more mobile than the rest of the interface residues.

Another observation is about the cystine residues in the interfaces. Cystine residues have capability to from disulphide bonds which is a covalent bond and it is stronger than the other interactions through the interfaces. For example, in 2btx Cys 33 forms a disulphide bond with another Cys which is not at the interface. It is predicted as hot spot residue by looking at the conservation score. It can be observed that it is 2 times less mobile than its neighbor residues. Different situation occurred in 1sbw although Cys 58 is not predicted as

hot spots; its side chain rmsd value is well under the average value because of the strength of disulphide bond.

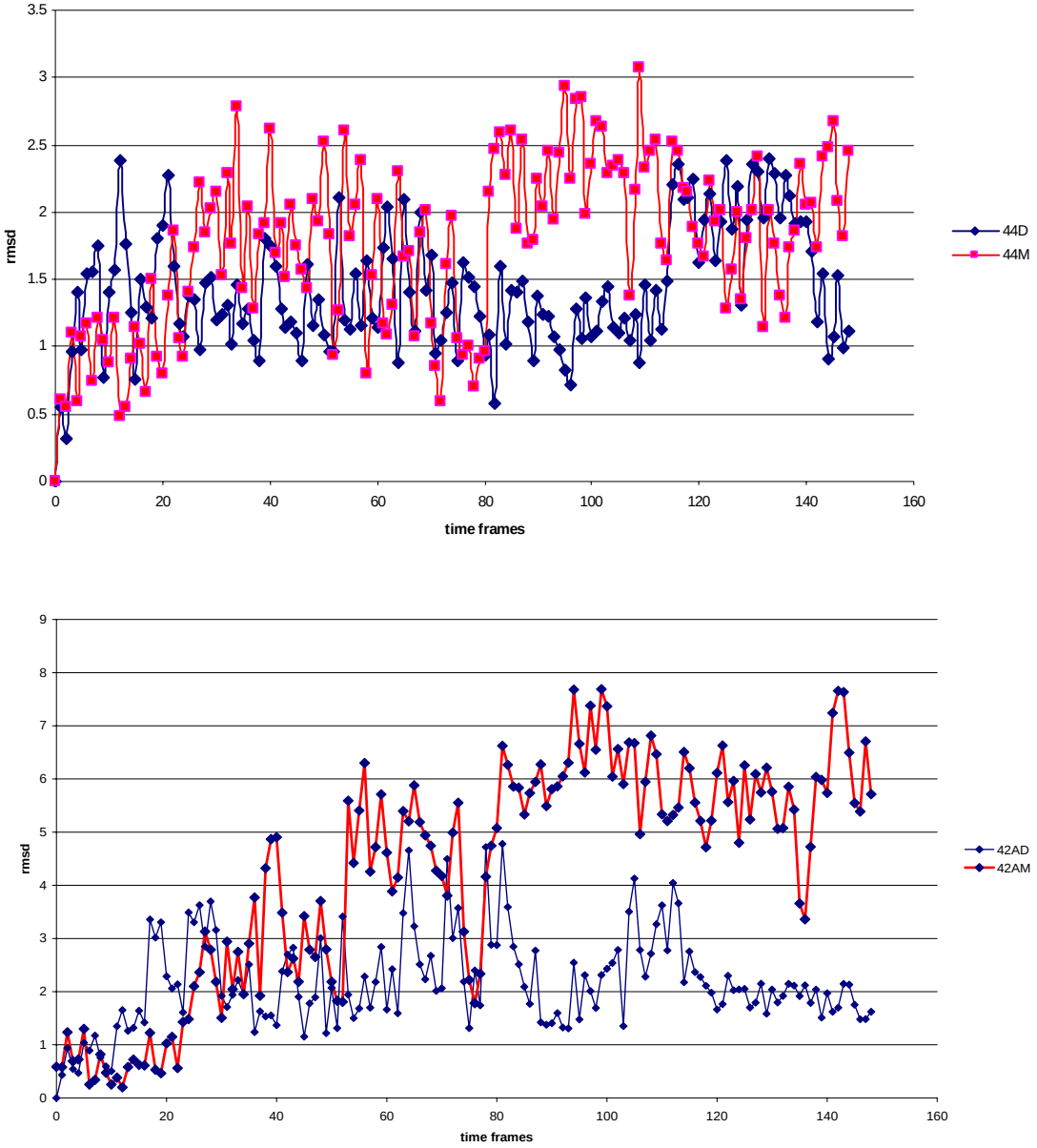


Figure 5.4 (a) side chain rmsd of non hot spot residues (PRO in 1CD0: A) (b) side chain rmsd of a hot spot residue (SER in 1CD0: A)

5.5.1 Side Chain Rmsd and Statistical Significance

In all the cases, the rmsd values of the hot spots are significantly smaller than the rest of the interface residues. These results are mostly extensively true for the homodimer complexes. The average over the complex types are as follows: for homodimers in Table 5.8.

We expect that the residues that we identified as hot spots will have less displacement from their initial coordinates compared to non-hot spots. To test this hypothesis, we calculated RMSD values of hot and non-hot spots for 4 different types of protein complexes: Homodimer, Antibody-Antigen, Enzyme-Inhibitor, and Protein-Peptide.

The RMSD values for hot and non-hot spots are not independent because both hot and non-spot residues of a protein complex are found in the same interface between proteins. The residues in this interface are likely to have a characteristic mobility that accounts for the motions of both hot and non-hot spots. If we think of an interface as a plane of interaction between proteins, hot and non-hot spots are scattered in this plane as a web of connected dots. This connectedness would lead to the dependence of average mobility values for hot and non-hot spots. Since RMSD is a measure of average mobility rates, we took these values as dependent in our statistical analysis.

For two columns of data that are dependent on each other, a paired t-test is in order. We implemented a separate paired t-test for each protein complex. The null and the alternative hypotheses in these tests were as follows:

Ho: RMSD values for hot-spots are equal to the RMSD values for non-hot spots
Ha: RMSD values for hot-spots are smaller than the RMSD values for non-hot spots

Using MINITAB, we obtained P-values for each test, which are shown in Table 5.6.

Table 5.6 statistical analysis of interface residues by using paired t test for hot spot data from CONSURF.

	RMSD Values							
	Homodimer		**Protein/Peptide**		**Antibody/Antigen**		**Enzyme/Inhibitor**	
	HS	**NHS**	**HS**	**NHS**	**HS**	**NHS**	**HS**	**NHS**
	1.27	1.78	1.05	1.73	0.91	0.96	0.97	1.38
	1.35	1.61	3.67	4.40	2.05	1.68	1.32	1.44
	1.29	1.83	1.92	2.33	1.03	2.05	1.63	2.12
	1.31	2.07	1.13	1.47				
	0.89	1.31						
	0.91	1.32						
	1.42	2.24						
	1.44	2.16						
P-value	0.000		0.006		0.314		0.047	
Conclusion	significant		significant		not significant		significant	

alpha level 0.05 for all tests

P-value for data pooled from all 4 groups = 0.000 (rmsd for hot spots significantly smaller)

At $\alpha = 0.05$ level, we were able to reject the null hypothesis for homodimer, protein-peptide and enzyme-inhibitor complexes (p-values = 0.000, 0.006 and 0.047 respectively). Therefore, we could conclude for these protein complexes that the rmsd values for hot-spots are significantly smaller than those of non-hot spots. For the antibody-antigen complex, we failed to reject the null hypothesis (p-value = 0.314) and concluded that there was not enough evidence to show that the rmsd values for hot-spots are significantly smaller than the values for non-hot spots. However it should be noted that the sample size for this protein complex can be considered too small (n = 6). Also, for one protein chain, the rmsd value for hot-spots

was greater than that for non-hot spots, which can be an outlier. If we had more data, we might have been able to show the significance of this test as well.

Finally, we pooled the data from all 4 protein complexes and implemented another paired t-test with the pooled data. Again, we were able to reject the null hypothesis (p-value=0.000) and conclude that the rmsd values for hot-spots are significantly smaller than the values for non-hot spots.

5.5.2 Comparison with the hot spots from Robetta Server

We repeated the same statistical tests with data obtained from Robetta, the computational alanine scanning mutagenesis software by the Baker Lab in the University of Washington. The results from these tests overlapped with our results only for homodimers and the pooled data in Table 5.7. For protein-peptide and enzyme-inhibitor complexes, data from Robetta were not able to show the significance of the difference between rmsd values of hot and non-hot spots (p-values= 0.059 and 0.355 respectively). For the antibody-antigen complex, the test with Robetta's data showed that the rmsd values for hot-spots were significantly smaller than those for non-hot spots (p-value = 0.037). This complex was the only one that our data did not show a significant difference, but it should be noted that the sample size of Robetta's data for this complex was greater (n = 10) and the data did not contain an obvious outlier as in the case with our data.

Table 5.7 statistical analysis of interface residues side chain rmsd by using paired t-test for hot spot data from Baker's computational scanning database information.

	RMSD Values							
	Homodimer		**Protein/Peptide**		**Antibody/Antigen**		**Enzyme/Inhibitor**	
	HS	**NHS**	**HS**	**NHS**	**HS**	**NHS**	**HS**	**NHS**
	1.07	1.87	0.93	1.64	1.35	1.80	1.07	1.18
	1.64	1.45	3.67	4.40	0.91	0.96	1.00	1.50
	1.64	1.52	1.97	2.30	1.28	1.83	2.09	1.78
	1.36	1.65	0.85	1.35	1.80	2.00		
	0.81	1.42			1.11	2.37		
	1.03	1.28						
	1.59	2.17						
	1.69	2.09						
P-value	0.032		0.059		0.037		0.355	
Conclusion	significant		not significant		significant		not significant	

alpha level 0.05 for all tests
P-value for data pooled from all 4 groups = 0.000 (rmsd for hot spots significantly smaller)

5.5.3 Hydrogen Bonding Ability

The hydrogen bonds that are formed by interface residues were monitored by using simple geometric criteria by using snapshots. This criteria are the cut off distance, which is 3Å, between donor and acceptor atoms and the bond angle (30^{0}) formed by donor, acceptor and hydrogen. The number of hydrogen bonds formed by each residue in the interface was calculated and averaged over all frames during simulations.

Hydrogen bond forming ability of hot spot residues does not differ from non – hot spot residues. This is consistent with the work done by Keskin et al. over a much larger dataset 21,686 two chains interfaces [38]. Among the residues in interfaces, charged residues

has greater contribution to the hydrogen bonding formation than the other residues because charged residues contribute through their both side chains and backbone.

The paired t-test is applied to the hydrogen bonding data in Table 5.8. The α-level is selected as 0.05 and for whole type of interfaces p –values are higher than α-level. That is, the difference between hot spot and non spot residues are not significant. However, when the test is applied on the pooled data, it is observed that numbers of hydrogen bonds formed by non hot spot residues are significantly less than non hot spot residues. It can be explained by O-ring theory, states that there are energetically less significant residues around hot spots prevent hot spots from bulk solvent by interacting with solvent molecules.

