8-2010

STRUCTURAL BIOINFORMATICS BASED METHOD FOR PREDICTING THE INITIAL ADSORBED PROTEIN ORIENTATION ON A SURFACE

Aby Thyparambil
Clemson University, abyabrahamt@hotmail.com

STRUCTURAL BIOINFORMATICS BASED METHOD FOR PREDICTING THE INITIAL ADSORBED PROTEIN ORIENTATION ON A SURFACE

A Thesis

Presented to

the Graduate School of

Clemson University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Bioengineering

---

by

Aby Abraham Thyparambil

August 2010

Accepted by:

Dr. Robert A. Latour, Committee Chair

Dr. Alexey Vertegel

Dr. Delphine Dean

ABSTRACT

In any molecular simulation of protein-surface interaction, the selection of the initial orientation with which the protein would interact with the surface must be first made and is found to be critical in the determination of the bioactive state of the adsorbed protein. While various molecular simulation methods have been developed to identify the preferred orientation, these methods are generally computationally expensive and time consuming, especially for large molecules thereby motivating the current study.

The computational implementation for identifying a preferred orientation was done in MATLAB® and directly addresses the current research problem by assuming the protein to be rigid and mapping the number of solvent accessible residues that would interact with the surface as a function of orientation, thereby yielding a topography map that would reveal the potential minimum energy orientations for a given protein-interface interaction system. The protein orientation prediction has been performed for a wide range of proteins (11kDa - 300kDa) and surfaces (hydrophobic, hydrophilic, charged, biological-membranes) with the total runtime involved usually averaging in minutes. These results were also found to be in good agreement with the experimental and simulation results reported in the literature for biological and man-made materials. Besides the intended application for the support to molecular simulations, this program also has the general application of surface design to control the bioactive state of adsorbed proteins and to selectively target and immobilize protein in a controlled orientation.

DEDICATION

I would like to dedicate this thesis

To the Creator all things big and small, and in whose wisdom everything was created. I thank You for thy benevolence in providing me with everything that I have, especially for the endurance, health and the understanding to comprehend things far too greater than I could, that was required to complete this thesis. Everything I am is thy mercy.

To all my teachers for their unfailing support and encouragement whose blessings, guidance and support have inspired me to seek knowledge and wisdom.

To my parents, Mathew and Gelly Abraham and my dear brothers Alex and Alan Abraham, for without their financial, love, prayers, sacrifices and support all this would not have been possible.

To my friends who have stood by me through every thick and thin situations in life.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Robert Latour for his patience, support, and guidance throughout this project. I would also like to thank my committee members, Drs. Alexey Vertegel and Delphine Dean for their advice and support.

I would also like to thank all past and present members of biomolecular interactions group at Clemson, especially Balakrishnan Sivaraman and Yang Wei who went way out of their means in providing continuous encouragement and support.

A special thanks to Kevin Champagne and Sriram Ravindren for introducing the basic concepts of MATALB® and providing the guidance to implement the algorithm required to meet the specific goals of this study.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER 1

INTRODUCTION

Material implants that are designed to augment the function of damaged tissue or function as scaffold for the regeneration of tissue are generally termed as biomaterials.[1] The success of a biomaterial depends on the ability to integrate into the implant environment which in turn would depend on the proteins adsorbed on the interface/surface of the implant, as cellular responses within the host tissues are elicited due the bioactive state of the adsorbed protein.[2-4] The adsorption and bioactivity of proteins on solid interfaces is also of great importance and is an active field of research in other fields such as biotechnology[5, 6], biosensors[7, 8], chromatography[9], drug delivery[10], microarrays[11] and nanotechnology[12].

Current material design strategies for controlling the non-specific adsorption and bioactivity for protein-interfaces are limited to educated trial and error methods[13] but, considering the system-specific nature of the interacting protein and the interface, and the variability involved in designing an interface, the probability of finding an optimal design by such an approach is very low. Given this scenario, a detailed level of understanding in the molecular mechanisms involved in protein adsorption to interface is essential.

Experimental techniques involving protein adsorption usually attempt to determine the amount, the conformation and the final orientation of the protein adsorbed on the surface[6, 14-17] but, these techniques do not give significant details on the molecular mechanisms mediating the adsorption process and, in particularly, the orientation of the

tertiary structure of adsorbed proteins. Currently the most direct way for theoretical understanding adsorption processes is by molecular simulation, [4, 18-20] the applications of which fall into one or more of the following categories (1) protein orientation, (2) protein conformation and (3) bioactivity.[13, 21, 22]

It is generally recognized that the conformation and orientation in which the protein adsorb onto a surface is the key factor that determines the bioactivity of adsorbed protein layer.[4] However, evidence suggests that the surface-induced conformational shifts in the protein occur much slower than the initial changes in the orientation that occur on the surface.[23] Thus an efficient control on the bioactivity of the protein and prediction of the conformational shifts that might occur in a protein on any given interface are largely dependent on the orientation in which the protein interacts most favorably with the surface (i.e. at a minimum free energy of interaction).

Molecular simulations approaches to understand the conformation and bioactivity of proteins on an interface, however, have been limited as these approaches require the selection of an initial orientation of the protein, which may be randomly selected or, more appropriately, be based on interaction energy minima between protein and surface.[21] While random selection of the initial orientation of the protein is easily applied, molecular simulation methods for selecting the initial orientations based on interaction energy minimums of protein and surfaces represents a much more challenging problem, which can be very computationally expensive.[21, 23] Thus the selection of the most favorable orientation by a computationally inexpensive and rapid method is needed.

The current computational study to determine the preferred orientation of a protein for a given surface aims to circumvent the limitations of energy minimum interaction methods, by mapping out the compositional distribution and characteristics of amino acids on the surface of a given protein that are close enough to interact with the interface which could be considered to be a function of the adsorption free energy. In order to achieve the current goal, a structural bioinformatics based approach has been investigated and applied.

This thesis is structured to first provide the reader with the background knowledge in the mechanisms involved in the protein adsorption to surfaces. The objectives of this study are then outlined in the chapter following the background. The third chapter deals with the methods used to achieve the specific research aim, which is followed by a chapter on the results and discussion of the current study with different proteins on various substrates/interfaces. An appendix section is provided after the results and discussion, which details the implemented algorithms and the results that were not included in the main manuscript. References cited in this thesis are then included in the last section following the appendix.

CHAPTER TWO

BACKGROUND

2.1 Adsorption of Proteins to Interfaces

A model system for protein adsorption comprises of three components namely the adsorbent surface, protein, and solvent [4]. Each of these components affects the adsorption process differently and to varying extent. Given below are some of the factors which are known to affect the adsorption system.

Table 2.1: Summary of the factors influencing protein adsorption

System components	Factors
Adsorbent surface[4, 6, 9, 24, 25]	Area, composition, electrochemical properties, structure and surface free energy
Protein[4, 6, 9, 26]	Concentration, molecular weight, molecular size , polarity and type
Solvent [9,25]	Buffer type, pH, temperature and ionic strength

The way a protein interacts with the adsorbent surface and the solvent environment depends on the folding pattern of the protein. The protein folding in turn would give these protein molecules the characteristic molecular surface, active site, and shape. However, irrespective of the composition of the system, the transport of a protein from a solvent environment to the interface is mediated by diffusion through a stagnant layer immediately above the surface[27]. Since the diffusivity of low molecular-weight

molecules is much faster than the high molecular-weight molecules, smaller proteins approach the interface at a much faster rate than larger ones.

As the concentration of the protein in the stagnant layer depletes as a protein adsorbs to a surface, the resulting concentration gradient drives the further diffusion of protein into the stagnant layer from the surrounding bulk solution in order to further saturate the interface with adsorbed proteins. The rate of this diffusion process in turn affects the amount of time that an adsorbed protein has, before neighboring proteins adsorb next to it, which influences an adsorbed protein's ability to undergo conformational changes and re-orientation after it adsorbs.[26] Furthermore, the longer the residence time of an adsorbed protein is on a surface, the greater the number of van der Waals bonds that it makes with the surface, thus leading to increased probability that the protein will adsorb in an irreversible manner.[4]

Given the complexity and variability involved in the adsorption system, one of the first steps in understanding protein adsorptive behavior is to first understand the thermodynamic parameters involved in protein folding in aqueous solution.

2.2 Thermodynamic Perspective on Protein Folding

Most proteins are composed of the set of the 20 naturally occurring L-amino acids that are coded for by DNA. The specific sequence of amino acids along a polypeptide chain is designated as the primary structure, with this chain then being folded into higher order structural elements, which are referred to as secondary structure, tertiary structure and quaternary structure. The folded structure of a polypeptide chain depends on the

primary sequence of the chain, the solvent environment (usually aqueous environment) and intra- and inter-polypeptide chain residue-residue interactions [4, 29]. The contribution to the stability of a protein structure in an aqueous environment could be broadly classified into:

a) Enthalpic[4] (e.g. van der Waals, hydrophobic or electrostatic interaction) and
b) Entropic[4] (e.g. Secondary structure packing, cavity area, degree of solvation)

The primary driving forces in protein folding processes is attributed to the hydrophobic effect and the formation of intramolecular hydrogen bonds[28, 29]. Intramolecular hydrogen bonds refer to the interactions between hydrogen-bondable groups of the polar amino acids constituting the protein. Water being the natural solvent and a major contributory factor to the structure and stability of the protein; it is also important to consider the interactions of water with proteins as the water molecule is a highly hydrogen bondable structure.

In an aqueous phase it has been proposed that for a protein with a molecular weight of 15 kDa, a monolayer of water usually requires 600-1000 molecules, thereby bringing the ratio of protein: hydration mass to approximately 1:1.[30] When a monolayer of water surrounds the protein molecules, the estimated van der Waal radius of water around polar molecules is 1.4 Å and 1.6 Å with non-polar molecules.[30] Interactions of the polar molecules with the solvent to form hydrogen bonds decreases the enthalpy, while

the interaction of the non-polar molecules with the solvent, results in an ordered packing of the water molecules thereby increasing the entropic contribution of free energy (i.e., $-T\Delta G$) involved with the system.[4] This in turn affects the folding behavior of the protein molecule in an aqueous environment thereby attaining a compact and densely packed conformation.[28] While these interactions result in the general structure of a folded protein, they do not necessarily result in all non-polar amino acids occupying the interior of the protein and all polar amino acids being on the protein surface. [4, 28, 29] On the contrary, proteins fold such that amino acids of each type are contained at both the surface and buried within the protein's tertiary structure. It's generally attributed that of the total surface buried during the protein structural make up, over three-fifth of the buried surface area occur within the secondary structures while the rest occur between them.[30] The buried surface area in a protein is directly related to the hydrophobic energy that would help to compensate for the loss in conformational entropy occurring during the formation of organized structures.[30] Reduction of solvent accessible surface area is applicable to both the polar and non-polar molecules, but this effect is more pronounced as the size of the protein increases. Thus it has been recognized that the protein folded state in an aqueous environment is at a global energy minimum state, but the introduction of an adsorbent surface shifts the free energy minima of the entire system, resulting in the adsorption of the protein and likely shifts in a protein's folded structure.[28, 29]

2.3 Adsorption Thermodynamics

On a general basis, the interface between a solid and an aqueous phase is usually characterized by a high ΔG (more positive) and hence the system tries to reduce the ΔG

(more negative) by adsorbing molecules onto the interface. Adsorption of proteins from aqueous phase to a solid interface is thus an intrinsic attempt by the system to reduce the overall ΔG. If the net change in the ΔG of the system is lower (more negative) upon adsorption, the adsorption process is considered favorable and should spontaneously occur.[4]

2.4 Molecular Modeling

Molecular simulations approaches are currently the best way to theoretically determine the enthalpy, entropy and ΔG of the interface-protein-solvent system as the experimental determination of these individual components is currently difficult.[31] All-atom empirical force field and united atom methods are the commonly applied simulation methods to model protein adsorption behavior. Proper force parameterization or the equation describing the force acting on each atom as a function of their relative position is important for the accurate description of the events involved in protein adsorptive processes.[13]

Since the solvent molecules are an active component of the system comprising of protein-interface-solvent, simulations of protein adsorption should account for the interaction of the solvent with the protein either explicitly or implicitly. Explicit solvent models provide a more realistic picture of how a protein would interact in a 3D environment compared to implicit representations of the solvent. In addition, accurate parameterization of individual atom-atom pair-wise interactions between protein-solvent, protein-surface, and solvent-surface are also necessary, which is accomplished by

adjusting parameters in the Coulomb's law and Lennard-Jones terms in the potential energy equation. However the computational overhead of an explicitly modeled system is very high when compared to an implicitly modeled system; because of which the latter is preferred for systems involving proteins with high molecular weight.[13]

Implicit modeling of solvent molecules is applicable to proteins in an aqueous solution by approximating the behavior of solvent around the solvent excluded volume of the protein, which in turn would describe the solvation free energy of the solute due to the solvent. Currently there are primarily two types of implicit modeled systems: (a) Solvent accessible surface area (SASA) and (b) continuum electrostatics model. [13]

The continuum electrostatics model involving the Poisson-Boltzmann equation accounts for the enthalpic contribution in the protein-solvent interaction and usually best describes protein adsorption behavior in dilute ion concentrations.[32] However, these models do not account for the geometric and non-polar contribution of protein's interaction with the solvent. SASA on the other hand is a linear function of the solvation energy and accounts for the geometric, hydrophobic and in some part even to the dispersive components involved in protein-solvent interaction but the ionization effects are poorly described.[32-34]

Through modeling studies performed by Agashe et al, it has been observed that as the protein approaches an interface, it undergoes many rotational and translational motions before adsorbing onto the interface with a favorable orientation.[23] Once the protein adsorbs on the interface in the favorable orientation, the protein may or may not

undergo conformational shifts depending upon how conformational changes will influence the overall ΔG of the system.[22] Selection of the initial orientation is also important for molecular simulations involving protein conformational shifts and bioactivity.[13, 22]

Many modeling approaches have predicted initial orientation through a 3D configurational search usually requiring a computational time ranging from 24-260 hours per surface, which requires models of both the protein and the specific surface of interest that it is to be adsorbed.[21, 35-38] Hence a more rapid method for determining the preferred orientation (i.e., one that takes minutes versus hours to complete) would be much more desirable. Another shortcoming in most of these systems is the lack of a standardized coordinate system that is used to define the orientation of a given protein on a surface that is easily interpreted so that results from different studies can be readily compared that is readily applicable to all protein.[21, 37-39]

Peptide based adsorption studies on different functionalized surfaces have characterized the adsorption free energy of different types of amino acids (e.g., polar, non-polar, or charged amino acids) to different types of surface chemistries.[31] Assuming that the contributions of individual amino acid-surface interactions with surface functional groups to be additive, these individual interactions could be used to predict the preferred initial orientation of a protein on a given surface based on the knowledge of the distribution of amino acid types on the surface of the protein. In other words, by mapping out the characteristics and compositional distribution of amino acids on the surface of a

protein, adsorption free energy values could be estimated for every face of the protein from which preferred initial orientations could be predicted for which the protein structural data is required which could aid in modeling the protein structure. Protein structural data are atomic positions in the form of Cartesian coordinates that could be retrieved from available databases, such as the Protein Data Bank.[40]

2.5 Protein Data Bank (PDB) File Format

RCSB PDB Data Bank is an online protein structural coordinate database, which hosts several tens of thousands of protein structures.[41, 42] According to the file format of the latest release of PDB[42], each protein structure is stored in a separate file represented by a four character identifier of the form [0-9][a-z,0-9][a-z,0-9][a-z,0-9]. Every line in the file contains 80 characters and is called as a record. Each of these records is further sub-divided into a list of fields. Records could be broadly classified into:

a. Title: This section contains the record used to describe the experiment and the biological macromolecules present in the entry. Sub-records include HEADER, TITLE, COMPND, SOURCE, AUTHOR, JRNL, REMARK etc.

b. Primary structure: Describes the amino acids sequence in each chain of the macromolecule model.
c. Heterogen: The complete description of non-standard residues in the entry.
d. Secondary structure: This section describes the helices, sheets, and turns found in the entry.

e. Connectivity annotation: Describes the existence and location of disulfide bonds and other linkages.

f. Miscellaneous features: Describes the properties in the molecule such as environments surrounding a non-standard residue or the assembly of an active site.

g. Crystallographic and coordinate transformation: This section describes the geometry of the crystallographic experiment and the coordinate system transformations.

h. Coordinate: Collection of atomic coordinates in X Y Z format

i. Connectivity: This section contains information on atomic connectivity.

As on March 2010, there are 59,323 PDB deposits in this database which are exclusively protein molecules, of which over 85% are X-ray resolved structures and around 12% of which are NMR structures. Of the X-ray resolved structures, over 79% of the structures are resolved at 1 Å - 2.5 Å, with a majority falling between 1.5 Å – 2.5 Å. Resolution, in the context of model quality, is an average value for the uncertainty of atomic positions in a crystallographic model. On a general scale, the uncertainty of the position of an atom is roughly one fifth to one tenth of the resolution. Hence the quality of the model becomes extremely important during the residue mapping especially if a rigid model of the protein is being used. However this limitation is overcome if only the backbone of the model is considered as in almost all structures, with the backbone of a given structure generally being clear even at a resolution of 5 Å. An increasing number of

similar protein models are being resolved by one or more techniques and are being deposited in the databases. For example, models of hen-egg white lysozyme are available in NMR and X-Ray crystallographic data. However many of these models could be similar in sequence yet different in structure, while some are similar in both sequence and structure. Hence a screening criterion is required to separate very dissimilar structures and to group similar structures. This type of approach would be extremely helpful if a comparison is to be performed on the influence of environmental condition or single point mutation on the topology of a protein.

2.6 Structural Bioinformatics

Traditionally structural bioinformatics is used to analyze and predict the 3D structure of a protein.[43] These predictions and analyses could be based on sequence data[44] or structural data.[40] Strategies such as homology modeling and threading typically require a structural template onto which a user-defined sequence could be modeled, for which the quality of the model is of prime importance. Quality of a model could be affected by various factors involved in the processing of the structural coordinate files[45]. Some of the online servers dedicated for the quality control purposes are: WHAT IF[46]/WHATCHECK[45], PROCHECK[47], Verify3D [48].

Another application of the structural bioinformatics is in the structural alignment of protein structures, which would measure the divergence of the input structure with a target structure in units of Å. When the divergence of the target and input structure is less than 1 Å, the structures are considered to be similar. Examples of some online servers

are: DALI[49], GANGSTA+[50], MAMMOTH[51].Secondary structure prediction of helices, beta-sheets and turns has also been an active field of research. Compared to the initial strategies using a knowledge-based approach (e.g.; CHOU-FAS[52], DSSP[53] and GOR[54]), modern approaches include machine learning (ALB[55], GORIII[56]) and evolutionary information (PSIPRED[57] and PHD[58]). Other properties that are predicted based on sequence information with reasonable accuracy include absorbance, diffusivity, flexibility, fractal index, pI, solubility, and many others.[44] Each of these method are being continually revised and being predicted with better accuracy. Modeling the solvent environment to predict the solvent accessibility of a residue[59-62] and pKa[63] of an individual residue from structural coordinate file are also fields of active interest.

2.7 Limitations of Current Software Packages

Existing modeling packages such as SPDBV[64], AMBER[65], CHARMM[66], XPLOR[67], and Materials Studio®, though useful for dynamics simulation and energy minimization, are difficult to be used or do not support data visualization, graphing, rotational, translational or distance-based structure modeling. Besides many of these types of packages, such as Materials Studio®, are proprietary and source codes are not modifiable to user-requirements. On the other hand, software packages such as MATHEMATICA and MATLAB® do not directly support the computation required for protein modeling but are amenable to be tailored to user-requirements.

CHAPTER THREE

RESEARCH OBJECTIVES

The objective of this research work was to develop a rapid method based on structural bioinformatics to predict the initial/preferred orientation in which a protein would interact with a given surface. It is hypothesized that by mapping out and quantifying the residues on the protein surface that could interact with an surface, minimum energy patches within the protein structure could be identified, which correspond to residue distributions that are most complementary to a given surface. As a direct application, this could help in the rational selection of the preferred initial adsorbed orientations of a protein on a given surface thereby decreasing the run time involved in the configurational space search prior to conducting a molecular simulation. The map of potential residue interaction on surface will be referred to as the *'topography map'* of protein to generate which the following research objectives were identified:

a) Develop an initial positioning system which could be applied to all proteins.
b) Implement solvent parameterization in order to exclude buried residues from being quantified.
c) Perform a configurational space search to quantify the types of residues that should be interactive with a surface in each individual orientation.
d) Visualize the topographical map as a function of individual orientations
e) Validate the results by comparison with existing simulation and experimental results.

CHAPTER FOUR

METHODS

This chapter details and verifies the algorithms implemented for achieving research objectives (a)-(d). The algorithm on '*Standardizing the Initial Positioning of the Protein'* establishes a local coordinate system that is clearly defined based on the structure of a protein. This algorithm forms the basis for the following algorithm on *'Configuration Space Search'* to define all the possible orientations of a protein on a solid interface. Implicit solvation of residues were modeled and verified using the algorithms described in the '*Modeling the Solvent Accessibility for Amino Acids'* based on which the topography map was generated using the algorithm described in the '*Quantification of the Residues Interacting with the Interface*.'

The implementation of these algorithms were done in MATLAB® (R2009a) language because of the relative ease with which MATLAB® handles matrix and vector formulations.

4.1 Standardizing the Initial Positioning of the Protein

A given structural coordinate file can be visualized either based on a *global coordinate system* or by a *local coordinate system*. When the atomic coordinates are used as provided in the structural files, this positioning is called the *global coordinate system*. However, if the positioning of the atomic coordinates is relative, or in other words

'*translated*' to a defined coordinate based on the protein itself, such a positioning is called a *local coordinate system*.

Irrespective of the coordinate system followed, visualization tools plot the corresponding coordinates in a 3D spatial unit cell called the bounding box and automatically adjust the view by selecting its center, even though the origin may be off-screen. The XYZ coordinates available in a structural coordinate file in itself is the result of a series of refinement processes involved to fit a protein model to the 3D electron density map obtained by X-ray or NMR studies, which may not just involve one protein molecule but many and could naturally vary from one refinement procedure to another.[68]

Hence it is evident that a standardized positioning scheme does not exist among the visualization tools currently in use, nor within the structural coordinate files themselves so as to generate a well-defined, uniform starting position for a given protein model. This creates visualizing-tool and model-dependant variability in the initial positioning of the protein model, which can make it difficult to interpret and directly compare analyses performed by different investigators who use different types of visualization tools and models.

These shortcomings were overcome by first implementing a *local coordinate system* by *translating* the origin (0, 0, 0) of the given global coordinate system of a PDB file to the centre of mass of the protein (COM). COM was calculated for the PDB file by considering only the heavy atoms of the protein and excluding all heteroatoms. A Cartesian coordinate system was then defined at this COM by the following steps.

Step 1. Since all structural coordinate files of a protein have the first amino acid residue of the polypeptide chain designated as the N-terminus (Nter) and last amino acid designated as the C-terminus (Cter), a direction vector is generated from Nter→Cter, with a vector parallel to this direction and passing through the COM defined as the local coordinate system X-axis. The coordinates of the Nter and Cter were defined as the alpha carbon (Cα) coordinates constituting the first and last amino acid positions of the protein, respectively.

Step 2. A unit vector normal to the plane containing the direction vectors Nter→Cter and Nter→COM is determined.

Step 3. The Z-axis of the local coordinate system is defined to be parallel to the normal vector defined in Step 2 and is in the direction of the vector passing through the COM and the unit normal defined in Step 2.

Step 4. Finally, the Y-axis is defined as the vector cross-product of the +Z-axis vector and the X-axis vector.

Step 5. A bounding box is now defined based on the local coordinate system with the initial image displayed such that the X-axis is oriented in horizontal position relative to the graphics window with the +X-axis pointing to the right, the Y-axis is vertical with the +Y-axis pointing upward, with the +Z-axis then pointing out of the front of the screen.

The above described positioning method has been verified with over 100 PDB models. For illustrative purposes, PDB models of lysozyme with different global coordinate systems but with identical primary amino acid sequences and comparable root-mean-square deviations (RMSD) to one another were used. RMSD is the measure of the average distance between the atoms of superimposed proteins, thus serving as a sensitive indicator of similarity between structures. The higher the RMSD value is, the greater the deviation of the queried structure is from the subject structure. RMSD comparisons in the current study have been performed using the online web server of DALI[49]. DALI allows for pairwise database searching of homologous 3D structures.

Table 4.1: Summary of the pairwise alignment of PDB models 1GXV and 2EPE

Model 1 (Query)	**Model 2 (Subject)**	**Z-score**	**RMSD (Å)**	**Sequence identity (%)**
1GXV	2EPE	22.1	1.3	100

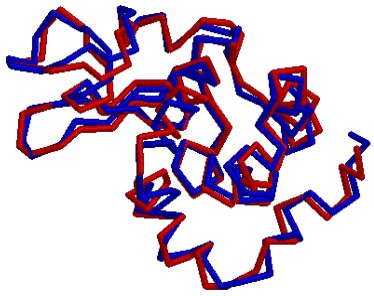
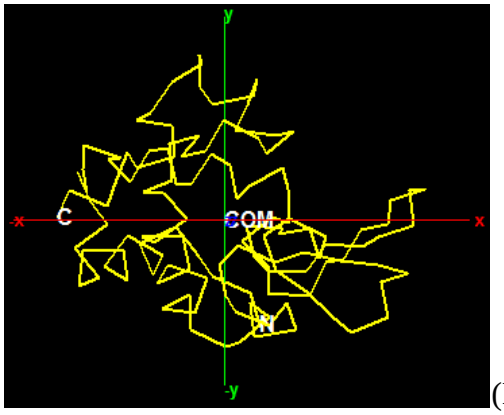


Figure 4.1: Superimposed PDB structures of lysozyme model (a) 1GXV (Red) and (b) 2EPE (Blue). The deviations in the structures are distinctly identified as regions where the protein structures do not exactly superimpose. As shown, the two lysozyme models represent very similar structures.

(A) 

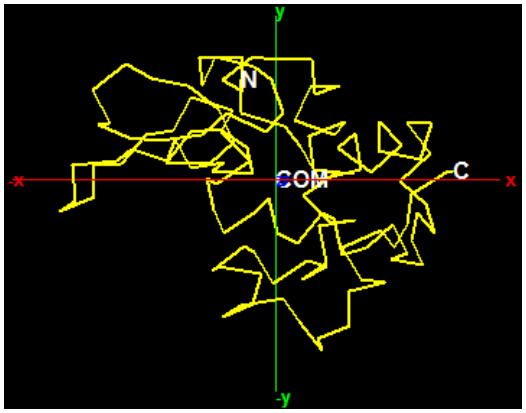
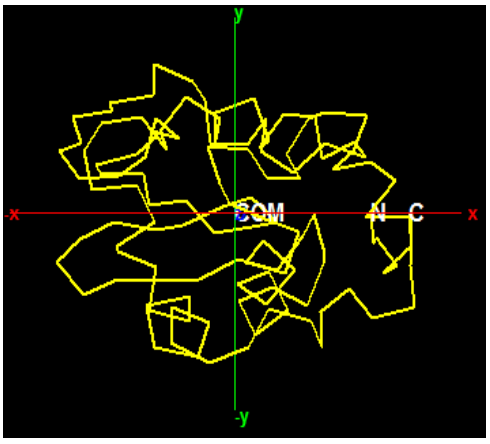
(B) 

Figure 4.2: PDB structures of lysozyme (a) 1GXV (b) 2EPE in the global coordinate system. 'N' represents the N-terminus and the 'C' represents the C-terminus. 'COM' represents the centre of mass. XYZ axes represented in red, green & blue respectively corresponds to the bounding box with the global coordinate system centered on the COM.

(A) 

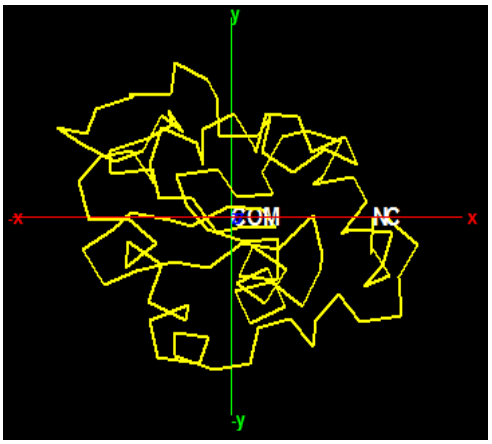
(B) 

Figure 4.3: PDB structures of lysozyme (A) 1GXV (B) 2EPE in the local coordinate system based on the method described above. Nter→Cter is oriented with the direction vector parallel to the X-axis, while the normal of the plane containing the Nter→Cter and Nter→COM is oriented parallel to the Z-axis of the bounding box. By orienting these two models using the local coordinate system, their similarity is clearly evident.

4.2 Configurational Space Search

Different orientations of the protein in a 3D space can be described by different combinations of rotation around the X-, Y-, and Z-axes. Depiction of the two Euler angles (Φ, θ) can be used to track the orientation of the protein, as shown in Fig 4.4. θ represents a clockwise rotation about the Y-axis, which rotates the X-axis to define the x'-axis, while Φ represents the rotation about the x'-axis.

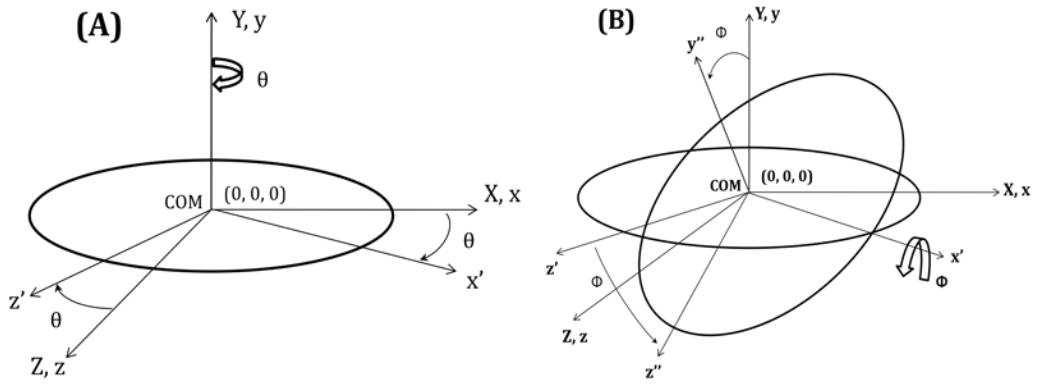


Figure 4.4: Depiction of the two Euler angles (Φ, θ) used to track the orientation of the protein relative to a local coordinate system. (A) Rotation of the protein about the Y-axis (Φ). (B) Rotation of the protein about the x'-axis (θ) following the Y-axis rotation.