Table 5.8 statistical anaylysis of interface residues hydrogen bonding ability by using paired t-test for hot spot data.

	Number of Hydrogen Bonds							
	Homodimer		**Protein/Peptide**		**Antibody/Antigen**		**Enzyme/Inhibitor**	
	HS	**NHS**	**HS**	**NHS**	**HS**	**NHS**	**HS**	**NHS**
	1.13	1.59	1.48	1.57	0.96	0.96	0.86	1.41
	1.05	1.43	0.81	1.20	1.07	1.29	1.26	1.28
	1.22	1.52	1.03	1.45	0.94	1.22	1.51	2.15
	1.14	1.50					0.95	0.63
	1.82	1.52						
	1.75	1.56						
	0.86	0.99						
	0.99	0.94						
P-value	0.110		0.199		0.052		0.095	
Conclusion	not significant		not significant		not significant (but borderline)		not significant	
alpha level	0.05 for all tests							

P-value for data pooled from all 4 groups = 0.006 (number of hydrogen bonds for hot spots significantly less)

Table 5.9 overall averages of rmsd and hydrogen bond values of hot spot and non hot spot residues for all complexes. HSHB: hot spot hydrogen bond, NHSHB: non-hot spot hydrogen Bond, HS-rmsd: hot spot rmsd, NHS-rmsd: non-hot spot rmsd

PDB ID	TYPE	HSHB	NHSHB	HS-rmsd	NHS-rmsd
	homodimer				
1CD0	**A**	1.13	1.59	1.27	1.78
	B	1.05	1.43	1.35	1.61
1BZD	**A**	1.22	1.52	1.29	1.83
	B	1.14	1.50	1.31	2.07
1CQK	**A**	1.82	1.52	0.89	1.31
	B	1.75	1.56	0.91	1.32
1MR8	**A**	0.86	0.99	1.42	2.24
	B	0.99	0.94	1.44	2.16
AVERAGE		**1.16**	**1.08**	**1.24**	**1.79**
	Antibody/antigen				
1VFB	**C**	1.48	1.57	0.91	0.96
1QNZ	**H**	0.81	1.20	2.05	1.68
1FCC	**A**	1.03	1.45	1.03	2.05
AVERAGE		**1.00**	**1.03**	**1.33**	**1.56**
	Enzyme/inhibitor				
2SNI	**E**	0.96	0.96	0.97	1.38
1SBW	**A**	1.07	1.29	1.32	1.44
1UGH	**E**	0.94	1.22	1.63	2.12
AVERAGE		**0.97**	**1.18**	**1.31**	**1.65**
	Protein/peptide				
1CJQ	**B**	0.86	1.41	1.05	1.73
2BTX	**A**	1.26	1.28	3.67	4.40
1DDH	**A**	1.51	2.15	1.92	2.33
1BJR	**A**	0.95	0.63	1.13	1.47
AVERAGE		**1.25**	**1.25**	**1.94**	**2.48**

Chapter 6

Conclusions

Proteins generally function through reciprocal cooperation; they bind together into protein complexes and help each other to perform their functions. Comprehending the protein – protein interaction extensively will provide great improvement in the conformational drug researches, prediction of protein-protein interaction and constitute protein interaction networks. Throughout the protein interfaces there are specific regions or amino acids called “hot spots” which play the most crucial role in protein association. The characterization of the interface and focusing on these essential residues has strong implication in design and development of drug compounds. Current drugs target the whole binding side of the protein complexes which causing undesired side effects. Focusing on important regions helps drug designers to produce small molecules that interfere or alter the specific function of proteins. Another and more critical implication of researches on protein – protein interaction is to understand mechanism of association.

There are several approaches to find the specific regions which regulate the association process. In this study, we have combined the sequence conservation information with the dynamics of the protein complexes by performing molecular dynamics simulations. As a starting point 14 different protein complexes which constituted 4 different types of protein complexes were extracted from PDB database according to their size and interface information. By performing ASA calculation the interface residues of the complexes are determined. To do this, the ASA of each residue calculated in both monomeric and dimeric form. If there is a difference between them it concluded that the residue is at the interface.

The number of residues in the interfaces, i.e. interfaces size, varies between 25 and 73 and the total accessible surface are of the interfaces between 800 Å^2 and 2597 Å ^{2} .Our computational hot spots were determined by the conservation scores calculated by software CONSURF because it is known that functionally important regions in proteins are conserved during evolution. Among total number of 617 residues, 153 residues predicted as computational hot spots.

The characteristics of the interfaces are analyzed with those interface residues. Interfaces are dominated by hydrophobic residues in almost all types of complexes. Then the second contribution is from residues that have aromatic side chains. Tyr, Ser, Phe, Gly and Thr existed more frequently than the other residues in the interfaces. Hot spots are more buried than the rest of the interface by observing the difference in their ASA values between hot spots and non spots residues. Paired t- test was applied to understand significance of the difference and the p-value (0.022) was found well under the α – value (0.05). It means the difference is statistically important.

It is expected that the hot spot residues should be less flexible than the rest of the interface. It is observed with a calculation from Molecular Dynamics simulation and proved by applying paired t-tests. At α = 0.05 level, we are able to reject the null hypothesis for homodimer, protein-peptide and enzyme-inhibitor complexes (p-values = 0.000, 0.006 and 0.047 respectively). Therefore, we could conclude for these protein complexes that the rmsd values for hot-spots are significantly smaller than those of non-hot spots. For the antibody-antigen complex, we failed to reject the null hypothesis (p-value = 0.314) and concluded that there is not enough evidence to show that the rmsd values for hot-spots are significantly smaller than the values for non-hot spots. However, in some interfaces the residues which are non hot spots are found to be less flexible because it depends on the location of the interface and type of the residue. Charged residues across the interfaces found to be more mobile because of they are relatively long side chains. Hydrogen bond forming ability of hot spot residues does not differ from non – hot spot residues. As a conclusion, across the interfaces

the interactions dominated by important residues are composed of mainly by hydrophobic interactions.

Recommendations for Future Work:

The free energy calculation is another identification method of hot spots across interfaces. The free energy calculations can be explored as an extension for future work and the results from free energy calculations can be discussed together with outcomes of our study.

APPENDIX

A.1. Prediction of Hotspots by Consurf

Given the 3D-structure of a protein or a domain as an input, the server extracts the sequence from the Protein Data Bank (PDB) [74] which has the three-dimensional structures information of biological macromolecules. Then it automatically carries out a search for close homologous sequences of the protein of known structure. The server uses the PSI-BLAST [75]heuristic algorithm with default parameters to collect homologous sequences of a single polypeptide chain of known 3D-structure. The search is carried out using the SWISS-PROT [76] database or the full UNI-PROT [77] knowledgebase (SWISS-PROT + TrEMBL [76]) databases and a default single iteration of PSI-BLAST with a E-value (expectation value) cutoff of 0.001. It then multiply aligns them by using CLUSTAL W [78] with default parameters and homologues extracted from the PSI-BLAST output file.

Finally, it builds a phylogenetic tree consistent with the multiple sequence alignment (MSA), and calculates the conservation scores using either the Maximum Likelihood [79] or the Maximum Parsimony [77] method. In the maximum Parsimony method, given the MSA of homologous proteins the method tries to find the tree which has minimum number of amino acid substitution. Together with the physicochemical similarity between amino acid, the conservation scores are calculated. The rate of evolution is not constant among amino acids, there are some regions on the protein surface which is relatively more conserved than the other regions. In this study, we have used Maximum Likelihood Method, because it allows taking into account the stochastic process underlying sequence evolution within protein families and the phylogenetic tree of the proteins in the family. The conservation

scores represent the evolutionary rate of aminoacids. The whole process is summarized as a flowchart in Figure A1.1.

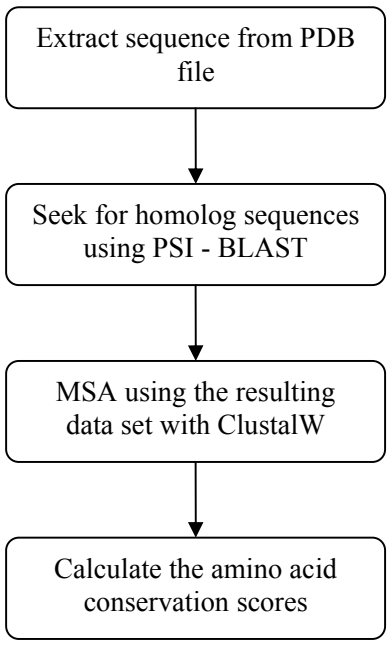


Figure A1.1 Flowchart of the methodology of CONSURF software

A1.2 Illustrative example

In this part, the detection of Hot Spot residues by using CONSURF will be elucidated with an example output file as in Figure 5.2. PDB code: 1mr8, which is a homodimer, has been selected as model complex. The procedure is straightforward and easy to apply:

- CONSURF program takes pdb code of the desired protein and its chain identifier as input and generates an output file which has amino acid conservation scores
- By applying accessible surface area calculations (ASA) the interface residues of the complex should be determined
- Finally by using a simple script, the amino acid which has high conservation scores and is located in the interface at the same time will be named as Hot Spot residues.

POS	SEQ	3LATOM	SCORE	COLOR	MSA DATA	RESIDUE VARIETY
1	M	MET1:A	-1.447	9	14/50	M
2	L	LEU2:A	0.557	3	22/50	A,C,E,L,M,P,S,T,V
3	T	THR3:A	-0.357	6	45/50	C,F,P,R,S,T
4	E	GLU4:A	-0.002	5	45/50	D,E,K,P,Q,S
5	L	LEU5:A	-0.788	7	46/50	L,M,T
6	E	GLU6:A	-0.837	8	48/50	D,E,K,L,Q
7	K	LYS7:A	0.862	2	48/50	D,E,H,K,M,Q,R,S,T,X
8	A	ALA8:A	-0.724	7	48/50	A,C,G,H,N,S
9	L	LEU9:A	-0.229	6	48/50	I,L,M,P,V
10	N	ASN10:A	1.466	1	48/50	A,D,E,G,I,N,S,T,V
11	S	SER11:A	0.213	4	48/50	A,D,E,G,L,M,N,S,T,V
12	I	ILE12:A	-0.129	5	49/50	I,L,M,V
13	I	ILE13:A	-0.909	8	50/50	I,K,L,M,V
14	D	ASP14:A	0.602	3	50/50	A,D,E,F,G,K,N,R,S,T
15	V	VAL15:A	-0.307	6	50/50	I,L,T,V
16	Y	TYR16:A	-1.293	9	50/50	F,Y
17	H	HIS17:A	-1.038	8	50/50	E,H,Q,S,Y
18	K	LYS18:A	-0.382	6	50/50	A,K,N,Q,R,T
19	Y	TYR19:A	-1.130	9	50/50	H,Y
20	S	SER20:A	-1.072	8	50/50	A,S
21	L	LEU21:A	-0.101	5	50/50	A,C,G,L,N,R,S,T,V
22	I	ILE22:A	0.882	2	48/50	E,I,K,N,Q,R,S,X
23	K	LYS23:A	0.897	2	49/50	A,D,E,H,K,L,Q,T,V,Y
24	G	GLY24:A	-0.963	8	50/50	G,K,R,S
25	N	ASN25:A	0.199	4	50/50	D,E,H,K,N,P,S,Y
26	F	PHE26:A	1.587	1	48/50	C,F,H,K,N,P,R,S,T,Y
27	H	HIS27:A	0.669	3	49/50	C,D,F,H,L,N,Q,R,T,V,Y
28	A	ALA28:A	-0.128	5	50/50	A,H,K,M,N,Q,S,T,Y
29	V	VAL29:A	-0.997	8	50/50	I,L,V
30	Y	TYR30:A	-0.195	6	50/50	E,I,K,N,S,T,Y
31	R	ARG31:A	-0.769	7	50/50	K,Q,R,X

Figure A1.2 An example of the consurf output file. POS is the position of the aminoaicds in the SEQRES derived sequence, SEQ is the SEQRES derived sequence in one letter code, SCORE is the normalized conservation scores, 3LATOM is the ATOM derived sequence in three letter code, including the amino acid's positions as they appear in the PDB file and the chain identifier, COLOR is the color scale representing the conservation scores (9 - conserved, 1 - variable), MSA DATA is the number of aligned sequences having an amino acid (non-gapped) from the overall number of sequences at each position and RESIDUE VARIETY is the residues variety at each position of the multiple sequence alignment.