Implementation of the above-described rotation scheme was incorporated in the custom programmed MATLAB® scripts through the direction cosine matrix (DCM) technique. Individual rotation about the Y-axis is represented by the following matrix:

$$\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad \textbf{(4.1)}$$

Individual rotation about the x'-axis is represented by the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Phi & -\sin\Phi \\ 0 & \sin\Phi & \cos\Phi \end{pmatrix} \quad \textbf{(4.2)}$$

For a combined rotation, the transformation matrix is given by the dot product of the rotation about Y- and X-axes.

$$\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Phi & -\sin\Phi \\ 0 & \sin\Phi & \cos\Phi \end{pmatrix} \quad \textbf{(4.3)}$$

In the current sampling method, the value of θ was varied from 0° to 360°, while Φ was varied from 0° to 180°. Φ is incremented by a user-defined angle 'Δ Φ', while θ was incremented by 'Δ θ', so as to maintain a constant area of analysis, which is given by the following expression:

$$\Delta\theta = \Delta\Phi / \sin(\Phi) \quad \textbf{(4.4)}$$

4.5 Modeling the Solvent Accessibility for Amino Acids

When a protein comes in contact with a nonpolar materials surface, dehydration of the interacting nonpolar groups of both the protein and materials surface is expected to occur.[4] This dehydration is due to fact that water molecules adjacent to the nonpolar

functional groups of both the protein and the materials surface tend to be in a higher free energy state compared to bulk water. Hence the dehydration associated with protein adsorption to a nonpolar surface result in a decrease in the free energy of the system. The more negative the ΔG value, the more strongly the protein should adsorb to the surface.[4]

A linear relationship has been demonstrated between the dehydration and the molecular weight of the protein (M) while adsorbing down to an air-water interface.[69] This relationship was shown to be equivalent to $M^{2/3}$ and the adsorption tendency was found to decrease by the same magnitude with increase in molecular weight of the protein for an air-water interface based adsorption system[69]. Molecular weight in turn, is directly proportional to the radii of the protein molecule.

The solvent accessible surface area (ASA) of a protein is the surface area of the molecule that is available for the solvent molecules to interact (Figure 4.5). The surface area of a protein is a linear function of its overall radius and has been empirically related to the molecular weight of the protein by the following expression[28, 30]:

$$ASA = 11.1 * M^{2/3} \quad \textbf{(4.5)}$$

If the reduction in ASA of a protein is associated with a decrease in ΔG, a larger drop in the ASA of a protein would imply a stronger tendency for the protein to be adsorbed. This would also indicate that the patches within the protein structure that experience more drop in the ASA on adsorption to a surface are more likely to be adsorbed strongly than the other patches with lower drop in ASA, thereby further

implying that the probability of a protein orienting in a given way when it contacts a surface is proportional to the drop in ASA for that particular orientation.

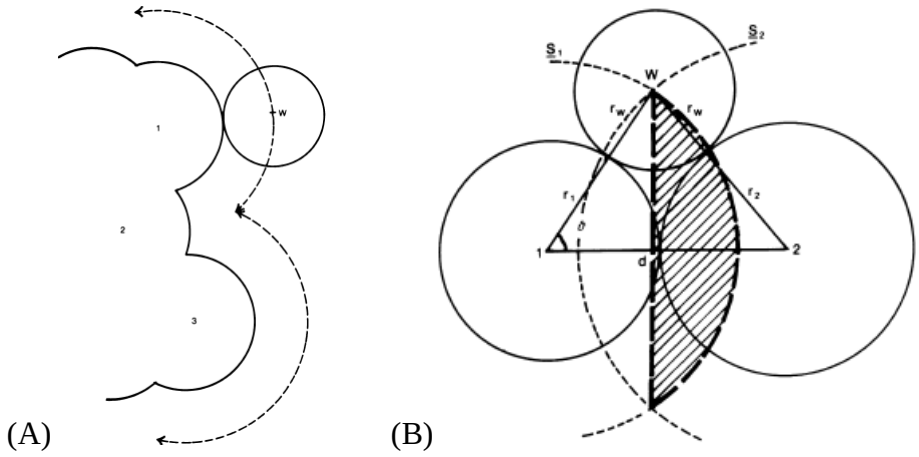


Figure 4.5: (A) ASA of the atoms (1, 2, and 3) formed by rolling the solvent probe 'w' with no penetration. (B) Surface area of '1' which is not accessible to probe 'w' due to '2' is represented by hatched region. 'r1' is the radius of atom '1' and 'r2' is radius of atom '2'. Adapted from ref [59]

ASA of the protein is usually modeled by rolling a solvent probe around the atoms constituting the protein's outer surface (Figure 4.5).[30] Since the initial algorithm for the calculation of ASA proposed by Richard and Lee[60] was very computationally intensive, various alternative approaches have been devised to calculate the ASA of individual amino acids forming the surface of a protein through analytical[59], geometrical[62] and statistical[63] techniques, with significant improvements in computation time and accuracy.

ASA calculation in the current implementation of the program is based on the analytical expression developed by Wodak and Janin.[59] The solvent accessibility is considered to be a function of only the inter-atomic distances between the COMs of the amino acids constituting the surface of the protein. Besides the improvement in the computation speed, other advantages in this model are:

a) The model is based on a pairwise-distance function of the COMs of the amino acid and is relatively independent of a lot of other variables involved in processing the structural file required to generate the XYZ coordinates of the atoms.
b) The model is reasonably accurate.
c) The model is adaptable to include minimization functions, as a result of which the decrease in the solvent accessibility per amino acids at the interfacial contact between protein-protein and protein-surface can be calculated with relative ease.

In the original algorithm proposed by Wodack and Janin, the van der Waal radii of amino acids were taken as inputs, which I have modified for my current implementation to accept van der Waal radii of atoms such as carbon, nitrogen, oxygen and sulfur (CNOS) as inputs. Such an alteration was done in order to predict the solvent accessibility of the protein and its amino acid residues that are exposed to different solvent environments, as the van der Waal radii of CNOS in different solvent environments are more readily available than the amino acids themselves. Table 4.3 through Table 4.5 validate the ASA values calculated by the current implementation for

the whole proteins using the inputs from Table 4.2 with probe radii = 1.4 Å and penetration depth (s) = 2.5 Å. The validity of this approach is indicated by comparing the ASA results obtained using my modified method to the results obtained by previously developed methods.

Table 4.2: User-defined van-der Waal radii

Carbon	**Nitrogen**	**Oxygen**	**Sulfur**
1.7 Å	1.52 Å	1.55 Å	1.80 Å

Table 4.3: Comparison of the standard and the calculated volumes of amino acids for the user-defined inputs given in Table 4.2. BL- stands for residue volumes of amino acids which are buried, not in contact with ligands and are not in contact with water.

Residue	**Volume (Å^3)**				
	Standard	**Computed**			
	Richards[30]	Harpaz[70]	Pontius[70]	BL-[70]	Present study
Gly	66.4 ± 4.7	63.8	67.5	64.9	67.2
Ala	91.5 ± 6.7	90.1	91.5	90	85.8
Val	141.7 ± 8.4	139.1	138.4	139	122.7
Leu	167.9 ± 10.2	164.6	163.4	164	143.3
Ile	168.8 ± 9.8	164.9	162.6	163.9	143.3
Met	170.8 ± 8.9	167.7	165.9	167	146.9
Pro	129.3 ± 7.3	123.1	123.4	122.9	124.1
His	167.3 ± 7.4	159.3	162.3	160	174.3
Phe	203.4 ± 10.3	193.5	198.8	191.9	196.7

Tyr	203.6 ± 9.6	197.1	209.8	197	211.4
Trp	237.6 ± 13.6	231.7	237.2	228.2	249.3
Cys	105.6 ± 6.0	103.5	102.4	103.3	107.7
Ser	99.1 ± 7.4	94.2	102	95.4	100.2
Thr	122.1 ± 6.7	120	126	121.5	118.9
Asn	135.2 ± 10.1	127.5	138.3	124.7	134.5
Gln	161.1 ± 13.0	149.4	156.4	149.4	153
Asp	124.5 ± 7.7	117.1	135.2	117.3	133.6
Glu	155.1± 11.4	140.8	154.6	142.2	152.1
Lys	171.3 ± 6.8	170	162.5	167.3	173.1
Arg	-	192.8	196.1	194	187.9

Table 4.4: Comparison of the standard and calculated van der Waal radii of amino acids computed from table 4.3 using the following equation $\sqrt[3]{0.75 * Volume\ of\ amino\ acid}$.

Residue	**Radius (Å)**				
	Standard	**Computed**			
	Richards[30]	Harpaz[71]	Pontius[71]	BL-[71]	Present study
Gly	2.51 ± 0.06	2.48	2.52	2.49	2.52
Ala	2.80 ± 0.07	2.78	2.80	2.78	2.74
Val	3.23 ± 0.07	3.21	3.21	3.21	3.08
Leu	3.42 ± 0.07	3.40	3.39	3.40	3.25
Ile	3.43 ± 0.07	3.4	3.39	3.40	3.25
Met	3.44 ± 0.06	3.42	3.41	3.42	3.27
Pro	3.14 ± 0.05	3.09	3.09	3.08	3.09

His	3.42± 0.05	3.36	3.38	3.37	3.47
Phe	3.65 ± 0.05	3.59	3.62	3.58	3.61
Tyr	3.65 ± 0.06	3.61	3.69	3.61	3.70
Trp	3.84 ± 0.07	3.81	3.84	3.79	3.90
Cys	2.93 ± 0.05	2.91	2.91	2.91	2.95
Ser	2.87 ± 0.06	2.82	2.90	2.84	2.88
Thr	3.08 ± 0.06	3.06	3.11	3.07	3.05
Asn	3.18 ± 0.08	3.12	3.21	3.10	3.18
Gln	3.37 ± 0.06	3.29	3.34	3.29	3.32
Asp	3.10 ± 0.07	3.04	3.18	3.04	3.17
Glu	3.33 ± 0.08	3.23	3.33	3.24	3.31
Lys	3.45 ± 0.10	3.44	3.39	3.42	3.35
Arg	-	3.58	3.60	3.59	3.55

Table 4.5: Comparison of the ASA calculated by the current method and other methods as described in the literature, for the user-defined inputs given in Table 4.2, Solvent radii = 1.4Å and penetration depth (s) = 2.5Å.

PDB Model	**Fragment**	**Accessible surface area (Å^2)**			**Ratio**	
		Lee and Richards[59] (a)	Wodack and Janin (b)[59]	Present study (c)	(c)/(a)	(c)/(b)
Phage T4 lysozyme (2LZM)	All	8962	8767	8633	0.96	0.98
Human deoxyhemoglobin (2W6V)	A	7829	7833	6972	0.89	0.89
	B	8226	8257	7657	0.93	0.93

	αβ Dimer	14425	14660	14162	0.98	0.97
Trypsin-BPTI complex (3BTK)	Trypsin	8902	9530	9558	1.07	1.00
	BPTI	3556	3635	3168	0.89	0.88
	Complex	11075	11930	12026	1.09	1.00
Lobster GPDH (4GPD)	Monomer	15470	16120	16118	1.04	1.00
	Red/Blue Dimer	28020	29290	29788	1.06	0.98
	Red/Green Dimer	28814	30100	30027	1.04	1.00
	Tetramer	50310	53880	59883	1.19	1.11
Dogfish apo-LDH (3LDH)	All	17120	17867	18267	1.07	1.02
Concanavalin A (2YZ4)	All	10600	11350	11592	1.09	1.02

As shown by the comparisons presented in Tables 4.3 to 4.5, the current implementation provides values that are in very reasonable agreement with previously published methods.

Another goal of modeling the ASA is to determine the change in solvent accessible surface area of the individual amino acid residues constituting the protein when the protein unfolds. ASA of residues could be related to the transfer energy function (ΔG) from the protein interior to the interface and is given by the following equation:

$$\mathbf{\Delta G} = \text{Solvation parameter } (\boldsymbol{\sigma})\mathbf{*ASA} \qquad \mathbf{(4.6)}$$

The current implementation however, does not compute the energy function associated with the residues using the above equation, but rather distinguishes the residues interacting with the solvent (surface forming) from the ones that would not interact with the solvent (buried) through the computed ASA of the individual amino acids, thereby not accounting for interactions of buried residues but only accounting for surface-forming residues. The rationale behind such an approximation is because the following:

a) The kinetics of adsorbent-surface induced conformational shifts that occur when a protein adsorbs to a surface have been shown to be slower than the kinetics of protein orientation[23], and
b) The shielding effect provided by the residues immediately surrounding the buried residue prevents the interactions of buried residues with the solvent.

Combining these two concepts, it can be assumed that the contribution to adsorption energy from buried residues is negligible compared to the solvent-accessible residues on the surface of the protein.

Currently, there are no formal definitions to determine if a particular residue would be buried or not based on the computed ASA. However for our objective, residues having relative ASA values less than 10% are considered to be buried, but the custom designed script is flexible to accept other user-defined cutoffs.

Comparison of the relative accessibility of the individual residues constituting the 1AKI structure of lysozyme has been made to compare the ASA method that have been implemented in the current program with other commonly used tools such as the SurfV tool of STING Millenium suite[63], ASAView[61] and GETAREA[62] (see Fig 4.6). ASA of the whole protein complex for the model 1AKI is 6523 $Å^2$, which is in good agreement with 6571 ± 81 $Å^2$ as reported in the literature[72]. While the ASA calculation in ASAView[61] is based on electrostatic interaction function, GETAREA[62] utilizes a modified lattice algorithm. SurfV[63], on the other hand, utilizes a rolling sphere model to determine the accessibility of the individual amino acids. The root-mean-square errors (RMSE) between the different tools are compared in Table 4.6. ASA was found to be independent of RMSD (Fig 4.7) and PDB-reported structural resolution (Fig 4.8)

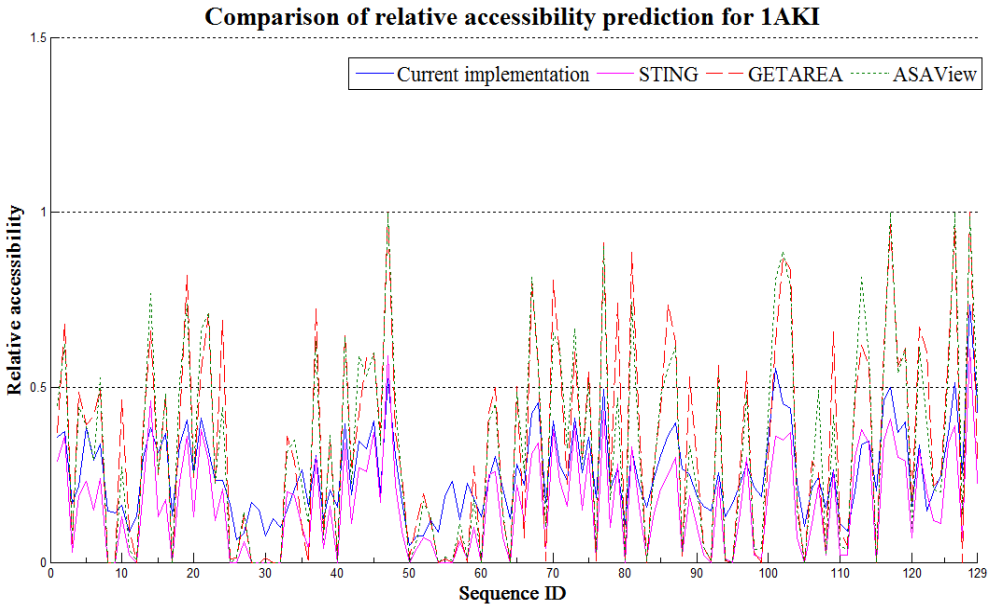


Figure 4.6: Comparison of the relative accessibility prediction by ASA of individual amino acids calculated for lysozyme model (1AKI) by the current method and other

commonly used tools. Solvent radii = 1.4Å, penetration depth (s) = 2.5Å and user-defined inputs as given in table 4.2.

Table 4.6: RMSE between the different tools used to predict ASA

Method	Current implementation	STING[63]	GETAREA[62]	ASAView[61]
Current implementation	NA	0.11	0.22	0.20
STING	0.11	NA	0.23	0.22
GETAREA	0.22	0.23	NA	0.08
ASAView	0.20	0.22	0.08	NA

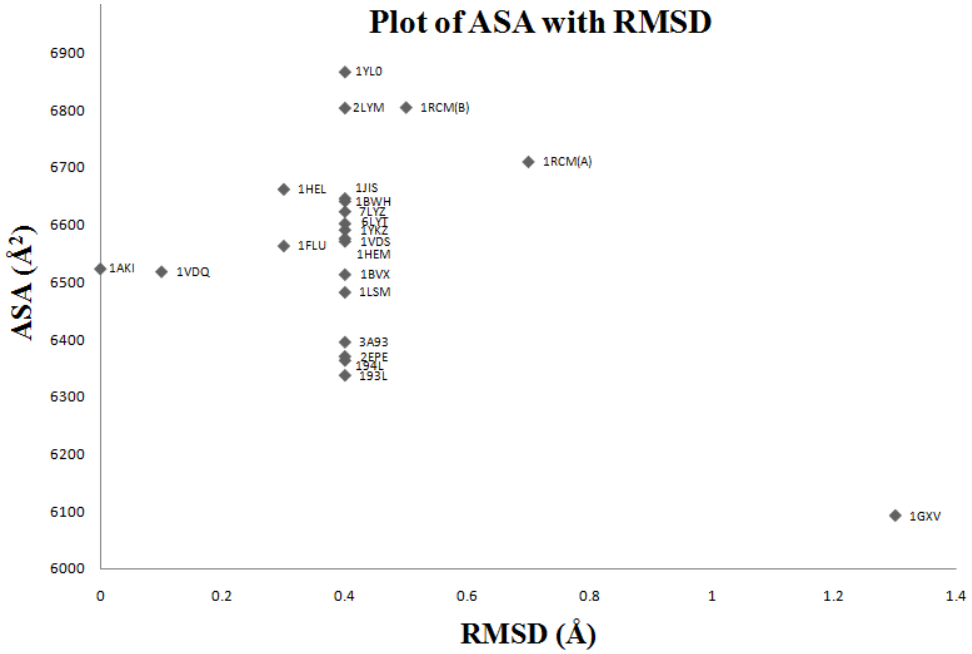


Figure 4.7: Dependence of ASA on RMSD in different lysozyme models, but which are identical in sequence. Reference structure: 1AKI, Solvent radii = 1.4Å, penetration depth (s) = 2.5Å and user-defined inputs as given in Table 4.2. Correlation (R^2) = 0.146.

Figure 4.8: Dependence of ASA on resolution in different lysozyme models, but which are identical in sequence. Reference structure: 1AKI, Solvent radii = 1.4Å, penetration depth (s) = 2.5Å and user-defined inputs as given in Table 2. Correlation (R^2) = 0.114. Note that a higher resolution values represents greater uncertainty in the atomic structure.

4.6 Quantification of the Residues Interacting with the Surface

In order to characterize which amino acids of a protein are interacting with an adsorbent surface when the protein approaches an adsorbent surface, it is required to rotate the protein along the three principal axes of the local coordinate system of the protein. By assuming the surface to be of homogenous chemical composition,

microscopically flat and devoid of any defects, translation along the axes normal to the surface would be sufficient to map out the residues interacting with the interface.

In the current implementation, the interface between the protein and a material surface is considered to be parallel to the xy plane with the outer normal of the surface being along the +Z-axis as depicted in the Fig 4.9. In this orientation, and assuming that the protein is resting on the surface, the solvent accessible amino acid residues that are considered to be close enough to the surface are expected to energetically contribute to the adsorption of the protein in a non-negligible manner and are characterized by identifying the residues that are positioned within a user-defined height 'D' from the interface. In the current study, the chemical characteristics of each type of amino acid residue are classified according to the Table 4.7.

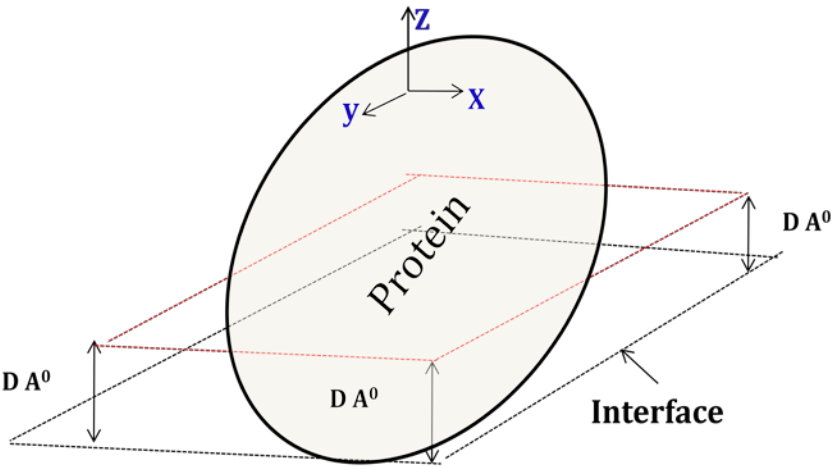


Figure 4.9: The protein is modeled as resting on the surface (outlined by blue lines) with the coordinates of the bounding box being represented by XYZ. The plane (Red) at

height ‘D’ from the surface is used as the cut-off for identifying the amino acids expected to be close enough to the surface to interact with the surface.

Table 4.7: Classification of amino acids

Classification	Amino acids
Main group	
Amphiphillic neutral	Glycine
Hydrophobic neutral	Alanine, Isoleucine, Leucine, Methionine, Phenylalanine, Proline, Tryptophan, Valine
Hydrophilic neutral	Asparagine, Cysteine, Glutamine, Serine, Threonine, Tyrosine
Hydrogen bondable	Asparagine, Cysteine, Glutamine, Serine, Threonine, Tyrosine, Arginine , Aspartic acid, Glutamic acid, Lysine, Histidine
Charged	Arginine , Aspartic acid, Glutamic acid , Lysine
Positively charged	Arginine , Lysine
Negatively charged	Aspartic acid, Glutamic acid
Neutral/positive charged	Histidine
More positively charged	(Arginine + Lysine) – (Aspartic acid, Glutamic acid)
More negatively charged	(Aspartic acid, Glutamic acid) - (Arginine + Lysine)

4.7 Modeling the Adsorbent Material Surface

The adsorbent material surface is modeled as a structureless, homogenous, flat plane without any charge or force parameters attributed that could contribute to the adsorption of the protein. In the current implementation for predicting the orientation of the protein on a material surface, the surface is thus modeled only as a visualizing aid to enable the user to view the orientation in which the protein would be adsorbed.

4.8 Visualization of Residue Distribution

The distributions of the residues for a given protein model are visualized using the contour plot function available within MATLAB®. Once a configurational space search was completed, residue distribution of any given orientation is obtained by the linear interpolation of the residue distribution corresponding to the nearest available data point

A topography map is generated for different proteins ranging from small molecular weight proteins (i.e., greater than 10 residues) to multi-chain and moderately high molecular weight proteins (i.e., less than 3000 residues), with computation time usually averaging in minutes per protein model. Parameters such as the (a) angle of rotation, (b) effect of local positioning and (c) exclusion of buried residues were found to affect the topography map significantly.

4.9 Application to a Protein Adsorption System

As a demonstration of the developed capabilities, a topography map was generated for a selected protein and used to predict the preferred initial orientations of the

protein on different types of surfaces. The orientation angle (θ, Φ) and the number of residues within the defined interaction distance (i.e., D; see Fig. 4.9) corresponding to each of these angles were used to identify the preferred orientation of a given protein model on surface models with different surface characteristics, which were characterized as hydrophobic, hydrophilic, positively charged, or negatively charged.

CHAPTER FIVE

RESULTS AND DISCUSSION

Different parameters such as the interaction height (D), residue exclusion, and angle of rotation were observed to affect the predicted orientation of the protein as these factors influence the topography map. Effects of each of these parameters are indicated in the attached appendices. Based on these results, the following parameters were set for predicting the orientation of protein on interface.

Table 5.1: Settings for predicting the orientation of protein

Parameter	**Setting**
Angle of rotation (α)	5º
Interaction height (D)	10 Å
Residue exclusion • Buried	Relative accessibility < 10%
Solvent environment • Solvent probe radii (r_w) • Penetration depth (s)	 1.4 Å 2.5 Å

Analyses have been performed for both large (> 100 kDa) and small (< 20 kDa) proteins with the total runtime (i.e., iterations required for rotating the protein and analyzing the amino acid composition over the surface) typically taking only a few minutes per protein. The results from the current approach have been tested with different

protein-surface systems on a variety of surfaces including those which are both biological and synthetic in origin.

5.1 Prediction of Orientation on Homogenous Surface

5.1.1 Prediction of orientation of proteins on uncharged homogenous surfaces

The adsorption process of proteins to hydrophilic or hydrophobic surfaces is considered to involve the displacement of the tightly or loosely bound layer of water interacting with the surface and the displacement of this water has been observed to be more on a hydrophobic surface than on a hydrophilic surface due to (a) lower number of hydrogen bonds between a hydrophobic surface and water and (b) resulting increase in free energy that occurs when water over a hydrophobic surface is displaced to bulk solution.[4] These effects are most prominent in polymers, and in particular uncharged polymers with nonpolar functional groups.[74]

CH_3-SAMs and OH-SAMs are considered as ideal model surfaces for mimicking the behavior of hydrophobic and hydrophilic surfaces, respectively. Simulation results have provided estimates of the potential mean force (PMF) between a peptide and these types of surface and have shown that these interactions primarily only acts over a range of about 10 Å of surface separation distance (SSD) for these types of surfaces under physiological solution conditions.[74] A cut-off value for 'D' of 10 Å is therefore used to approximate the interaction distance over which the amino acids on the solvent accessible

surface of a protein are close enough to be influenced by the functional groups of a material surface. A summation of the PMF contributions for the interactions between these amino acids and the surface for a given orientation of the protein provides a means to estimate the overall change in free energy for adsorption and to predict the preferred orientations of a protein on hydrophilic and hydrophobic SAMs.

5.1.1.1 Prediction of orientation on a hydrophilic surface

An adsorption process involving an OH-SAM and a protein is considered to be mediated through hydrogen bonds involving the hydrogen bondable groups between the SAM surface and the protein surface. Hence the prediction of protein orientation on this type of hydrophilic surface was based on the total set of hydrogen bondable residues as per the classification given in table 4.7.

5.1.1.1.1 Human serum albumin (HSA, PDB Model: 1E78)

HSA is the most abundant protein in the human blood plasma and has been recently been observed to mediate platelet attachment beyond a critical degree of adsorption-induced unfolding.[74] Hence the orientation in which these protein adsorb is of particular interest in view of biomaterial design. For the current PDB model, the 'A' chain of HSA has been chosen for analysis.

The topographical map of hydrophilic neutral residues in HSA suggests a unique site ($\Phi = 95^o$, $\theta = 231^o$) within the HSA where the maximum number of residues is seen to interact with the surface (Fig. 5.1) while the site ($\Phi = 30^o$, $\theta = 330^o$) corresponds to

maximum number of hydrogen bondable residues are positioned in the protein (Fig.5.2) with the orientation being that of 'back-on' and 'Side AB-on' respectively(Fig. 5.3), which was found to be in agreement to the predicted orientations of this adsorbed protein as reported in literature.[37] The residue distribution for the aforementioned orientations on contacting the hydrophilic surface have been depicted in Fig 5.4.

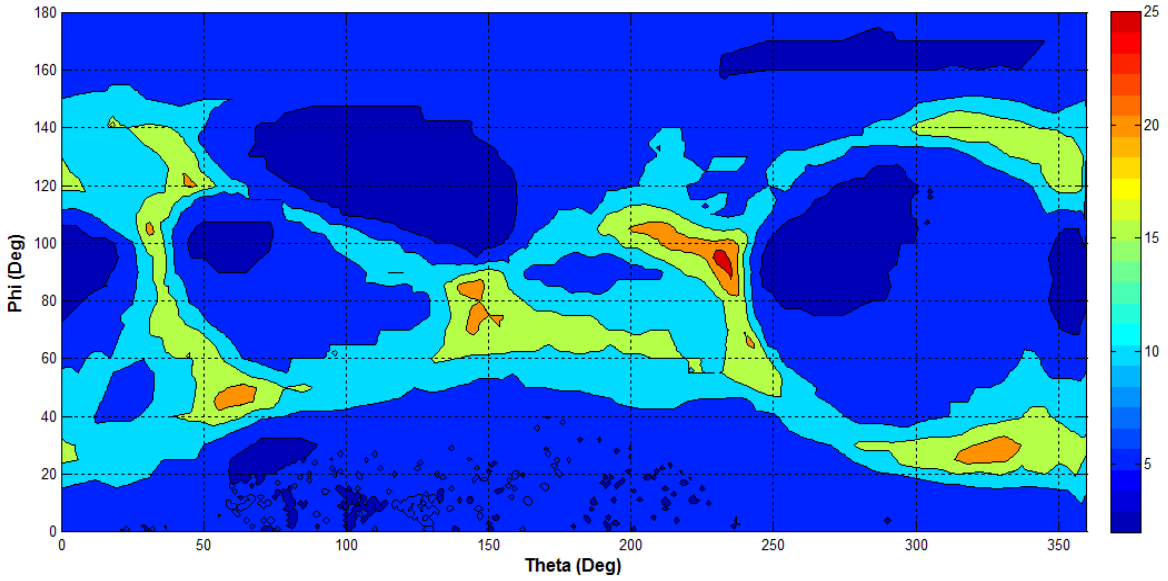


Figure 5.1: Distribution of hydrophilic neutral residues in HSA (1E78). The dark red colored region corresponds to maximum residue concentration for which the orientation angles are $\Phi = 95^{o}$, $\theta = 231^{o}$.