A.2. Solvent Accessible Surface Area (SASA) Calculation by Naccess

The Naccess software calculates the accessible area of a molecule from a PDB (Protein Data Bank) format file. It can calculate the atomic and residue accessibilities for both proteins and nucleic acids, and is available for free from this site for researchers at academic and non profit-making institutions. It can be downloaded from the NACCESS (http://wolf.bms.umist.ac.uk/naccess/) website.

The program uses the Lee & Richards [80] method, whereby a probe of given radius is rolled around the surface of the molecule, and the path traced out by its centre is the accessible surface. In general, the probe has the same radius as water (1.4 Angstroms) and hence the surface described is often referred to as the solvent accessible surface. The calculation makes successive thin slices through the 3D molecular volume to calculate the accessible surface of individual atoms. The connection of the solvent sphere with a given z-slice appears as arcs. The exposed regions are sum of these arc lengths over all z-slices. The overlapping arcs representing the atoms of the same molecule are eliminated.

The accessible surface area (ASA) can be calculated for each individual residue. The solvent accessibility of each individual residue can be quantified with a biologically more meaningful measure, called its relative accessibility. Relative accessibility of a residue is defined as the percent accessibility compared to the accessibility of that residue in an extended ALA-X-ALA tripeptide.

The ASA calculation is performed both monomers and protein complex individually. In order to identify a residue as a interface residue, there is difference between ASA of a amino acid between monomer and complex form. The interface residues are detected by means of this definition.

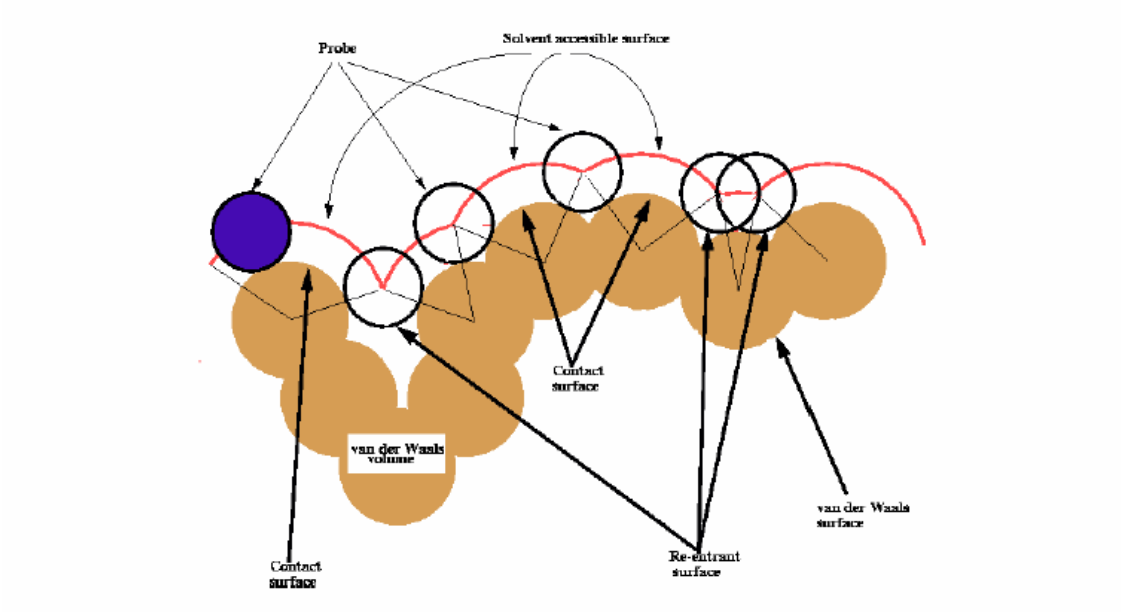


Figure A.2.1 Envelope of solvent accessible surface per slice.[81]

Naccess program produces three different output files of the, .asa .rsa and .log files. Here are three examples of the output files. The output files are in PDB format. The first line after the coordinates represents the accessible surface area of the corresponding atom and the following line the assigned vdw radius of the atom.[75]

```
ATOM      1  N   ASN A   1      12.550  10.609  21.364  31.155  1.65
ATOM      2  CA  ASN A   1      13.155  10.863  20.025   2.070  1.87
ATOM      3  C   ASN A   1      14.251   9.846  19.722   0.218  1.76
ATOM      4  O   ASN A   1      15.341   9.908  20.295  11.493  1.40
ATOM      5  CB  ASN A   1      13.751  12.269  19.980  33.195  1.87
ATOM      6  CG  ASN A   1      14.314  12.624  18.613   9.033  1.76
ATOM      7  OD1 ASN A   1      14.889  13.697  18.429  41.205  1.40
ATOM      8  ND2 ASN A   1      14.143  11.726  17.645  25.938  1.65
ATOM      9  N   PHE A   2      13.972   8.908  18.825   0.000  1.65
ATOM     10  CA  PHE A   2      14.980   7.918  18.481   1.857  1.87
ATOM     11  C   PHE A   2      14.807   7.377  17.074   0.000  1.76
ATOM     12  O   PHE A   2      13.771   7.570  16.444   3.535  1.40
ATOM     13  CB  PHE A   2      14.973   6.776  19.508   0.547  1.87
ATOM     14  CG  PHE A   2      13.737   5.923  19.477   0.000  1.76
```

The beginning and end of an example .rsa file are shown below. There are five different types of data group: Total (all atoms), Non Polar Side chain (all non-oxygens and

non-nitrogens in the side chain), Polar Side chain (all oxygens and nitrogens in the side chain), total side chain, and main chain.

```
REM  Relative accessibilites read from external file "standard.data"
REM  File of summed (Sum) and % (per.) accessibilities for
REM RES _ NUM      All-atoms   Total-Side   Main-Chain    Non-polar    All polar
REM                ABS   REL    ABS   REL    ABS   REL    ABS   REL    ABS   REL
RES ASN A   1   154.31 107.2 111.44 104.9  42.87 113.7  44.52  96.3 109.79 112.4
RES PHE A   2     8.00   4.0   4.47   2.7   3.54  10.0   4.47   2.7   3.54  10.3
RES MET A   3   114.42  58.9 114.42  73.0   0.00   0.0 114.42  72.5   0.00   0.0
RES LEU A   4     3.62   2.0   0.00   0.0   3.62   9.6   0.00   0.0   3.62  10.0
RES ASN A   5    85.07  59.1  81.85  77.0   3.22   8.5  31.46  68.1  53.61  54.9
RES GLN A   6    24.22  13.6   0.89   0.6  23.33  62.2   4.83   9.2  19.39  15.4
RES PRO A   7    46.31  34.0  45.85  38.2   0.45   2.8  45.85  37.9   0.45   3.0
RES HIS A   8   152.25  83.3 142.24  96.7  10.01  27.9 104.62 107.7  47.63  55.6
RES SER A   9    68.41  58.7  47.16  60.4  21.25  55.3  38.42  79.1  29.99  44.1
 .   .      .      .      .     .      .     .      .     .      .     .      .
 .   .      .      .      .     .      .     .      .     .      .     .      .
 .   .      .      .      .     .      .     .      .     .      .     .      .
RES LEU B 104     2.77   1.5   0.00   0.0   2.77   7.4   0.55   0.4   2.22   6.1
RES THR B 105    62.47  44.9  62.47  61.4   0.00   0.0  35.22  46.5  27.24  42.9
RES VAL B 106    13.09   8.6   0.96   0.8  12.13  32.6   0.96   0.8  12.13  33.7
RES LEU B 107    97.86  54.8  46.30  32.8  51.55 137.4  59.82  42.0  38.04 104.7
END  Absolute sums over single chains surface
CHAIN  1 A     5078.4       4191.2        887.2       2723.4       2354.9
CHAIN  2 B     5148.3       4234.3        914.0       2671.4       2477.0
END  Absolute sums over all chains
TOTAL         10226.7       8425.5       1801.2       5394.8       4831.9
```

Here is an example of .log file which includes information about the Naccess parameters and the operation progress:

```
ACCALL - Accessibility calculations
 MAX RESIDUES     5000
 MAX ATOMS/RES     100
 PDB FILE INPUT 1CD0.pdb
 PROBE SIZE       1.40
 Z-SLICE WIDTH   0.050
 VDW RADII FILE vdw.radii
 EXCL HETATOMS
 EXCL HYDROGENS
 EXCL WATERS
 READVDW  32 residues input
 ADDED VDW RADII
 CHAINS       2
 RESIDUES   221
 ATOMS     1670
 SOLVA: PROGRAM ENDS CORRECTLY
 CALCULATED ATOMIC ACCESSIBILITES
 RELATIVE (STANDARD) ACCESSIBILITIES READFOR  20 AMINO ACIDS
 SUMMED ACCESSIBILITIES OVER RESIDUES
```

A-3 Side chain rmsd calculations

The analysis of the side chain flexibility was done by using root mean square analysis on the molecular dynamic simulation output. The trajectories were taken out and analyzed by using VMD and appropriate scripts written by Tcl. First of all, the pre-defined interface residues' backbones were superimposed to the initial step of the simulation and then rmsd calculations were carried out for each residues side chains in the interface.
By looking at the side chain rmsd, we observe flexibility of the side chain residues.