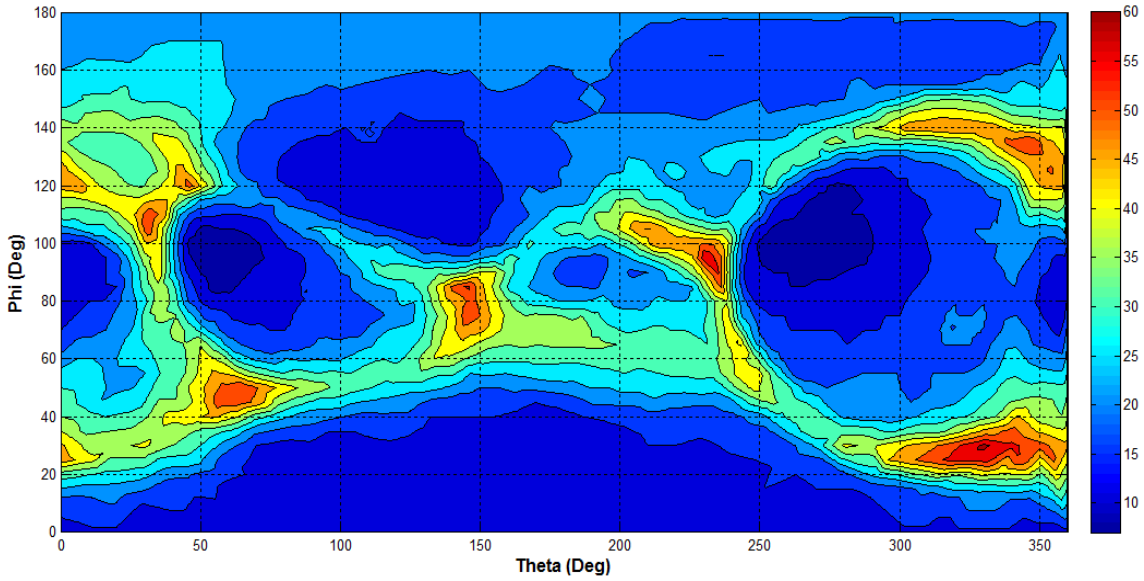


Figure 5.2: Distribution of hydrogen bondable residues in HSA (1E78). The dark red region corresponds to maximum residue concentration for which the orientation angles are Φ = 30º, θ = 330º.

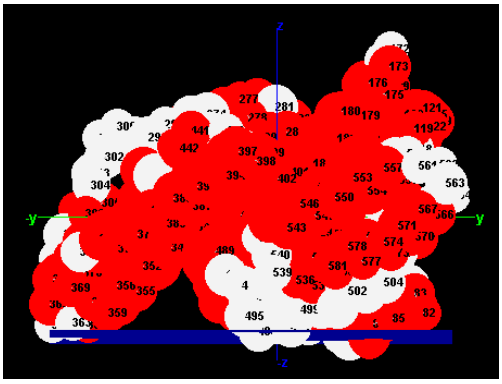
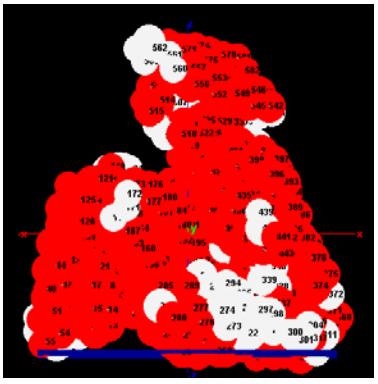
(a)  (b) 

Figure 5.3: Preferred orientation of HSA (1E78) on hydrophilic surface for the orientation angle where the maximum concentration of (a) Hydrophilic neutral residues (Φ = 95º, θ = 231º) and (b) hydrogen bondable residues (Φ = 30º, θ = 330º) are observed. Dark blue indicates the adsorption surface plane. Residues are colored according to

secondary structure. Red- helices; white-loops; Olive green-beta sheets. Numbers correspond to the primary amino acid sequence.

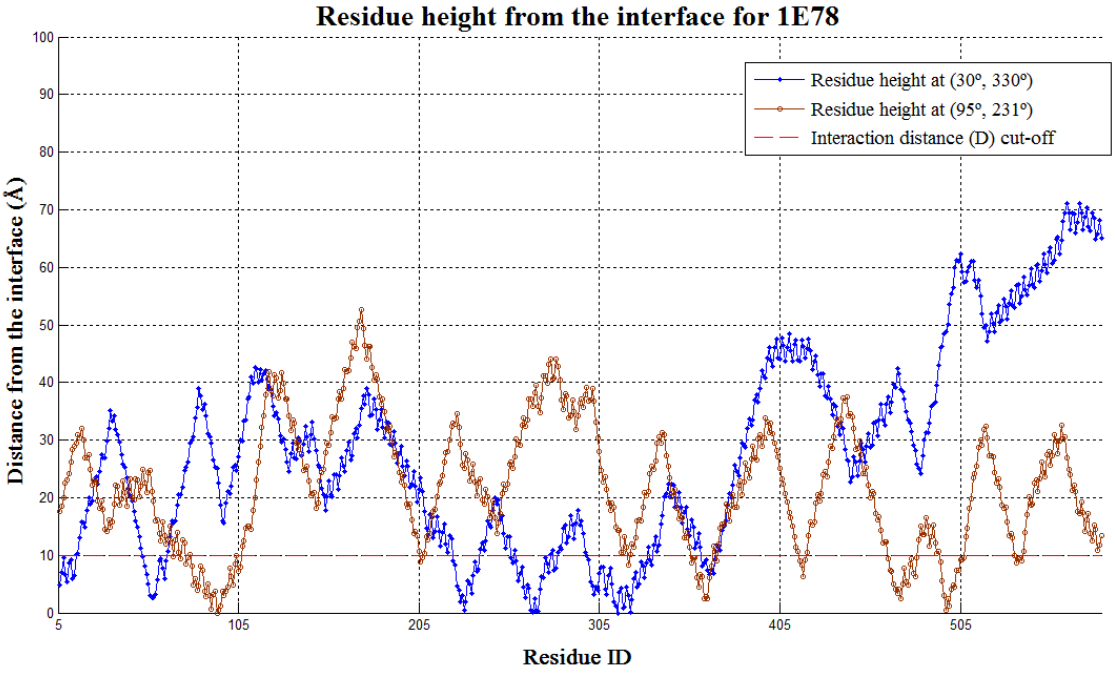


Figure 5.4: Residue height from the surface plane for HSA (1E78) corresponding to orientation angles (Φ, θ) of (95º, 231º) and (30º, 330º).

5.1.1.1.2 Hen egg white lysozyme (HEWL, PDB Model: 7LYZ)

HEWL is an enzyme that catalyzes the hydrolysis of the glycosidic bond between N-acetylglucosamine and N-acetylmuramic acid in bacterial cell wall through the catalytic sites Glu35 and Asp 52.

The topographical map of hydrophilic neutral residues suggests multiple θ, Φ sites of (120º, 162º), (120º, 167º) and (120º, 173º) within the HEWL where the maximum number of residues is seen to interact with the interface (Fig. 5.5). In the (120º, 162º) orientation, the protein is found to be adsorbed with its long axis being parallel to the surface and with its catalytic site face being oriented upwards towards the solution (Fig. 5.6 and 5.7).

Alternatively, the maximum number of hydrogen bondable residues is highest at orientation angle corresponding to Φ, θ site (1ÞÞ 229º) (Fig.5.8), at which the orientation is predicted to be that of side-on with the active site being partially stearically hindered by the surface plane (Fig 5.9 and Fig. 5.10).

The adsorptions of lysozyme on different substrates has been studied by simulation[21, 72] and experimental techniques[14, 75], and have indicated the preferred orientation of lysozyme on hydrophilic surfaces to be side-on, especially at low bulk concentrations, thereby suggesting the preferred orientation could correspond to the maximum concentration of hydrogen bondable residues and not the hydrophilic neutral residues.

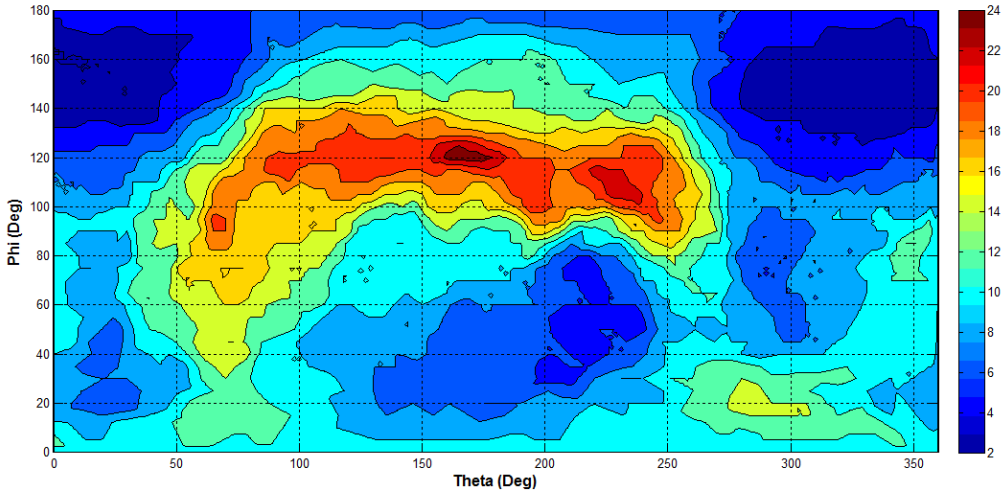


Figure 5.5: Distribution of hydrophilic neutral residues in HEWL (7LYZ). The dark red circled region corresponds to maximum residue concentration for which the orientation angle are Φ = 120º, θ = 161º.7´.

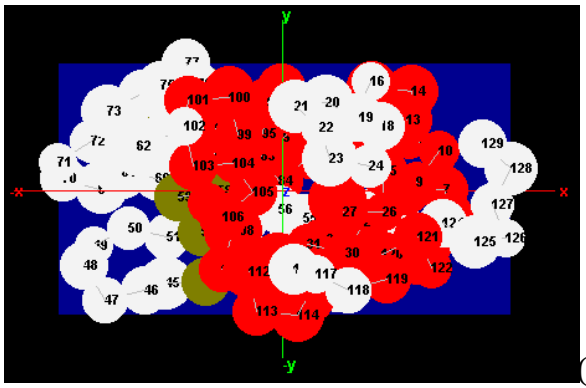
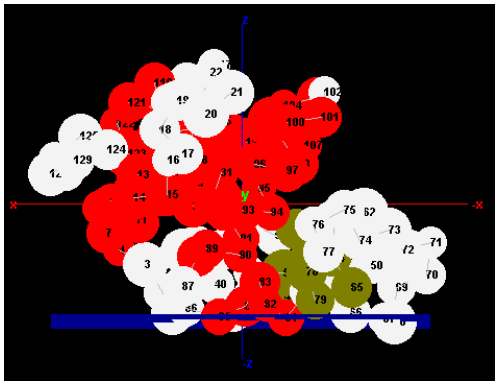
(a)  (b) 

Figure 5.6: Preferred orientation of HEWL (7LYZ) on hydrophilic surface for the orientation angle where the maximum concentrations of hydrophilic neutral residues are observed. (a) Side and (b) Top view corresponding to Φ = 120º, θ = 162º. Dark blue indicated the adsorption surface plane. Residues are colored according to secondary structure. Red- helices; White-loops; Olive green-beta sheets.

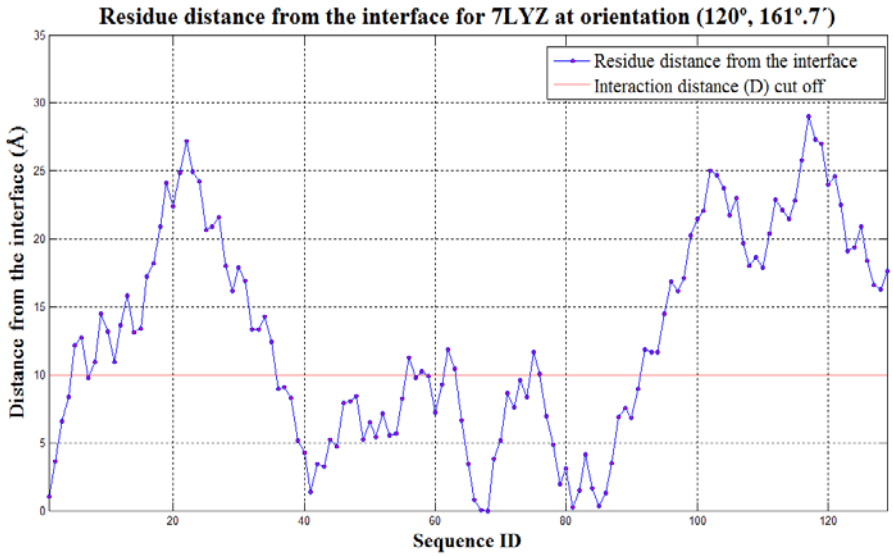


Figure 5.7: Residue height from the interface for HEWL (7LYZ) corresponding to corresponding to $\Phi = 120^{o}, \theta = 161^{o}.7'$. The residue contacting the surface is R68. The bioactive site is made up of residues E35 and D52.

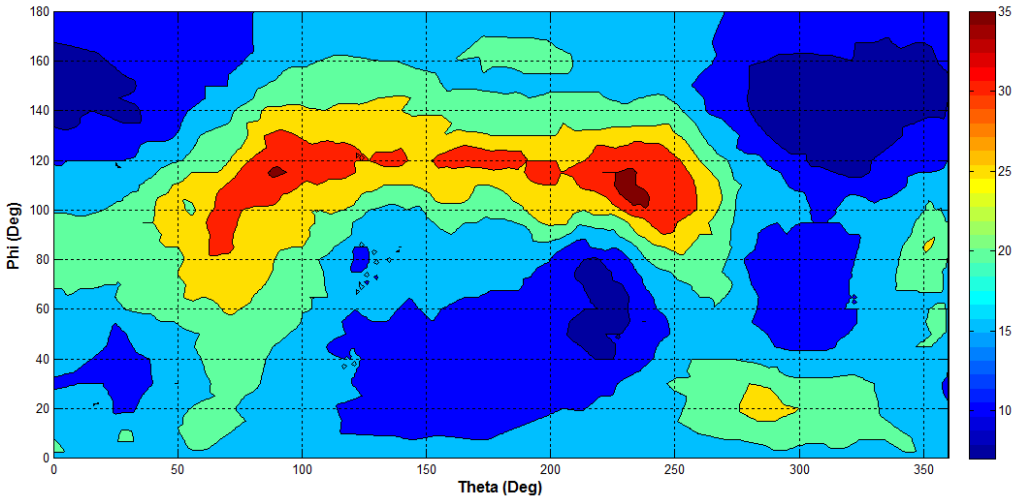


Figure 5.8: Distribution of hydrogen bondable residues in HEWL (7LYZ). The dark red circled region corresponds to maximum residue concentration of 38 for which the orientation angle is $\Phi = 110^{o}$, $\theta = 228^{o}.8'$.

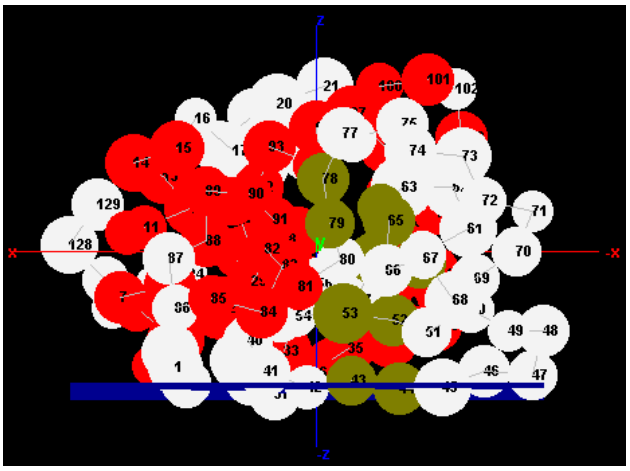
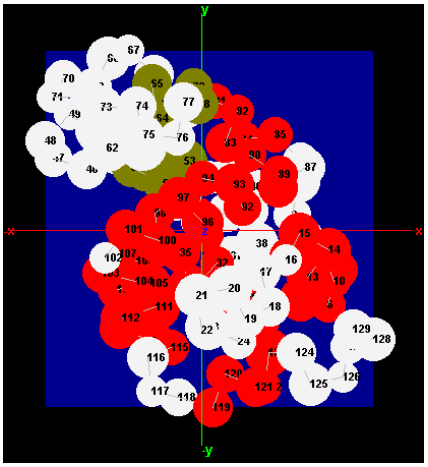
(a)  (b) 

Figure 5.9: Preferred orientation of HEWL (7LYZ) on hydrophilic surface for the orientation angle where the maximum concentrations of hydrogen bondable residues are observed. (a) Side and (b) Top view corresponding to Φ = 110º, θ = 228º.8´. Dark blue indicated the adsorption surface plane. Residues are colored according to secondary structure. Red- helices; White-loops; Olive green-beta sheets

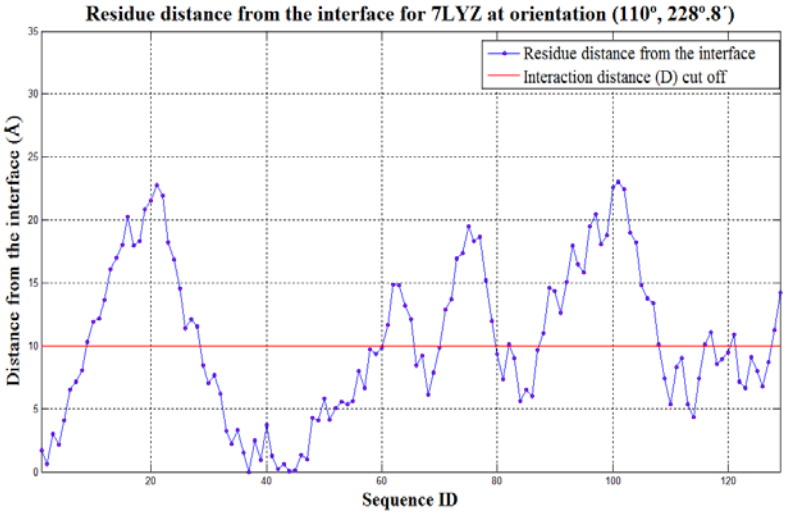


Figure 5.10: Residue height from the interface for HEWL (7LYZ) corresponding to corresponding to Φ = 110º, θ = 228º.8´. The residue contacting the surface is F37 and the bioactive site is made up of residues E35 and D52.

5.1.1.2 Prediction of protein orientation on a hydrophobic surface

5.1.1.2.1 Hen egg white lysozyme (HEWL, PDB Model: 7LYZ)

The topographical map of hydrophobic neutral residues suggest a unique site at θ, Φ (110º, 254º) within the HEWL where the maximum number of residues is seen to interact with the interface (Fig. 5.11), with the protein oriented in a sideways orientation (Fig 5.12, Fig 5.13). The preferential adsorption of lysozyme in a side-way orientation has been reported by experimental and simulation techniques.[73, 77, 78] The N-C termini along with several of the key residues such as Lys1, Arg5, Lys13, Lys33, Arg14 and Arg128, are involved in the HEWL adsorption to the surface at this orientation (Fig 5.13), which corresponds to the reported sites of close contact with the hydrophobic surface.[73]

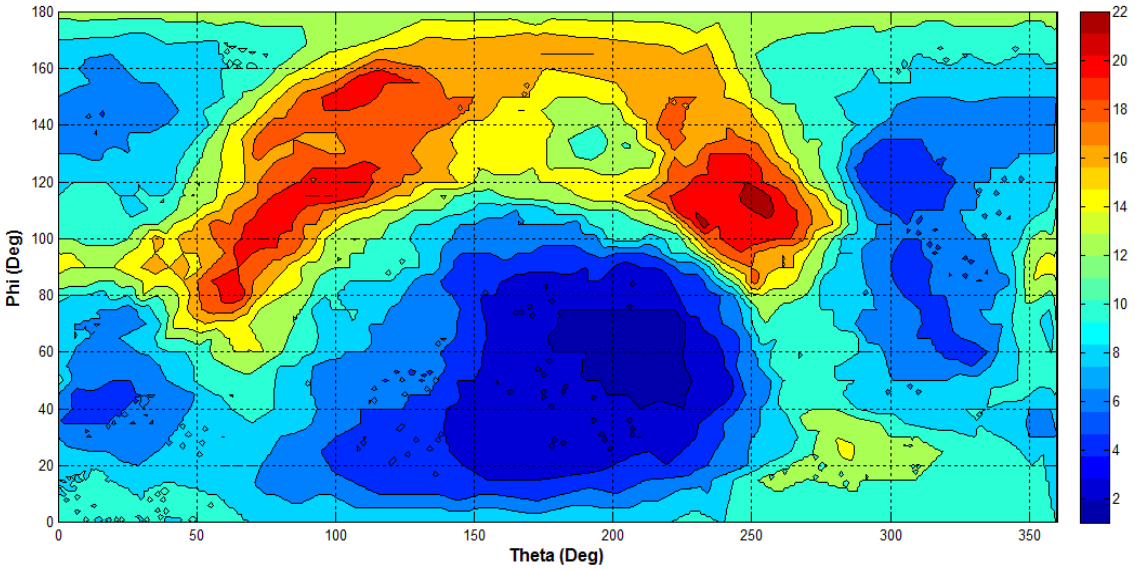


Figure 5.11: Distribution of hydrophobic neutral residues in HEWL (7LYZ). The dark red circled region corresponds to maximum residue concentration for which the orientation angle are Φ = 110º, θ = 254º.

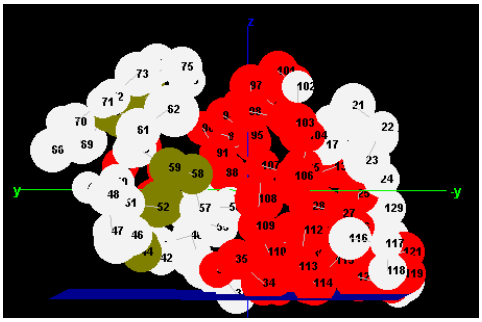
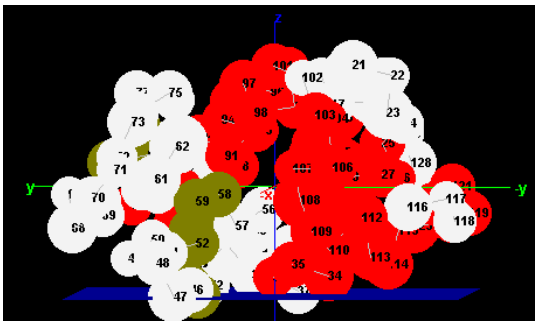
(a)  (b) 

Figure 5.12: Preferred orientation of HEWL (7LYZ) on hydrophobic surface for the orientation angles (Φ, θ) where the maximum number of hydrophobic residues (23) was observed: (a) (110º, 254º), (b) (105º, 233º). Dark blue indicates surface plane. Residues are colored according to secondary structure. Red- helices; White-loops; Olive green-beta sheets.

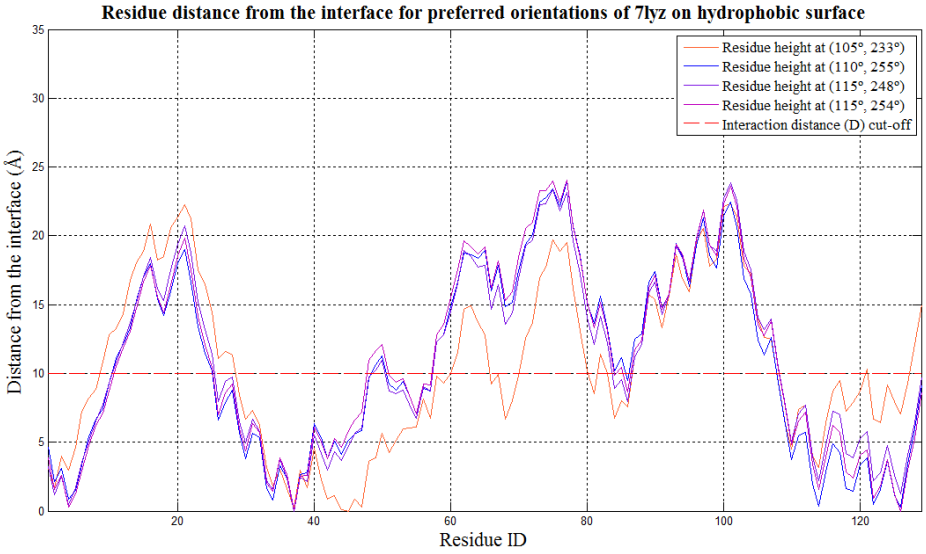


Figure 5.13: Residue height from the interface for HEWL (7LYZ) corresponding to Φ, θ (110º, 255º). The residue contacting the surface is F37 in most of the orientations except (105º, 233º) in which the residue contacting the surface is R45. The bioactive site is made up of residues E35 and D52.

5.1.1.2.2 Mitochondrial cytochrome c (CYTC, PDB Model: 1HRC)

CYTC is a peripheral membrane protein with a covalently attached heme as a redox-active cofactor and finds many applications in the field of protein chromatography, drug delivery on solid substrates, biosensors, biofuel cells and bioelectronic devices. The primary application of these proteins however has been limited by the orientation in which these proteins adsorb on interface, which in turn affect the electron transport process and the performance of the intended application.[38]

The topographical map of hydrophobic neutral residues in CYTC suggests four major sites where the numbers of interacting residues are high. These sites correspond to Φ, θ sites of (80º, 127º), (100º, 315º), (100º, 320º) and (130º, 235º) as depicted in Fig 5.14.

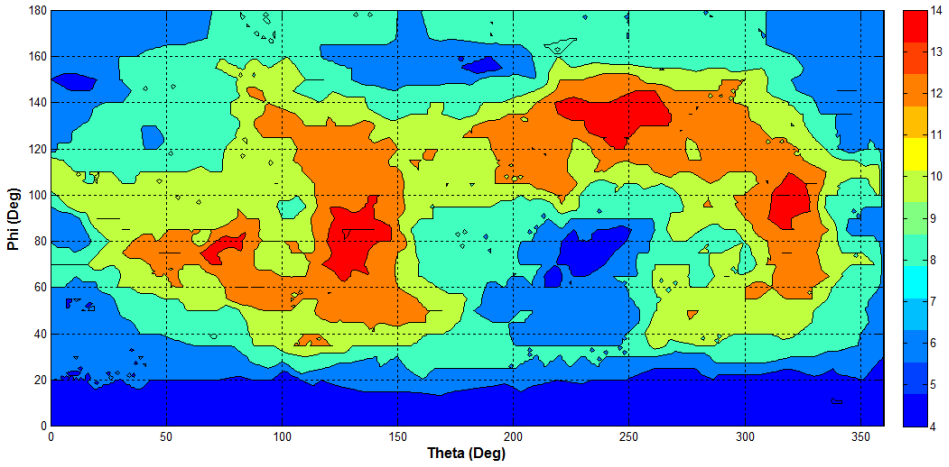


Figure 5.14: Distribution of hydrophobic neutral residues in mitochondrial cytochrome C (1HRC). The red region corresponds to maximum hydrophobic residue concentration of

16 for which the orientation angles (Φ, θ) are (80º, 127º), (100º, 315º), (100º, 320º) and (130º, 235º).

The residue height corresponding to these orientations is depicted in Fig 5.15.

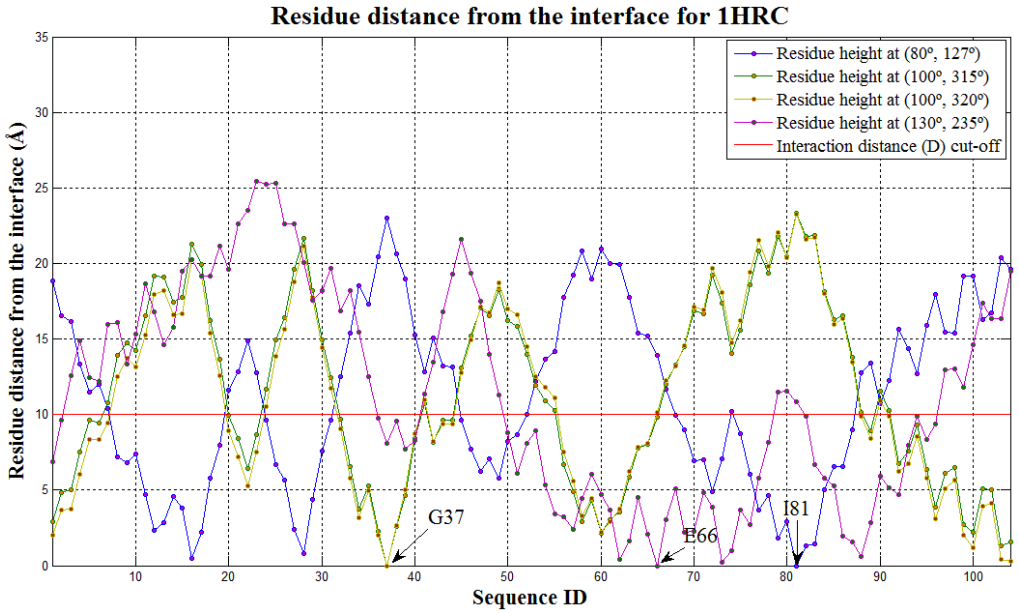
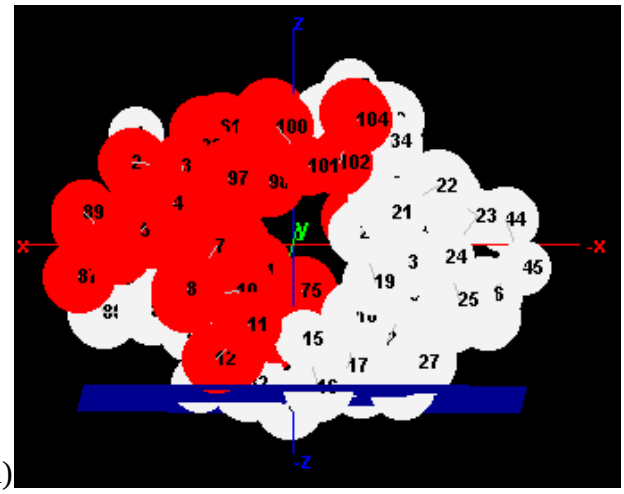


Figure 5.15: Residue height from the interface for mitochondrial cytochrome C (1HRC) for preferred orientations on a hydrophobic surface. The residues interacting with the interface to each of the orientation angle are labeled in the above. Acyl region of the cardiolipin is hypothesized to dock with the membrane through the hydrophobic channel comprising of residues 67-71 and 82-85.[89]

Based on computational and simulation studies reported in the literature, the orientation corresponding to Φ, θ (80º, 127º) is the favored orientation with which the

CYTC is considered to interact with CH_3-SAM (Fig 5.16) with Ile81 being the amino acid closest to the surface.[79] The predicted orientation also corresponds to the orientation predicted in a cell membrane by the OPM database.[80, 81] The other two orientations (Fig 5.17) with Gly37 and Glu66 as the interacting amino acid have not been reported in previous studies in the literature.