The procedure can be illustrated in a flow chart as follows:

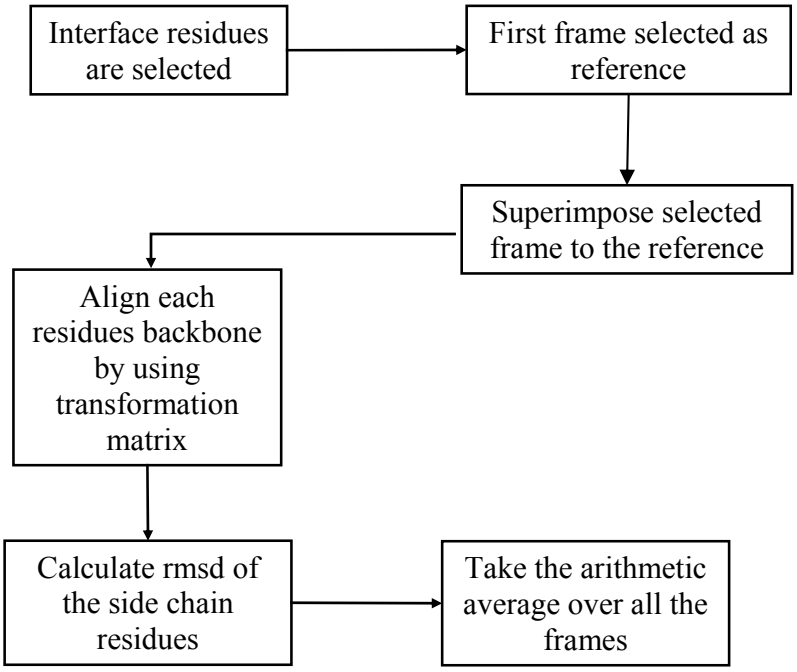


Figure A.3.1 flowchart of the side chain rmsd calculation

Formally, given N atom positions from structure x and the corresponding N atoms from structure y with a weighting factor w (i), the RMSD is defined as in Equation [1]:

$$RMSD\ (N; x, y) = \frac{\left[\sum_{i=1}^{N} wi \left\| x_i - y_i \right\|^2 \right]^{1/2}}{\left[N \sum_{i=1}^{N} w_i \right]^{1/2}} \qquad \text{(A.1)}$$

A-4 List of ASA, Side Chain Rmsd and Hydrogen Bond of All Complexes

1vfb Chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 30		54.94	12.71	42.23	1.10	1.96
GLY 31		62.83	31.89	30.94	0.80	-
TYR 32		5.96	3.77	2.19	1.77	1.71
GLY 33		81.34	27.56	53.78	0.75	-
TRP 52		16.93	6.70	10.23	1.95	0.78
GLY 53		113.22	77.17	36.05	0.53	-
ASP 54		116.50	67.52	48.98	0.27	1.66
ARG 99		95.62	34.00	61.62	4.06	1.88
ASP 100		169.44	76.55	92.89	0.21	1.89
THR 101		122.00	94.03	27.97	1.68	1.40
ARG 102		85.17	77.12	8.05	3.15	1.35

1vfb ChainC						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
GLY 22		61.09	23.37	37.72	0.44	-
TYR 23		19.66	10.39	9.27	1.63	0.97
SER 24	√	55.46	29.02	26.44	2.33	0.97
ASN 27		15.95	5.48	10.47	2.60	0.57
LYS 116		116.99	55.50	61.49	3.38	2.50
GLY 117		83.89	15.00	68.89	0.85	-
THR 118		80.49	36.34	44.15	1.45	0.53
ASP 119	√	82.91	38.07	44.84	0.63	0.86
VAL 120		18.74	4.99	13.75	0.20	0.68
GLN 121		95.95	63.98	31.97	2.03	0.56

1cjq Chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ALA 4		73.83	22.62	51.21	0.67	1.35
ALA 5		112.88	5.85	107.03	0.72	1.06
PHE 8		104.63	53.65	50.98	0.91	1.07
GLU 9		157.54	70.96	86.58	0.85	2.47
ARG 10		121.73	33.30	88.43	5.51	1.71
GLN 11		118.83	16.52	102.31	1.89	1.26
HIS 12		97.68	12.65	85.03	1.83	0.99
MET 13		130.08	35.23	94.85	0.80	2.67
ASP 14		156.65	100.62	56.03	0.75	3.09
SER 15		90.71	53.74	36.97	1.23	4.33

1cjq Chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
TYR 25		81.28	73.84	7.44	1.57	1.64
MET 29		43.56	31.91	11.65	0.99	1.79
ARG 33		143.55	54.18	89.37	4.35	1.46
ASN 34		82.21	61.73	20.48	1.12	3.20
LEU 35	√	35.50	1.88	33.62	0.85	1.62
ARG 39		114.86	108.14	6.72	4.70	6.08
LYS 41		69.97	54.09	15.88	2.40	1.50
ASN 44	√	23.30	4.13	19.17	1.59	0.60
THR 45	√	29.75	13.12	16.63	1.42	0.63
PHE 46		19.32	0.00	19.32	0.89	0.83
VAL 47		45.45	0.00	45.45	0.92	0.95
HIS 48	√	82.98	50.60	32.38	1.97	0.82
GLU 49	√	68.29	50.50	17.79	0.08	1.36
SER 50		58.63	48.02	10.61	1.78	1.50
LEU 51		123.60	56.63	66.97	0.80	1.48
VAL 54	√	17.64	0.00	17.64	0.56	1.28
GLN 55		61.78	51.49	10.29	1.81	2.72
ILE 106		0.11	0.00	0.11	0.50	0.67
VAL 108		5.79	0.00	5.79	0.85	0.59
GLU 111		93.05	80.38	12.67	0.60	2.11
VAL 116		44.29	25.10	19.19	0.80	1.02
PRO 117		36.45	2.71	33.74	0.04	1.10

VAL 118		61.59	13.30	48.29	0.86	1.58
HIS 119		124.44	119.37	5.07	1.28	1.51
PHE 120		73.86	36.98	36.88	0.64	1.14

2sni Chain E						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
SER 33	√	0.78	0.64	0.14	1.92	0.71
ASN 62		48.36	37.54	10.82	1.71	0.85
SER 63		20.79	17.27	3.52	1.32	0.90
HIS 64	√	40.94	0.00	40.94	2.46	0.72
LEU 96	√	14.09	0.00	14.09	0.67	1.36
ASP 99		119.78	67.47	52.31	0.17	1.41
GLY 100	√	35.84	0.67	35.17	0.41	-
SER 101	√	62.21	6.81	55.40	1.60	1.68
GLY 102	√	25.83	1.46	24.37	0.48	-
GLN 103		114.66	71.04	43.62	1.22	2.89
TYR 104		80.72	32.71	48.01	1.14	1.75
ILE 107	√	14.65	0.00	14.65	0.15	1.27
SER 125	√	3.35	0.00	3.35	1.20	0.78
LEU 126		27.10	0.20	26.90	0.05	1.35
GLY 127	√	61.56	7.81	53.75	0.88	-
GLY 128		15.75	5.55	10.20	0.58	-
PRO 129		132.04	96.64	35.40	0.10	1.21
LEU 135		5.53	0.00	5.53	0.51	1.01
ALA 152	√	8.91	3.06	5.85	0.65	0.89
GLY 154	√	19.01	4.03	14.98	0.19	-
ASN 155	√	78.58	20.67	57.91	1.87	0.84
GLU 156		124.18	97.79	26.39	0.78	1.43
TYR 167		43.41	41.40	2.01	1.53	1.12
PRO 168	√	2.19	0.00	2.19	0.04	1.52
PHE 189	√	57.48	19.85	37.63	0.11	0.68
TYR 217		79.65	37.31	42.34	1.70	1.33
ASN 218	√	91.28	54.30	36.98	1.55	1.00
GLY 219	√	2.58	0.00	2.58	0.87	-
THR 220	√	1.48	0.00	1.48	1.35	0.46
SER 221	√	15.86	0.00	15.86	1.61	1.10
MET222	√	3.21	0.40	2.81	0.23	0.70

2sni Chain I						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ILE 49		67.56	48.83	18.73	0.99	1.07
VAL 50		60.97	58.83	2.14	0.84	1.19
LEU 51		38.22	2.11	36.11	1.09	1.34
PRO 52		81.84	57.82	24.02	0.16	1.16
VAL 53		57.14	54.70	2.44	0.40	1.51
GLY 54		74.88	44.18	30.70	0.89	-
THR 55		63.79	5.58	58.21	1.39	2.01
ILE 56		174.83	0.09	174.74	0.81	1.70
VAL 57		49.45	14.16	35.29	1.15	1.04
THR 58		87.54	1.03	86.51	1.95	1.69
MET 59		178.36	15.26	163.10	0.44	1.21
GLU 60		68.24	3.59	64.65	0.04	1.24
TYR 61		164.48	58.24	106.24	1.79	1.20
ARG 62		131.35	85.08	46.27	4.51	2.23
ARG 67		33.98	27.32	6.66	4.25	2.61
GLU 78		71.32	46.75	24.57	0.20	1.91
ARG 81		125.80	105.44	20.36	5.43	2.08

1sbw chainA						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
TYR 39	√	128.82	105.19	23.63	1.75	0.83
HIS 40	√	34.17	18.83	15.34	1.10	1.69
PHE 41	√	33.28	3.75	29.53	0.91	0.51
CYS 42		7.74	0.01	7.73	0.31	0.75
HIS 57		69.33	10.08	59.25	1.75	0.65
CYS 58		4.72	3.94	0.78	0.35	2.45
LYS 60		60.80	59.66	1.14	2.17	0.78
TYR 94		52.46	48.89	3.57	1.85	1.81
SER 96		79.29	57.44	21.85	1.38	1.74
ASN 97		135.82	96.45	39.37	1.43	1.32
THR 98		69.67	52.86	16.81	1.41	1.01
LEU 99		39.58	2.92	36.66	0.31	0.70
TRP 141		2.46	2.15	0.31	1.87	1.76

TYR 151		133.91	109.19	24.72	1.77	0.99
TYR 172	√	6.75	4.55	2.20	1.84	1.79
GLN 175	√	94.60	28.06	66.54	1.61	1.50
ASP 189	√	6.37	4.32	2.05	0.86	0.78
SER 190		10.95	0.08	10.87	1.35	1.06
CYS 191		10.94	2.22	8.72	0.97	4.10
GLN 192		117.1	26.43	90.67	1.05	1.03
GLY 193	√	17.31	0.32	16.99	0.62	-
SER195		17.66	0.00	17.66	1.64	0.51
VAL 213		8.20	0.13	8.07	0.95	0.56
SER 214		10.12	2.34	7.78	1.13	0.53
TRP 215	√	49.49	0.56	48.93	1.31	2.16
GLY 216	√	30.74	1.58	29.16	0.54	-
SER 217		51.60	20.32	31.28	1.71	1.72
GLY 219	√	50.84	37.17	13.67	0.20	-
CYS 220		7.43	5.86	1.57	0.78	4.76
LYS 224		68.51	57.43	11.08	2.79	0.59
GLY 226		3.11	0.11	3.00	0.64	-
VAL 227		0.92	0.76	0.16	0.84	1.60