(a)



(b)

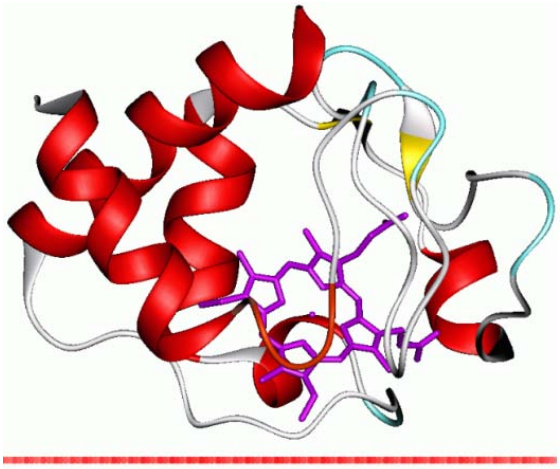
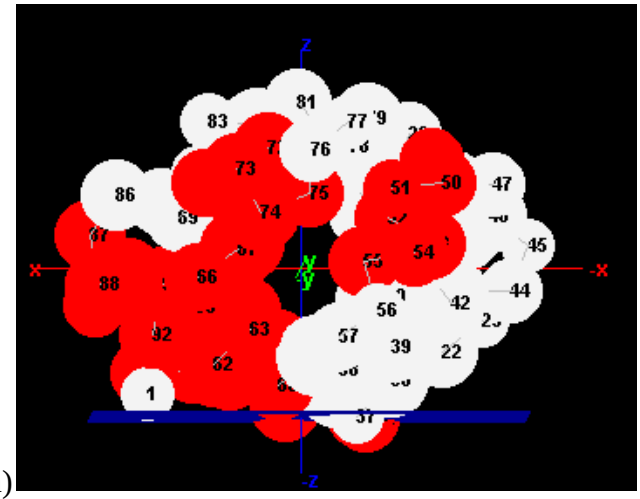


Figure 5.16: Preferred orientation of CYTC (1HRC) on a hydrophobic surface for the orientation angle Φ, θ (80º, 127º). (a) Orientation prediction by current tool. Blue plane

indicates adsorbent surface. (b) Orientation prediction by the OPM database[80]. Red horizontal line indicates adsorbent surface. Residues are colored according to secondary structure. Red- helices; White-loops; Olive green-beta sheets

(a)



(b)

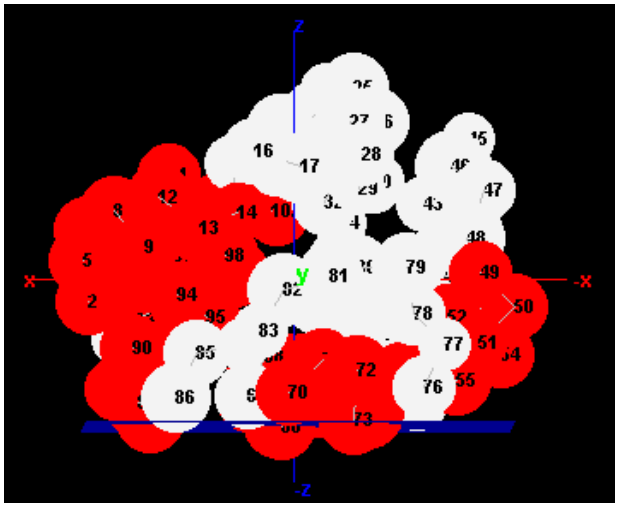


Figure 5.17: Orientation of CYTC (1HRC) on hydrophobic surface for the Φ, θ orientation angle set (a) (100º, 315º), (100º, 320º) (b) (130º, 235º). Blue plane indicates adsorbent surface. Residues are colored according to secondary structure. Red- helices; White-loops; Olive green-beta sheets

5.1.2 Prediction of orientation of proteins on charged homogenous surfaces

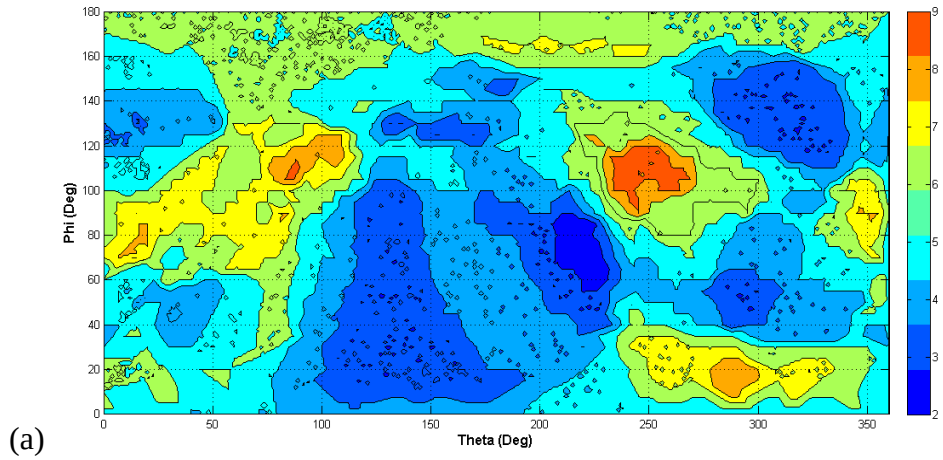
The adsorption processes of proteins to negatively or positively charged surfaces are primarily mediated by electrostatic interactions. These types of interactions are most prominent on metallic and ceramic surfaces.[39, 72] Electrostatic interactions mediating the adsorption of protein to interfaces are complex and are seen to be strongly influenced by the concentration of ionic species in solution.[35, 39, 72] Electrostatic interactions of charged residues with oppositely charged surface groups tend to be shielded by the ions at ionic concentrations above 0.5M. NH_2-SAMs and COOH-SAMs are considered as ideal model surfaces for replicating the behavior of positively and negatively charged surfaces, respectively. Simulation results have indicated that the potential mean force (PMF) of interactions between charged amino acid residues and surface groups generally also occur within a surface separation distance (SSD) of approximately 10 Å in a physiological saline environment on these model surfaces.[73] Hence the 'D' cut-off at 10 Å can be considered to provide an appropriate cutoff distance to be used to approximate the orientation of the proteins on negatively and positively charged SAMs.

5.1.2.1 Prediction of orientation on a negatively charged surface

The orientation of protein on a negatively charged surface can be predicted by the topographic mapping of the positively charged residues on the protein surface. For the determination of the most probable site of adsorption, a topographic map of the 'More positive regions' were performed to also take into account the counteracting repulsive effect caused by the negatively charged residues with the negatively charged interface.

5.1.2.1.1 Hen egg white lysozyme (HEWL, PDB Model: 7LYZ)

The topographical map of positive residues in HEWL (Fig 5.18) indicates side-on orientation to be preferred (Fig 5.19) with the orientation angle corresponding to Φ, θ (110º, 255º) and the residue distribution at the interface being similar to that observed for the hydrophobic interface (Fig 5.20). The predicted orientation is also in line with the reported experimental and simulation result on SAMs and ceramic surfaces such as silica and mica.[14, 21, 73, 82]



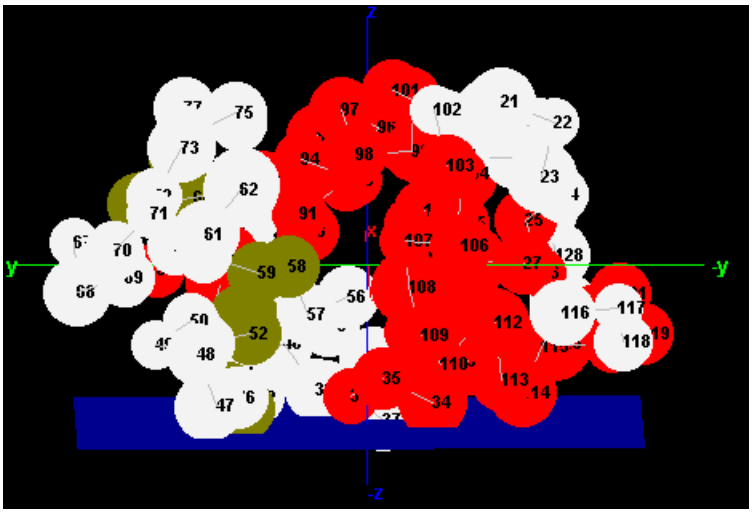
(a)

(b)

Figure 5.18: Distribution of positively charged residues in HEWL (7LYZ). (a) The dark orange region corresponds to the positive residue concentration of 10 for which the

orientation angle Φ, θ is (110º, 255º). (b) The dark red region corresponds to the 'More positive regions' of residue concentration equivalent to 7 for which the orientation angle Φ, θ is (10º, 288º). Combining the analysis from graphs (a) and (b), topography site (110º, 255º).is the most ideal for HEWL adsorption to negatively charged surfaces

(a)



(b)

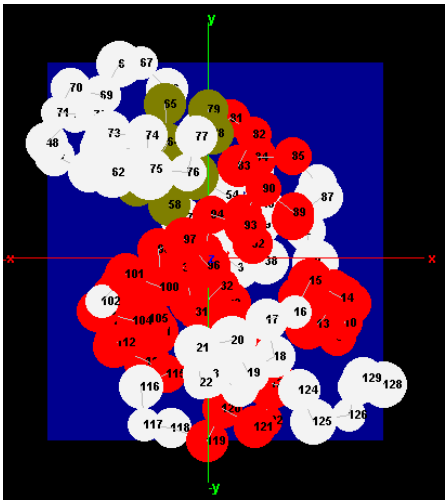


Figure 5.19: Orientation of HEWL (7LYZ) on negatively charged surface for the orientation angle Φ, θ (110º, 255º) (a) Side view (b) Top view. Residues are colored

according to secondary structure. Dark blue indicates adsorbent surface plane. Red-helices; White-loops; Olive-green-beta sheets.

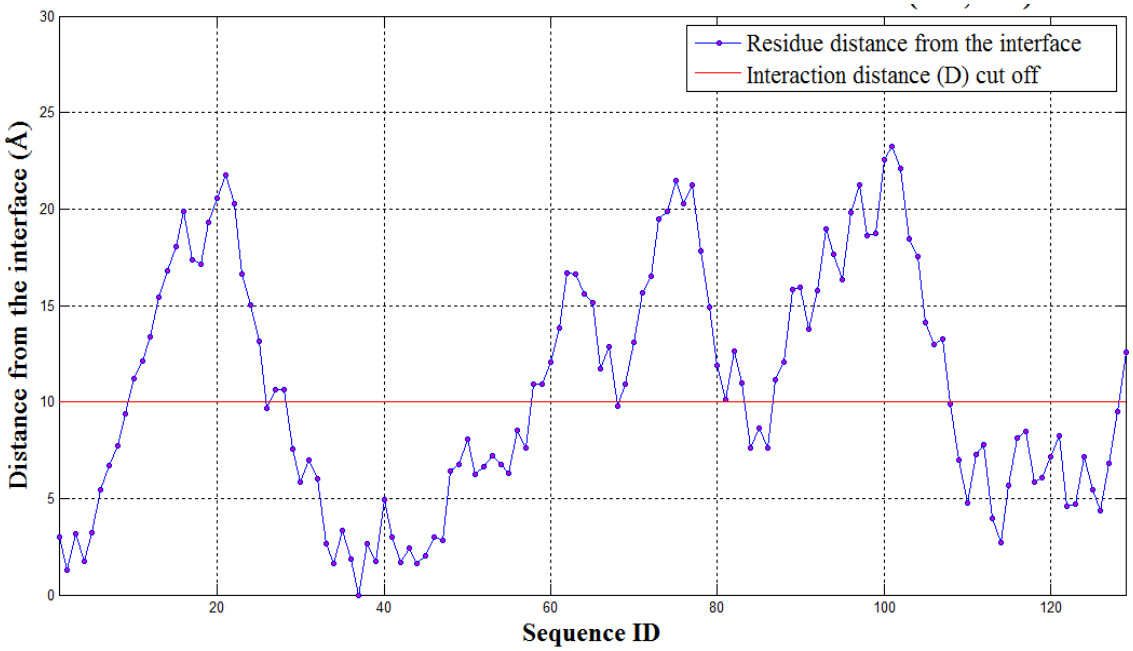


Figure 5.20: Residue height from the interface for HEWL (7LYZ) for preferred orientations on a negative surface. The bioactive site is made up of residues E35 and D52.

5.1.2.1.2 Mitochondrial cytochrome c (CYTC, PDB Model: 1HRC)

The orientation of CYTC could be best described based on the orientation of the heme-cofactor (Fig 5.21). For the best performance of adsorption systems utilizing CYTC, it is desired that fast transport of electron transport be achieved by orienting the heme-cofactor near and perpendicular (vertical) to the contacting interface (Fig 5.21b).

Based on the understanding gained through several published studies, negative substrates are widely considered to be the ideal surfaces for immobilizing CYTC. Hence it would be interesting to use the developed program to identify the predicted preferred orientation of CYTC on a negative substrate for comparison with these published results.

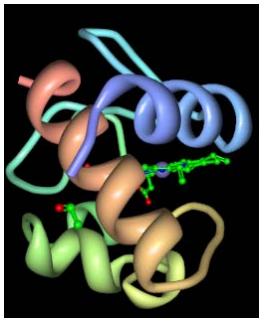
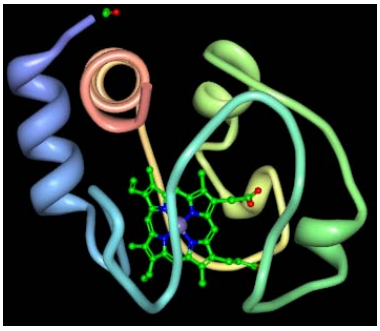
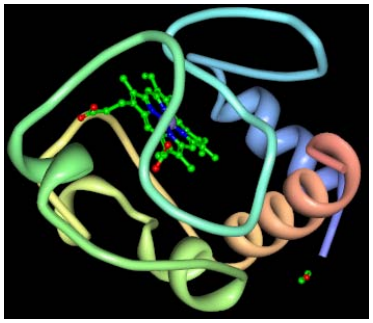
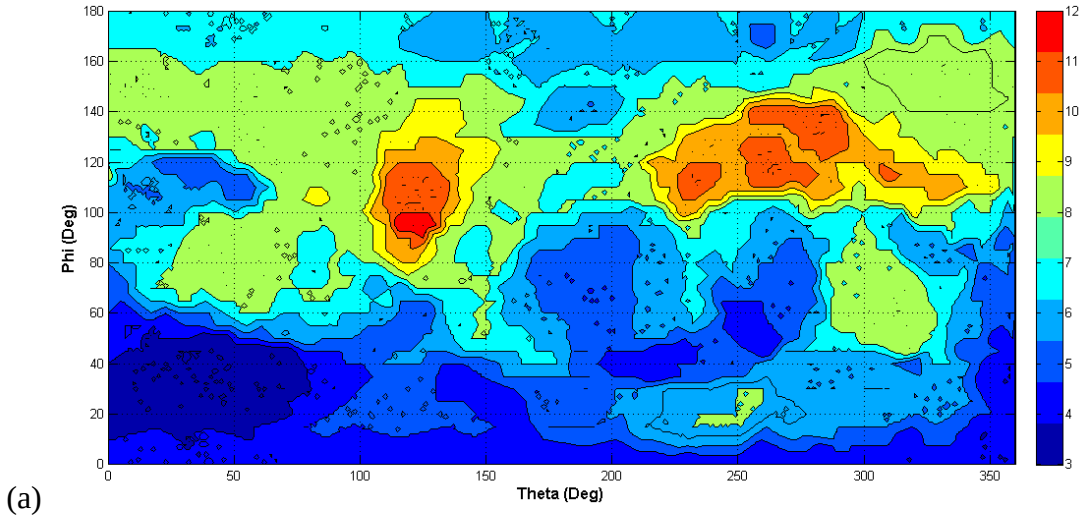
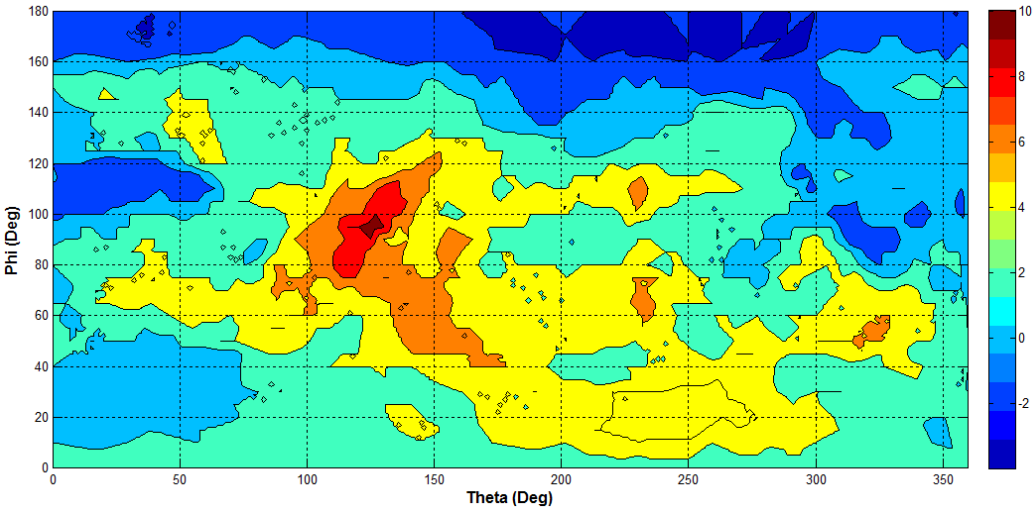
(a)  (b)  (c) 

Figure 5.21: Orientations of the heme group for CYTC adsorbed on a surface. (a) Horizontal (b) Vertical and (c) Inclined. 3D model generated in RCSB-SimpleViewer.