1sbw chainI						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
SER 15		117.43	75.45	41.98	2.84	2.21
CYS 16		98.12	77.60	20.52	0.90	1.54
ARG 17		163.45	24.93	138.52	5.07	1.99
CYS 18		63.27	21.16	42.11	0.69	1.09
THR 19		59.33	0.00	59.33	1.27	0.54
LYS 20		207.89	0.96	206.93	2.42	1.65
SER 21		49.23	4.50	44.73	0.84	0.67
ILE 22		176.86	39.52	137.34	0.76	0.85
PRO 23		104.24	97.60	6.64	0.06	1.46
PRO 24		77.74	59.73	18.01	0.09	1.22
GLN 25		109.04	68.06	40.98	1.80	2.31
HIS 27		104.36	54.19	50.17	1.31	1.69

1cd0 chainA						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
GLN 34		12.14	4.93	7.21	1.46	0.86
TYR 36		42.15	4.47	37.68	1.65	0.54
GLN 38	√	52.31	17.37	34.94	1.43	1.27
GLY 41		95.96	93.78	2.18	0.71	-
SER 42	√	55.20	48.49	6.71	1.10	2.23
ALA 43		74.84	23.65	51.19	0.36	2.48
PRO 44		74.40	7.62	66.78	1.06	1.30
ILE 45		90.75	87.49	3.26	0.97	1.21
THR 46		47.31	5.53	41.78	1.22	0.83
TYR 49	√	79.72	15.09	64.63	1.82	1.22
GLN 53		79.81	65.23	14.58	1.27	2.18
ARG 54		106.71	89.21	17.50	3.65	2.09
PRO 55	√	40.94	7.88	33.06	0.13	1.24
SER 56		118.15	109.69	8.46	1.51	3.93
TYR 87	√	58.38	9.97	48.41	1.79	1.02
GLN 89		2.16	1.91	0.25	1.36	0.61
TYR 91	√	77.77	47.21	30.56	0.86	0.64
ARG 94		91.21	43.24	47.97	4.74	5.93
ASN 95		100.19	32.88	67.31	1.51	1.44
VAL 96		65.65	7.07	58.58	0.25	0.66
PHE 98		111.19	22.36	88.83	0.96	0.95
GLY 99		5.64	0.39	5.25	1.43	-
GLY 100		68.47	55.23	13.24	1.28	-

1cd0 chainB						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
HIS 8		144.95	143.24	1.71	1.61	1.25
GLN 34		10.81	3.71	7.10	1.57	1.82
TYR 36		42.21	3.72	38.49	1.90	0.61
GLN 38	√	46.56	8.75	37.81	0.97	1.37
GLY 41		88.00	86.28	1.72	0.66	-
SER 42	√	67.62	62.00	5.62	1.30	2.54
ALA 43		77.31	21.36	55.95	0.24	1.30
PRO 44		73.77	9.47	64.30	1.08	0.98
ILE 45		86.25	85.55	0.70	1.00	1.25
THR 46		47.02	6.86	40.16	1.41	0.62
TYR 49	√	80.95	34.48	46.47	1.79	1.32
GLU 50		71.57	51.71	19.86	0.86	2.03
PRO 55	√	34.86	12.59	22.27	0.18	1.13
TYR 87	√	58.10	12.62	45.48	1.53	1.02
GLN 89		2.62	1.47	1.15	1.34	0.92
TYR 91	√	91.43	57.04	34.39	0.76	1.28
ALA 93		90.72	86.96	3.76	0.65	1.80
ARG 94		219.19	119.36	99.83	4.47	4.52
ASN 95		97.89	46.72	51.17	1.85	2.34
VAL 96		65.88	6.76	59.12	0.48	1.05
PHE 98		114.11	22.08	92.03	0.73	1.01
GLY 99		4.64	0.19	4.45	1.59	-
GLY 100		67.24	52.73	14.51	1.66	-

1mr8 chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 3		74.96	21.09	53.87	2.00	1.94
GLU 4		132.67	89.27	43.40	0.56	2.06
LEU 5		139.14	0.28	138.86	0.83	1.58
GLU 6		84.47	38.99	45.48	0.22	2.13
ALA 8		52.37	6.66	45.71	0.69	0.97
LEU 9		96.64	7.13	89.51	0.75	1.77
ASN 10		70.04	60.63	9.41	1.70	1.60
SER 11		61.78	21.91	39.87	1.68	1.07
ILE 12		50.54	2.07	48.47	0.58	2.12
ILE 13	√	63.57	6.64	56.93	0.43	1.04
VAL 15		38.90	6.57	32.33	0.99	1.32
TYR 16	√	7.74	3.11	4.63	1.27	1.25
HIS 17	√	85.72	35.51	50.21	1.72	1.32
HIS 27		142.60	112.92	29.68	1.52	5.69
LEU 37	√	3.27	0.47	2.80	0.75	1.32
THR 40		86.80	75.08	11.72	1.15	3.39
GLU 41	√	65.98	13.15	52.83	0.70	1.98
CYS 42		14.48	4.94	9.54	1.25	3.07
PRO 43		43.49	11.59	31.90	0.14	3.71
TYR 45		52.01	50.55	1.46	1.12	3.26
PHE 68	√	63.08	8.93	54.15	0.43	1.17
GLN 69		126.05	96.79	29.26	1.54	3.16
PHE 71		9.61	1.41	8.20	0.89	1.19
LEU 72		52.96	11.50	41.46	0.91	1.59
LEU 74		1.40	0.06	1.34	0.73	2.06
VAL 75		36.74	13.75	22.99	1.00	0.84
ILE 76		110.24	49.34	60.90	0.92	2.52
MET 78		49.32	4.07	45.25	0.86	2.08
GLY 79		40.39	0.50	39.89	0.53	-
VAL 80		91.35	57.13	34.22	0.85	1.70
ALA 82		48.82	17.78	31.04	0.82	1.21
HIS 83	√	122.95	16.59	106.36	1.39	1.90
SER 86		92.24	83.54	8.70	0.92	3.38
HIS 87		103.93	82.00	21.93	0.88	2.98

1mr8 chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 3		79.71	36.02	43.69	1.83	2.90
GLU 4		140.82	83.37	57.45	0.83	1.94
LEU 5		135.21	0.90	134.31	0.64	1.30
GLU 6		72.92	23.35	49.57	0.54	3.44
ALA 8		47.11	5.02	42.09	0.66	0.88
LEU 9		100.33	7.35	92.98	0.89	1.51
ASN 10		67.83	62.17	5.66	1.87	1.59
SER 11		60.76	31.19	29.57	1.41	1.55
ILE 12		49.93	1.77	48.16	0.73	1.46
ILE 13	√	57.95	13.99	43.96	0.68	1.06
VAL 15		44.26	8.42	35.84	0.97	0.89
TYR 16	√	1.38	1.19	0.19	1.40	0.82
HIS 17	√	48.13	29.12	19.01	1.84	1.63
HIS 27		118.75	80.28	38.47	1.48	3.81
LEU 37	√	1.71	0.30	1.41	0.84	1.46
THR 40		88.22	78.09	10.13	1.43	2.72
GLU 41	√	57.38	10.79	46.59	0.35	1.49
CYS 42		8.71	0.23	8.48	0.62	1.94
PRO 43		50.17	20.02	30.15	0.11	4.36
TYR 45		95.43	90.75	4.68	0.98	3.58
PHE 68	√	55.04	3.53	51.51	0.65	1.80
GLN 69		124.43	96.60	27.83	1.28	4.10
PHE 71		9.61	0.28	9.33	0.72	1.07
LEU 72		62.62	12.17	50.45	0.88	2.68
LEU 74		1.40	0.08	1.32	0.35	0.86
VAL 75		34.96	10.61	24.35	0.89	1.10
ILE 76		106.13	49.25	56.88	0.76	0.82
MET 78		52.13	0.26	51.87	0.97	2.90
GLY 79		38.51	1.18	37.33	0.55	-
VAL 80		97.81	52.59	45.22	0.95	1.63
ALA 82		57.49	20.54	36.95	0.82	1.45
HIS 83	√	116.29	33.05	83.24	1.22	1.84
SER 86		88.11	66.89	21.22	1.33	1.52
HIS 87		136.53	83.34	53.19	1.06	4.23

1bzd chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
TRP 41		79.30	78.05	1.25	1.48	1.28
ILE 68		66.80	49.81	16.99	0.93	0.93
LYS 70		38.13	34.91	3.22	3.42	2.12
LYS 76		65.71	58.44	7.27	3.43	0.90
PHE 87	√	132.52	14.94	117.58	1.00	1.29
HIS 88		54.34	6.54	47.80	1.12	3.06
GLU 89	√	108.35	29.90	78.45	0.94	2.09
HIS 90		66.25	45.28	20.97	0.77	3.78
GLU 92		77.86	5.08	72.78	0.75	4.48
VAL 93		21.89	1.94	19.95	0.96	0.81
VAL 94		84.76	3.36	81.40	1.05	0.67
PHE 95	√	38.33	0.05	38.28	0.94	1.04
THR 96		71.69	43.19	28.50	1.59	1.05
TYR 105	√	18.66	12.72	5.94	1.69	0.89
ILE 107	√	2.21	0.46	1.75	1.03	1.36
TYR 114		150.22	93.28	56.94	1.45	1.60
SER 115		63.39	30.36	33.03	1.33	1.29
TYR 116	√	62.44	6.68	55.76	1.31	1.54
SER 117	√	67.58	35.83	31.75	1.36	1.51
THR 118	√	67.69	3.75	63.94	1.35	0.89
THR 119	√	79.83	50.96	28.87	1.44	1.00
ALA 120	√	60.16	8.30	51.86	0.24	0.71
VAL 122	√	84.25	57.24	27.01	0.93	1.20

1bzd chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ILE 68		62.05	46.88	15.17	0.82	1.49
LYS 76		67.73	64.21	3.52	3.36	2.65
PHE 87	√	134.53	16.60	117.93	1.00	1.32
HIS 88		47.36	7.89	39.47	0.91	1.62
GLU 89	√	111.13	34.49	76.64	0.91	1.79
HIS 90		61.59	40.70	20.89	0.89	3.86
GLU 92		76.21	9.97	66.24	0.96	1.18
VAL 93		21.70	0.66	21.04	1.07	1.28
VAL 94		88.96	5.08	83.88	0.90	1.44
PHE 95	√	37.09	0.00	37.09	0.94	1.84
THR 96		75.73	43.21	32.52	1.72	1.43
TYR 105	√	21.46	15.17	6.29	1.69	0.84
ILE 107	√	3.35	1.05	2.30	1.01	1.21
TYR 114		150.72	92.57	58.15	1.41	2.40
SER 115		61.25	29.45	31.80	1.41	2.41
TYR 116	√	71.99	6.26	65.73	1.74	1.41
SER 117	√	62.37	28.11	34.26	1.50	1.25
THR 118	√	64.75	1.03	63.72	0.92	0.70
THR 119	√	81.82	53.09	28.73	1.50	1.03
ALA 120	√	61.82	7.83	53.99	0.70	1.37
VAL 122	√	61.08	34.93	26.15	0.68	1.66

1cqk chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
GLN 8		80.19	60.00	20.19	1.94	2.44
TYR 10		101.03	7.34	93.69	1.87	1.03
THR 11	√	34.71	28.35	6.36	0.88	1.19
ILE 12		66.29	16.37	49.92	0.87	1.00
PRO 15		75.06	17.03	58.03	0.25	0.79
GLU 17		151.42	95.82	55.60	0.85	1.43
GLN 18		46.49	9.08	37.41	1.16	0.56
LYS 21		123.12	85.22	37.90	2.97	1.86
SER 25	√	15.49	6.22	9.27	1.56	0.41
THR 27		32.40	2.81	29.59	1.54	0.52
MET 29		60.09	1.35	58.74	0.91	1.24
THR 31		18.45	2.69	15.76	2.01	0.78
ASP 32	√	65.27	55.98	9.29	0.29	1.30
ASN 51		96.25	67.30	28.95	1.78	1.78
LYS 53		151.44	61.76	89.68	2.63	1.55
ASN 54		62.59	50.30	12.29	1.64	0.98
THR 55		73.92	6.19	67.73	1.64	0.68
GLN 56		144.40	128.86	15.54	1.63	2.41
ILE 58		89.42	31.54	57.88	1.03	1.39
MET 59		121.96	109.07	12.89	0.79	1.70
ASP 60		59.94	14.11	45.83	1.02	1.27
THR 61		142.08	84.00	58.08	1.72	2.04
PHE 66	√	50.31	0.71	49.60	0.95	0.59
TYR 68	√	81.80	0.43	81.37	1.68	0.64
SER 69		0.11	0.02	0.09	1.71	0.57
LYS 70	√	97.40	4.11	93.29	3.72	0.48
ASN 72		62.03	36.24	25.79	2.00	1.65
LYS 100	√	88.33	68.85	19.48	3.68	1.64

1cqk chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
GLN 208		79.35	59.42	19.93	2.15	0.94
TYR 210		105.22	8.21	97.01	2.04	0.54
THR 211	√	34.67	27.60	7.07	0.80	0.57
ILE 212		67.21	14.79	52.42	1.11	0.73
PRO 215		72.99	15.81	57.18	0.17	1.12
GLU 217		152.77	96.52	56.25	0.66	2.56
GLN 218		45.00	8.61	36.39	1.14	1.01
LYS 221		121.58	81.75	39.83	3.13	2.57
SER 225	√	15.06	6.20	8.86	1.47	0.66
THR 227		31.00	0.94	30.06	1.54	0.62
MET 229		60.91	1.32	59.59	0.90	0.64
THR 231		17.57	2.55	15.02	1.94	0.51
ASP 232	√	64.60	57.07	7.53	0.89	1.33
ASN 251		94.89	68.16	26.73	1.70	2.42
LYS 253		152.35	62.20	90.15	2.83	1.54
ASN 254		61.70	50.29	11.41	1.66	0.70
THR 255		73.71	6.21	67.50	1.60	0.51
GLN 256		143.06	129.24	13.82	1.75	2.68
ILE 258		89.20	34.65	54.55	0.92	2.04
MET 259		118.41	106.59	11.82	0.84	1.44
ASP 260		58.88	12.84	46.04	1.07	2.03
THR 261		143.79	81.75	62.04	1.37	1.83
PHE 266	√	47.65	0.63	47.02	0.66	0.59
VAL 267	√	2.93	2.90	0.03	1.18	1.28
TYR 268	√	81.45	0.37	81.08	1.77	0.64
SER 269		0.08	0.02	0.06	1.58	0.49
LYS 270	√	98.31	5.75	92.56	3.13	0.90
ASN 272		63.07	34.54	28.53	2.78	0.90
LYS 300	√	87.86	68.07	19.79	4.10	1.38

2btx chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 6		32.42	3.00	29.42	1.26	3.93
THR 8	√	108.10	107.47	0.63	1.04	4.85
SER 9		63.25	58.70	4.55	1.01	3.12
PRO10		111.57	85.47	26.10	0.12	4.34
ILE 11		86.98	54.08	32.90	0.49	5.49
MET27		67.49	52.71	14.78	1.07	3.16
ASP 30		93.78	36.40	57.38	0.26	5.49
CYS 33	√	72.55	72.30	0.25	0.56	1.87
SER 34		70.10	63.17	6.93	1.03	2.54
ARG 36	√	216.25	164.90	51.35	3.47	4.31
GLY 37	√	37.73	19.22	18.51	0.29	-
LYS 38		101.48	73.48	28.00	3.2	3.41
VAL 39		45.46	0.81	44.65	0.84	2.29
VAL 40		24.20	2.41	21.79	0.9	2.61
GLU 41		24.80	21.05	3.75	0.99	5.18
HIS 68		22.04	6.31	15.73	1.47	4.39
PRO 69		118.65	108.50	10.15	0.11	3.17
LYS 70		80.05	16.57	63.48	3.05	3.24
GLN 71		162.22	122.04	40.18	1.26	7.77
ARG 72		209.24	196.61	12.63	3.24	10.41

2btx chain B						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ARG 2		141.89	92.33	49.56	5.63	5.07
TYR 3		173.21	37.93	135.28	1.55	7.99
TYR 4		143.94	15.96	127.98	1.69	6.41
GLU 5		164.87	66.26	98.61	0.48	7.48
SER 6		97.81	68.25	29.56	1.73	2.02
SER 7		52.13	27.47	24.66	1.13	2.91
LEU 8		114.37	24.35	90.02	0.52	6.19
LYS 9		164.60	155.42	9.18	3.76	6.51
SER10		75.14	71.72	3.42	1.49	5.50
ASP 13		188.31	186.38	1.93	0.55	9.11

1qnz chain H						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 142	√	28.82	25.00	3.82	1.37	2.32
THR 143		89.12	14.30	74.82	1.01	1.63
TYR 144		44.82	33.31	11.51	1.8	1.70
PRO 145		45.94	0.87	45.07	0.1	2.56
GLU 147		27.05	6.58	20.47	0.91	1.21
ASN 162		37.69	2.93	34.76	2.73	0.63
PHE 163	√	3.22	1.76	1.46	0.17	2.60
HIS 164		58.34	5.23	53.11	1.44	1.07
TYR 166		163.98	92.88	71.10	1.15	2.14
SER 167		67.99	67.64	0.35	1.2	1.17
ASP 169		88.39	37.34	51.05	0.24	2.48
THR 170		39.96	39.57	0.39	0.98	1.35
ASN 171		78.42	68.04	10.38	1.62	1.39
ILE 210	√	0.10	0.00	0.10	0.91	1.23
HIS 211		50.46	2.22	48.24	0.58	1.01
TYR 212		58.49	57.30	1.19	1.89	2.16
ALA 215		93.17	31.69	61.48	0.72	1.10
TYR 216		199.24	184.62	14.62	1.69	3.61

1qnz chain P						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
LYS 233		120.94	102.48	18.46	3.14	2.91
SER 234		52.91	42.80	10.11	1.37	0.90
ILE 235		82.58	11.79	70.79	0.96	1.38
ARG 236		149.62	1.75	147.87	3.18	1.77
ILE 237		47.43	38.21	9.22	1.15	1.11
GLN 238		152.47	7.93	144.54	1.85	1.39
ARG 239		182.01	150.91	31.10	2.69	4.16
PRO 241		140.81	140.68	0.13	0.1	1.77
GLY 242		35.44	11.15	24.29	0.94	-
ARG 243		119.37	43.28	76.09	5.04	1.47
ALA 244		59.08	50.10	8.98	0.13	1.35
PHE 245		132.48	103.69	28.79	0.95	1.67
ILE 248		85.35	60.01	25.34	0.51	3.98

1ddh Chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
LEU 5		1.95	0.00	1.95	1.1	1.67
TYR 7	√	20.60	1.24	19.36	1.69	1.99
VAL 9		3.24	2.96	0.28	0.95	0.76
GLU 24		0.76	0.49	0.27	0.85	0.80
TYR 59		13.75	11.33	2.42	1.2	3.45
ARG 62		174.45	130.51	43.94	3.77	2.10
GLU 63		20.25	3.78	16.47	0.97	1.95
ARG 65		190.53	181.33	9.20	4.55	3.79
ARG 66		76.53	30.96	45.57	6.67	1.26
GLY 69		33.16	14.76	18.40	0.61	-
ASN 70		43.07	6.54	36.53	1.42	1.40
GLN 72		110.17	105.73	4.44	1.77	2.24
SER 73		68.24	5.39	62.85	1.7	2.45
PHE 74		9.62	1.13	8.49	0.93	2.29
VAL 76		71.46	49.98	21.48	0.36	1.98
ASP 77		56.45	0.02	56.43	0.83	2.50
THR 80		46.82	34.06	12.76	1.79	2.33
ALA 81		8.43	1.85	6.58	0.85	1.57
TYR 84		87.57	66.76	20.81	1.77	2.22
LEU 95		15.08	5.74	9.34	0.95	2.03
TRP 97		38.07	2.15	35.92	1.38	1.79
ALA 99		4.44	0.28	4.16	1.04	0.59
TRP 114		32.81	5.90	26.91	1.74	1.90
PHE 116		34.52	15.78	18.74	0.81	1.72
TYR 123	√	24.27	9.23	15.04	1.68	1.11
ILE 142	√	58.73	58.36	0.37	0.46	2.34
THR 143	√	14.14	0.69	13.45	1.93	1.59
LYS 146	√	132.03	70.40	61.63	3.85	2.10
TRP 147		44.51	4.57	39.94	1.75	1.71
ALA 150		77.81	69.05	8.76	0.83	2.36
ALA 152		25.27	6.07	19.20	0.67	2.72
ARG 155		177.69	132.92	44.77	5.25	7.08
ASP 156		7.50	2.12	5.38	0.93	5.42
TYR 159	√	42.93	5.66	37.27	0.96	2.93
GLU 163		84.64	53.18	31.46	0.4	2.80
CYS 164	√	0.13	0.00	0.13	1.1	1.20
GLU 166		73.33	70.84	2.49	0.83	2.37
TRP 167		57.66	21.61	36.05	1.63	2.01

1ddh Chain P						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ARG 1		286.99	50.23	236.76	3.30	3.30
GLY 2		62.00	2.57	59.43	1.99	-
PRO 3		142.31	3.20	139.11	1.53	1.99
GLY 4		61.30	13.47	47.83	1.10	-
ARG 5		156.89	4.64	152.25	2.51	1.53
ALA 6		78.87	34.55	44.32	1.93	1.10
PHE 7		195.29	132.59	62.70	1.93	2.51
VAL 8		98.82	27.01	71.81	2.56	1.93
THR 9		111.46	33.69	77.77	2.51	1.93
ILE 10		222.54	19.23	203.31	0.49	2.56

1fcc chain A						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
LEU 251	√	12.27	8.29	3.98	0.7	1.34
MET 252		48.06	3.41	44.65	0.84	2.85
ILE 253		131.07	24.18	106.89	1.58	0.86
SER 254		113.24	32.98	80.26	1.84	1.04
ARG 255		103.26	90.45	12.81	3.83	2.30
GLN 311		140.05	87.85	52.20	1.63	3.67
GLU 380		36.46	23.61	12.85	0.92	1.55
GLU 382		41.36	13.25	28.11	1.05	2.57
GLY 385		79.08	64.24	14.84	0.78	-
PRO 387		84.88	76.97	7.91	0.16	2.68
SER 426		2.03	0.37	1.66	2.07	0.84
MET 428		12.42	4.06	8.36	1.1	1.26
HIS 433		125.85	79.30	46.55	1.36	3.62
ASN 434		139.43	6.80	132.63	2.06	1.26
HIS 435		45.31	36.72	8.59	0.82	0.98
TYR 436		71.70	2.26	69.44	1.47	1.08
THR 437	√	34.59	31.59	3.00	1.36	0.71
GLN 438		88.98	53.44	35.54	1.84	4.22

1fcc chain C						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ALA 23		30.72	19.90	10.82	0.73	1.39
ALA 24		62.20	17.66	44.54	0.39	1.79
THR 25		57.34	29.75	27.59	2.39	1.86
GLU 27		61.00	1.39	59.61	0.78	1.06
LYS 28		153.10	11.61	141.49	4.28	1.80
LYS 31		76.79	3.17	73.62	2.44	1.14
GLN 32		105.43	46.75	58.68	1.57	2.67
ALA 34		1.66	0.00	1.66	0.84	1.18
ASN 35		112.65	32.21	80.44	1.96	1.57
VAL 39		3.27	2.66	0.61	0.96	1.88
ASP 40		111.62	70.55	41.07	0.53	4.14
GLY 41		29.61	23.94	5.67	0.45	-
GLU 42		135.43	96.58	38.85	0.98	2.59
TRP 43		65.69	10.16	55.53	1.67	0.68
THR 44		76.29	62.69	13.60	1.55	5.44
TYR 45		72.05	45.55	26.50	0.73	1.76
PHE 52		6.02	0.60	5.42	0.39	2.01

1bjr chainE						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
THR 40	√	0.15	0.00	0.15	1.79	0.49
GLY 66		31.53	28.80	2.73	0.32	-
ASN 67		37.61	3.52	34.09	1.78	0.93
HIS 69	√	39.89	0.00	39.89	1.05	0.82
LEU 96	√	12.20	0.00	12.20	1.06	1.83
ASP 98		43.43	38.72	4.71	0.18	1.53
ASN 99		127.75	108.54	19.21	1.59	1.28
GLY 100		30.84	4.81	26.03	0.91	-
SER 101		61.72	56.68	5.04	1.84	2.71
GLY 102	√	32.96	30.28	2.68	0.36	-
ILE 107		6.22	3.33	2.89	0.47	1.08
SER 132	√	6.57	2.19	4.38	1.38	1.36
LEU 133		19.57	0.15	19.42	0.15	1.94
GLY 134	√	52.09	6.80	45.29	0.71	-
GLY 135		28.76	22.00	6.76	0.82	-
VAL 157		0.46	0.00	0.46	1.12	1.80
ALA 158	√	8.72	0.00	8.72	0.22	1.10
ALA 159	√	0.06	0.00	0.06	0.44	1.11
GLY 160	√	16.72	0.00	16.72	0.08	-
ASN 161	√	81.62	46.68	34.94	2.46	0.97
ASN 162	√	72.29	60.28	12.01	1.06	1.75
TYR 169		50.03	41.22	8.81	1.47	1.56
PRO 171	√	0.69	0.00	0.69	0.04	1.15
ALA 172	√	4.36	0.00	4.36	0.2	0.83
TRP 212		47.05	34.26	12.79	1.33	0.93
ARG 218		118.30	107.46	10.84	3.87	1.70
ILE 220		38.36	21.20	17.16	0.88	1.51
SER 221		57.74	47.12	10.62	1.56	0.70
GLY 222	√	0.36	0.20	0.16	0.88	-
THR 223	√	1.72	0.00	1.72	1.98	0.77
SER 224	√	18.02	0.00	18.02	1.72	1.23
MET 225	√	3.33	0.00	3.33	0.86	1.31

1bjr Chain I						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
VAL 280		153.06	44.70	108.36	0.42	1.23
ALA 281		76.77	11.40	65.37	0.60	0.55
GLN 282		169.11	88.12	80.99	1.42	2.72
GLY 283		59.50	12.61	46.89	0.61	-
GLY 284		81.21	0.00	81.21	0.55	-
ALA 285		81.44	37.58	43.86	0.55	1.03
ALA 286		87.40	11.05	76.35	0.62	1.11
GLY 287		34.83	0.19	34.64	0.65	-
LEU 288		185.13	0.00	185.13	0.37	2.77
ALA 289		124.55	18.98	105.57	0.50	1.34

1ugh chainE						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
GLN 144	√	82.32	23.00	59.32	1.78	3.64
ASP 145	√	29.98	21.37	8.61	0.02	5.21
TYR 147	√	51.51	25.45	26.06	1.63	2.24
HIS 148	√	71.00	19.37	51.63	1.38	2.54
GLN 152	√	43.16	12.92	30.24	1.59	1.66
PHE 158	√	22.79	19.55	3.24	0.96	1.37
VAL 164		21.55	18.48	3.07	0.69	0.72
PRO 165		98.22	79.99	18.23	0.14	0.97
PRO 166	√	18.02	15.71	2.31	0.05	0.90
PRO 167	√	17.07	0.00	17.07	0.07	1.71
PRO 168	√	88.96	11.41	77.55	0.14	0.82
SER 169	√	15.96	0.70	15.26	2.96	1.30
ASN 172	√	22.81	22.55	0.26	1.75	0.52
ARG 210		53.57	49.95	3.62	3.96	2.52
HIS 212		136.46	116.58	19.88	1.16	4.49
GLN 213		83.74	51.99	31.75	1.17	2.49
ALA 214	√	62.96	3.28	59.68	0.64	1.50
ASN 215		76.25	9.20	67.05	1.87	1.35
SER 216	√	22.07	15.70	6.37	0.87	1.10
LYS 218		88.91	48.70	40.21	3.64	1.95
GLY 246	√	11.60	4.41	7.19	0.64	-
SER 247		92.14	31.20	60.94	1.88	1.42
TYR 248		93.23	50.29	42.94	1.77	2.00
LYS 251		160.14	137.13	23.01	3.33	3.96
ALA 267		21.67	20.81	0.86	0.28	0.86
HIS 268	√	45.00	8.59	36.41	1.42	2.50
SER 270	√	32.48	0.00	32.48	1.05	0.73
PRO 271	√	77.38	1.60	75.78	0.06	0.60
LEU 272	√	170.96	3.71	167.25	0.03	1.07
SER 273	√	27.33	2.42	24.91	1.47	0.97
VAL 274	√	2.82	1.26	1.56	0.4	0.67
TYR 275		184.62	131.07	53.55	1.29	1.29
ARG 276		200.46	125.37	75.09	4.61	3.63

1ugh Chain I						
RESİDUE	**HOT SPOT**	**Chain ASA**	**Complex ASA**	**Interface ASA**	**HBOND**	**SIDE CHAIN RMSD**
ASN 3		138.49	117.52	20.97	0.84	6.07
ILE 18		51.46	30.26	21.20	1.97	1.10
GLN 19		137.94	36.71	101.23	0.39	3.17
GLU 20		87.88	3.34	84.54	1.61	1.60
SER 21		52.07	10.63	41.44	0.73	0.56
ILE 22		81.48	17.56	63.92	0.94	0.83
LEU 23		85.48	11.86	73.62	0.88	0.74
MET 24		55.07	1.12	53.95	1.05	0.76
LEU 25		96.18	31.15	65.03	0.31	1.53
GLU 28		85.48	28.31	57.17	0.55	1.30
VAL 29		4.75	0.00	4.75	0.68	1.10
GLU 31		147.38	104.13	43.25	0.63	2.00
VAL 32		91.81	28.91	62.90	0.88	1.62
ILE 33		17.32	12.46	4.86	1.07	1.41
LEU 42		21.56	1.50	20.06	1.05	1.68
VAL 43		12.66	0.00	12.66	1.52	0.57
HIS 44		12.02	6.72	5.30	1.24	0.74
THR 45		19.85	0.00	19.85	0.73	0.90
ALA 46		0.79	0.01	0.78	1.87	0.74
TYR 47		93.41	83.93	9.48	1.77	1.66
ASN 54		52.13	27.82	24.31	1.49	1.83
MET 56		27.72	0.00	27.72	1.05	1.61
LEU 58		5.95	0.00	5.95	0.36	1.88
ASP 61		54.46	21.56	32.90	0.73	2.19
ALA 62		52.92	12.27	40.65	0.07	1.22
PRO 63		122.12	109.26	12.86	1.94	1.23
TYR 65		77.21	28.06	49.15	1.18	0.90
VAL 71		4.94	0.05	4.89	2.77	0.82
GLN 73		51.45	10.40	41.05	1.19	1.11
SER 75		87.34	82.50	4.84	1.76	3.58
ASN 76		117.89	116.86	1.03	1.12	2.12
GLY 77		47.51	19.58	27.93	1.46	-
ASN 79		62.39	53.20	9.19	1.94	1.34

Bibliography

[1]. Keskin, O., et al., Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci*, 1998. 7(12): p. 2578-86.

[2]. Glaser F., S., D. M., Vakser, I. A. & Ben-Tal N, Residue frequencies and pairing preferences at protein – protein interfaces. *Proteins: Struct. Funct. Genet.*, 2001. 43: p. 89-102.

[3]. Jones, S., A. Marin, and J.M. Thornton, Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng*, 2000. 13(2): p. 77-82.

[4]. Lo Conte, L., C. Chothia, and J. Janin, The atomic structure of protein-protein recognition sites. *J Mol Biol*, 1999. 285(5): p. 2177-98.

[5]. Jones, S. and J.M. Thornton, Analysis of protein-protein interaction sites using surface patches. *J Mol Biol*, 1997. 272(1): p. 121-32.

[6]. Jones, S. and J.M. Thornton, Principles of protein-protein interactions. *Proc Natl Acad Sci U S A*, 1996. 93(1): p. 13-20.

[7]. Ofran, Y. and B. Rost, Analysing six types of protein-protein interfaces. *J Mol Biol*, 2003. 325(2): p. 377-87.

[8]. Ma, B., et al., Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A*, 2003. 100(10): p. 5772-7.

[9]. Bogan, A.A. and K.S. Thorn, Anatomy of hot spots in protein interfaces. *J Mol Biol*, 1998. 280(1): p. 1-9.

[10]. Clackson, T. and J.A. Wells, A hot spot of binding energy in a hormone-receptor interface. *Science*, 1995. 267(5196): p. 383-6.

[11]. Lowman, H.B., et al., Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry*, 1991. 30(45): p. 10832-8.

[12]. Kortemme, T. and D. Baker, A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A*, 2002. 99(22): p. 14116-21.

[13]. Hu, Z., et al., Conservation of polar residues as hot spots at protein interfaces. *Proteins*, 2000. 39(4): p. 331-42.

[14]. Valdar, W.S. and J.M. Thornton, Conservation helps to identify biologically relevant crystal contacts. *J Mol Biol*, 2001. 313(2): p. 399-416.

[15]. R.E. Bell and N. Ben-Tal., In silico identification of protein interfaces" Comparative and Functional Genomics. 2003. 4: p. 420-423.

[16]. Fraser, H.B., et al., Evolutionary rate in the protein interaction network. *Science*, 2002. 296(5568): p. 750-2.

[17]. Glaser, F., et al., ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 2003. 19(1): p. 163-4.

[18]. Thomas E. Creighton, *Proteins: structures and molecular properties*. 1993: New York: W.H. Freeman.

[19]. Jeong, H., et al., Lethality and centrality in protein networks. *Nature*, 2001. 411(6833): p. 41-2.

[20]. Kleanthous, C., *Protein-protein recognition*. 2000: Oxford; New York: Oxford University Press.

[21]. Halperin, I., et al., Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 2002. 47(4): p. 409-43.

[22]. Janin, J. and C. Chothia, The structure of protein-protein recognition sites. *J Biol Chem*, 1990. 265(27): p. 16027-30.

[23]. Tsai, C.J., D. Xu, and R. Nussinov, Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Sci*, 1997. 6(9): p. 1793-805.

[24]. Lo Conte L., C.C., Janin J., an evolutionary trace method defines bindin surfaces common to protein families. *J. Mol. Biol*, 1996. 257: p. 342-358.

[25]. Jones, S. and J.M. Thornton, Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol*, 1995. 63(1): p. 31-65.

[26]. Nooren, I.M. and J.M. Thornton, Diversity of protein-protein interactions. *Embo J*, 2003. 22(14): p. 3486-92.

[27]. Jin, L., B.M. Fendly, and J.A. Wells, High resolution functional analysis of antibody-antigen interactions. *J Mol Biol*, 1992. 226(3): p. 851-65.

[28]. Brooijmans, N., K.A. Sharp, and I.D. Kuntz, Stability of macromolecular complexes. *Proteins*, 2002. 48(4): p. 645-53.

[29]. Nooren, I.M. and J.M. Thornton, Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*, 2003. 325(5): p. 991-1018.

[30]. Young, L., R.L. Jernigan, and D.G. Covell, A role for surface hydrophobicity in protein-protein recognition. *Protein Sci*, 1994. 3(5): p. 717-29.

[31]. Sheinerman, F.B., R. Norel, and B. Honig, Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol*, 2000. 10(2): p. 153-9.

[32]. McDonald, I.K. and J.M. Thornton, Satisfying hydrogen bonding potential in proteins. *J Mol Biol*, 1994. 238(5): p. 777-93.

[33]. Kollman PA, Calculating structures and free energies of Complex Molecules: combining molecular mechanics and Continuum Models. *Acc. Chem. Res*, 2000. 33: p. 889 – 897.

[34]. Massova I, K.P., Computational alanine scanning to probe protein – protein interations: a novel approach to evaluate binding free energies. *J Am. Chem. Soc.*, 1999. 121: p. 8122-8139.

[35]. Kimura, S.R., et al., Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys J*, 2001. 80(2): p. 635-42.

[36]. Rajamani, D., et al., Anchor residues in protein-protein interactions. *Proc Natl Acad Sci U S A*, 2004. 101(31): p. 11287-92.

[37]. MacKerell, A.D., Jr.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.;Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M., *J Phys Chem B*, 1998. 102: p. 3586.

[38]. Ma, B., H.J. Wolfson, and R. Nussinov, Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr Opin Struct Biol*, 2001. 11(3): p. 364-9.

[39]. Keskin, O., B. Ma, and R. Nussinov, Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol*, 2005. 345(5): p. 1281-94.

[40]. Thorn, K.S. and A.A. Bogan, ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, 2001. 17(3): p. 284-5.

[41]. Kortemme, T., D.E. Kim, and D. Baker, Computational alanine scanning of protein-protein interfaces. *Sci STKE*, 2004. 2004(219): p. pl2.

[42]. Guerois, R., J.E. Nielsen, and L. Serrano, Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*, 2002. 320(2): p. 369-87.

[43]. Lichtarge, O., H.R. Bourne, and F.E. Cohen, An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol*, 1996. 257(2): p. 342-58.

[44]. Shatsky, M., R. Nussinov, and H.J. Wolfson, Flexible protein alignment and hinge detection. *Proteins*, 2002. 48(2): p. 242-56.

[45]. Bhat, T.N., et al., Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc Natl Acad Sci U S A*, 1994. 91(3): p. 1089-93.

[46]. Sauer-Eriksson, A.E., Kleywegt, G. J., Uhlen, M., Jones, T. A., Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. Structure. 1995. 15(3): p. 265-278.

[47]. Tugarinov, V., et al., NMR structure of an anti-gp120 antibody complex with a V3 peptide reveals a surface important for co-receptor binding. *Structure Fold Des*, 2000. 8(4): p. 385-95.

[48]. Pokkuluri, P.R., et al., Tertiary structure of human lambda 6 light chains. *Amyloid*, 1999. 6(3): p. 165-71.

[49]. Schormann, N., J.R. Murrell, and M.D. Benson, Tertiary structures of amyloidogenic and non-amyloidogenic transthyretin variants: new model for amyloid fibril formation. *Amyloid*, 1998. 5(3): p. 175-87.

[50]. Thies, M.J., et al., Folding and association of the antibody domain CH3: prolyl isomerization preceeds dimerization. *J Mol Biol*, 1999. 293(1): p. 67-79.

[51]. Ishikawa, K., et al., The structure of human MRP8, a member of the S100 calcium-binding protein family, by MAD phasing at 1.9 A resolution. *Acta Crystallogr D Biol Crystallogr*, 2000. 56 (Pt 5): p. 559-66.

[52]. McPhalen, C.A. and M.N. James, Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry*, 1988. 27(17): p. 6582-98.

[53]. Zhu, Y., Q. Huang, and C. Chi, Crystal structure of mung bean inhibitor lysine active fragment complex with bovine beta-trypsin at 1.8A resolution. *J Biomol Struct Dyn*, 1999. 16(6): p. 1219-24.

[54]. Mol, C.D., et al., Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, 1995. 82(5): p. 701-8.

[55]. Ratnaparkhi, G.S. and R. Varadarajan, X-ray crystallographic studies of the denaturation of ribonuclease S. *Proteins*, 1999. 36(3): p. 282-94.

[56]. Li, H., et al., Three-dimensional structure of H-2Dd complexed with an immunodominant peptide from human immunodeficiency virus envelope glycoprotein 120. *J Mol Biol*, 1998. 283(1): p. 179-91.

[57]. Singh, T.P., et al., Crystal structure of a complex formed between proteolytically-generated lactoferrin fragment and proteinase K. *Proteins*, 1998. 33(1): p. 30-8.

[58]. Scherf, T., et al., Three-dimensional solution structure of the complex of alpha-bungarotoxin with a library-derived peptide. *Proc Natl Acad Sci U S A*, 1997. 94(12): p. 6059-64.

[59]. el-Bastawissy, E., M.H. Knaggs, and I.H. Gilbert, Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature. *J Mol Graph Model*, 2001. 20(2): p. 145-54.

[60]. Leach, A.R., *Molecular Modelling Principles and Applications*. Second Edition ed. 2001: Pearson Prentice Hall. 26-264.

[61]. Tomas-Vert, F., Artificial neural network applied to the discrimination of antibacterial activity by topological methods. *J. Mol. Struc.*, 2000. 504: p. 249.

[62]. D.A. Pearlman, D.A.C., J.W. Caldwell, W.R. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun*, 1995. 91: p. 1-41.

[63]. B. R. Brooks, R.E.B., B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.*, 1983. 4: p. 187-217.

[64]. H. J. C. Berendsen, D.v.d.S.a.R.v.D., GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comp. Phys. Commun*, 1995. 91: p. 43-56.

[65]. Cornell, W.D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M. Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 1995. 117: p. 5179-5197.

[66]. Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F., & Postma, J. P. M, A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers*, 1984. 23.

[67]. Damm, W., A. Frontera, J. Tirado-Rives and W. L. Jorgensen, OPLS All-Atom Force Field for Carbohydrates. *J. Comp. Chem.*, 1997. 18: p. 1955-1970.

[68]. Dudek, M.J.R., K.; Ponder, J. W., *J Comput Chem*, 1998. 19: p. 548.

[69]. M., C. *http://www.lsbu.ac.uk/water/models.html*. 2004 [cited.

[70]. Guillot, B., A reappraisal of what we have learned during three decades of computer simulations on water. *J. Mol. Liq.*, 2002. 101: p. 219-260.

[71]. Laxmikant Kalé, R.S., Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten, NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 1999. 151: p. 283-312.

[72]. Humphrey, W., Dalke, A. and Schulten, K., VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 1996. 14: p. 33-38.

[73]. Darden, T., D. York, and L. Pedersen, Particle mesh Ewald: a N • log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 1993. 98: p. 10089-10092.

[74]. Berman, H.M., et al., The Protein Data Bank. *Nucleic Acids Res*, 2000. 28(1): p. 235-42.

[75]. Altschul, S.F., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997. 25(17): p. 3389-402.

[76]. Bairoch, A. and R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*, 2000. 28(1): p. 45-8.

[77]. Bairoch, A., et al., The Universal Protein Resource (UniProt). *Nucleic Acids Res*, 2005. 33(Database issue): p. D154-9.

[78]. Thompson, J.D., D.G. Higgins, and T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 1994. 22(22): p. 4673-80.

[79]. Felsenstein, J., Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol*, 1996. 266: p. 418-27.

[80]. Richards, l.a., *J. Mol. Biol.*, 1971. 55: p. 379-400.

[81]. Aytuna, A.S., *High Performance Algorithm for Prediction of Protein – Protein Interactions*, in *Electrical and Computer Engineering*. 2005, Koc University: Istanbul.

Vita

Sabri Bora Erdemli was born in Ankara, Turkey, on August 19,1981. He received his B.Sc. degree in Chemical Engineering from Middle East Technical University, Ankara, in 2003. He worked as a visiting researcher in his first year of graduate school for National Cancer Institute (NCI). From September 2003 to August 2005, he worked as a teaching and research assistant in Koc University, Istanbul, Turkey and studied to develop *"ANALYSIS OF HOT SPOTS IN PROTEIN PROTEIN INTERACTIONS"* project. He currently lives in Istanbul, Turkey and next year he will join Institute of Multiscale Modelling of Biological Interactions at Johns Hopkins University, Maryland USA to complete his PhD degree.