The topographical map of positive residues in CYTC suggests multiple sites with high concentration of positively charged residues (Fig. 5.22) for which the orientation angles are tabulated in Table 5.1. Combing the analysis from Table 5.1 and Figure 5.22 vertical orientations[38, 78] were preferred more than the horizontal ones[35]. Most importantly, the orientations of the heme group were were mostly near the interface in vertical orientations while the majority of the horizontal orientations were further away from the interface, thus supporting that the vertical orientations would provide the best conditions for electron transfer between the protein and the surface.



(a)



(b)

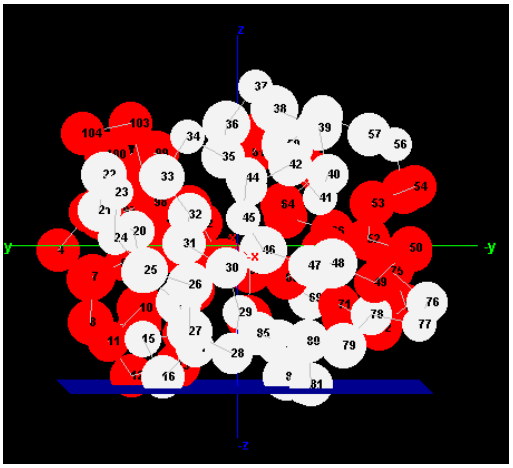
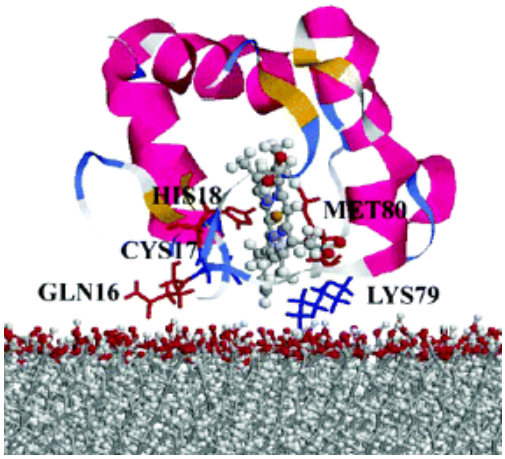
Figure 5.22: Distribution of positive residues in CYTC (1HRC). (a) The dark orange regions correspond to the positive residue concentration of 12 for which the orientation angle pairs (Φ, θ) are listed in Table 5.1. (b) The dark red region corresponds to the maximum residue concentration of 11 for 'More positive region' of orientation angles with the cofactor orientation corresponding to vertical as listed in Table 5.1.

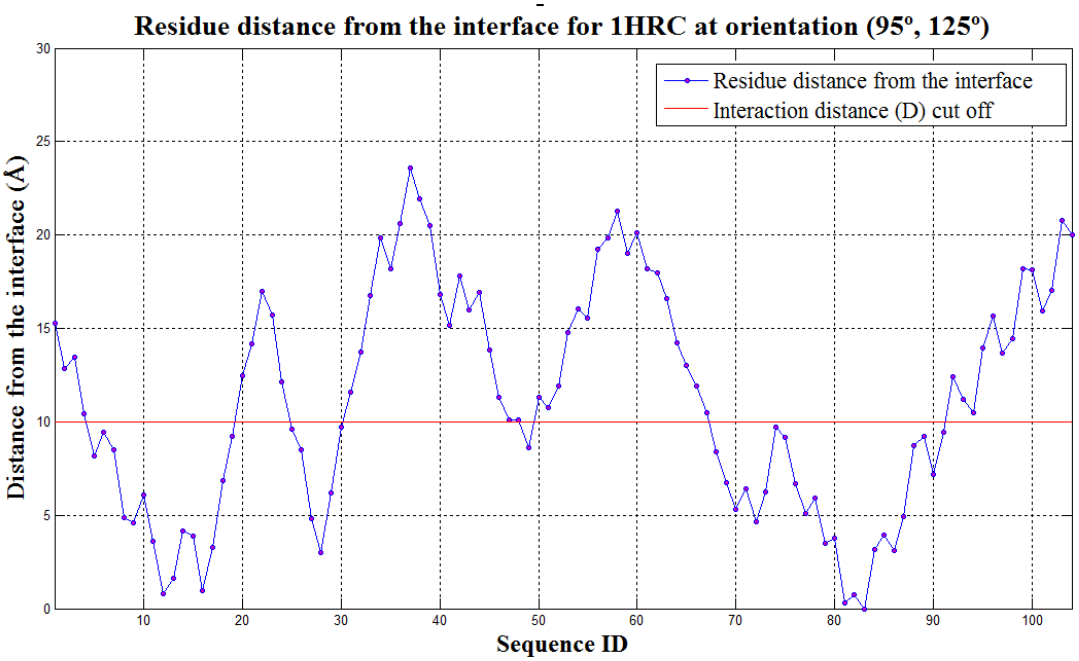
Table 5.2: Orientation angles corresponding to maximum positively charged residue concentrations on a negative interface. The residues with the strongest adsorption is highlighted in yellow

Phi(degrees)	Theta(degrees)	Interacting residue	Cofactor orientation
90	125	Ile 81	Vertical
95	115	Gln 12	Vertical
	120	Gln 12, Ala 83	
	125	Ala 83	
110	112	Gln 12	Vertical
	117	Gln 12	Vertical
	122	Ala 83	Vertical
	128	Ala 83	Vertical
	133	Lys 73	Horizontal
	229	Lys 73	Horizontal
120	260	Glu 62	Inclined
	266	Glu 62	

130	281	Glu 62	Inclined
	287	Glu 62	
140	257	Glu 62, Glu 88	Inclined
	265	Glu 62	Inclined
	272	Glu 62	Horizontal
	280	Glu 62	Horizontal
	288	Glu 62	Horizontal

The orientation and the residue distribution at the interface for CYTC on a COOH –SAM (Fig 5.23) corresponding to the orientation angle of Φ, θ (95º, 125º) is well corroborated with that reported in literature.[38, 78]

(a)  (b) 



(c)

Figure 5.23: Orientation of CYTC (1HRC) on negatively charged SAM for the orientation angle Φ, θ (95º, 125º) (a) Prediction by the current study. Dark blue indicates adsorbent surface plane. Residues are colored according to secondary structure. Red-helices; White-loops; Olive-green-beta sheets. (b) Prediction by molecular dynamic simulation. (c) Residue height from the interface for CYTC (1HRC) for orientation angle (95 º, 125 º) on a negative surface. The residues Cys14, Ala15, and His18 are covalently linked to the heme group and in the current orientation are adsorbed down.

5.1.2.1.3 Immunoglobulin G (IgG) antibody

IgG is the most abundant type of immunoglobulin found in humans, which are synthesized and secreted by the plasma B cells. These large protein molecules exist as

dimers composed of a large heavy and a small light chain with each monomer having an antigen binding site. On the basis of the distribution of IgG in serum, IgG molecules are further classified into four other subclasses such as IgG1, IgG2, IgG3 and IgG4, with IgG1 being the most prevalent and IgG4 being the least abundant type of IgG antibody in serum.[82]

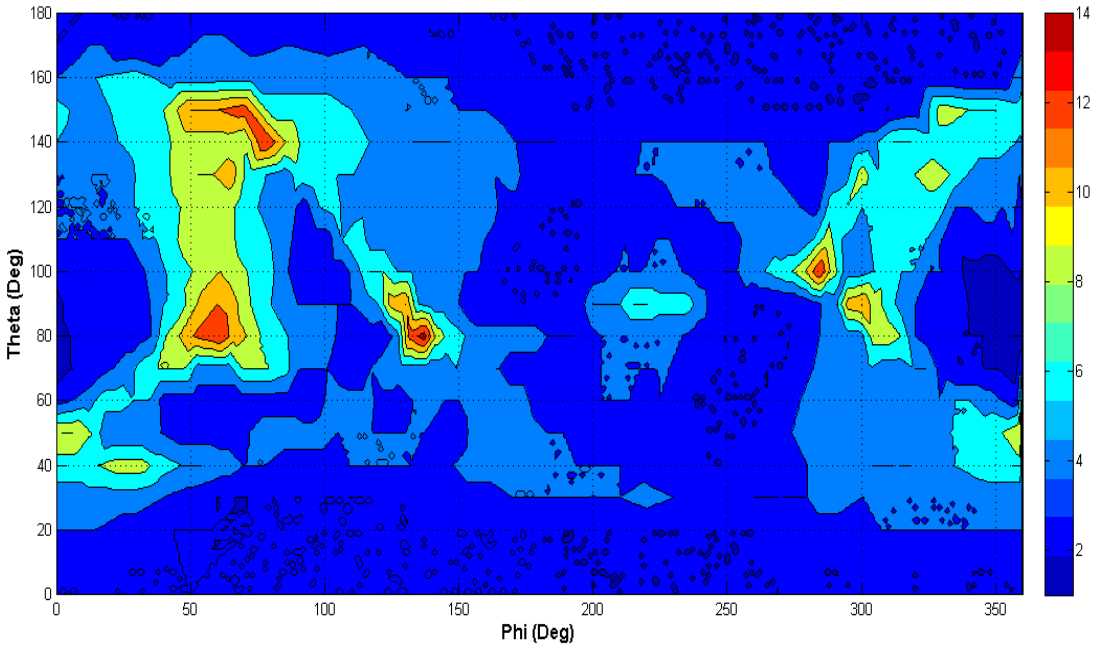
Since these proteins have broad range specificity against various pathogens and toxins, these molecules are used in many rapid assays and detection kits. However, the substrates on which these protein molecules are immobilized affect the orientation and the final bioactivity of the antibody[83]. The orientation of IgG1 and IgG2 on charged surfaces have been reported and have shown a preferred orientation at low ionic solution concentration and surface charge density[84, 85]. In order to verify the existence of sites corresponding to maximum positive residue concentrations, topography maps for positively charged residues in IgG1 (PDB Model: 1IGY) and IgG2 (PDB Model: 1IGT) were generated

5.1.2.1.3.1 IgG1 (PDB Model: 1IGY)

This PDB model is constituted of four chains namely the A chain, B chain, C chain and D chain. The A and C chains are each 213 residues long and form the light-chain fragments of the antibody, while the B and D chains are each 434 residues long and constitute the heavy-chain segments of the antibody. Within these chains the residue

positions 1-107 (A and C) and 1-113 (B and D) form the antigen binding site, which contribute to the bioactivity of the antibody.

As can be observed from the topographical map of positive residues for 1IGY (Fig 5.24a), there exists a unique site corresponding to an orientation angle of Φ, θ, (80º, 137º) where the antigen binding site is oriented towards the interface with the Fab fragment oriented above as depicted in Fig 5.25a, b, c. However, the topography map of the 'more positive residues' in 1IGY (Fig 5.24b), predicts the adsorption in this orientation to be unfavorable due to the domination of the number of negative residues (repulsive interaction) when compared to the positive residues (attractive interaction) at the site corresponding to (80º, 137º). The orientation corresponding to the maximum 'more positive' interaction indicates the orientation in Fig 5.25d



(a)

(b)

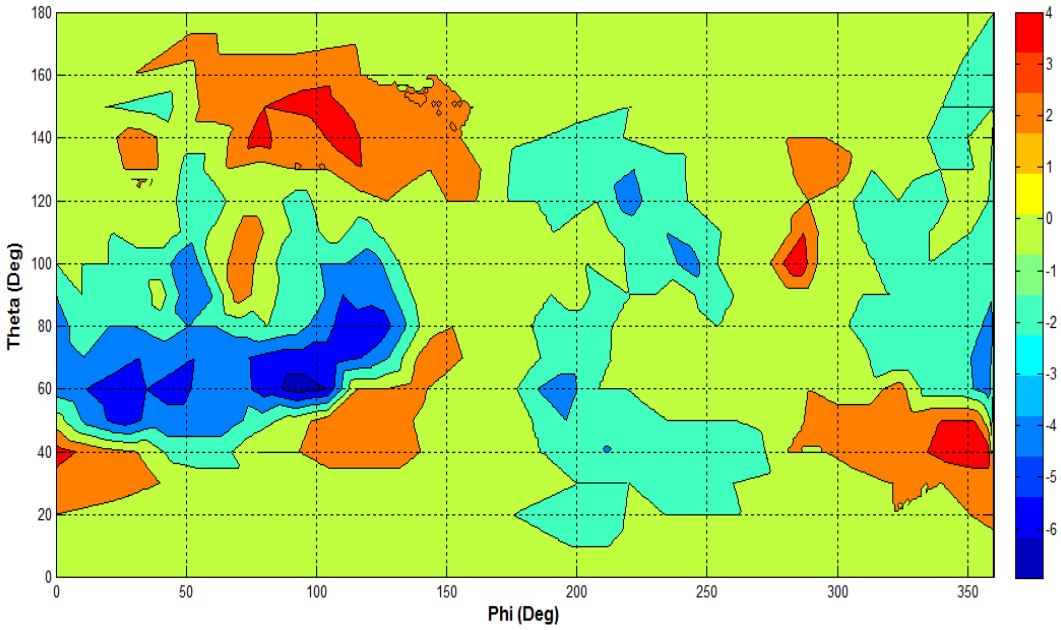
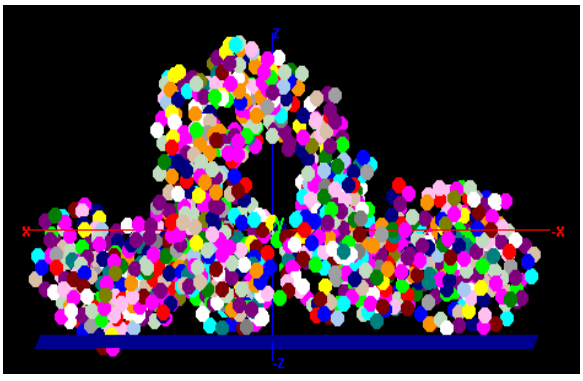
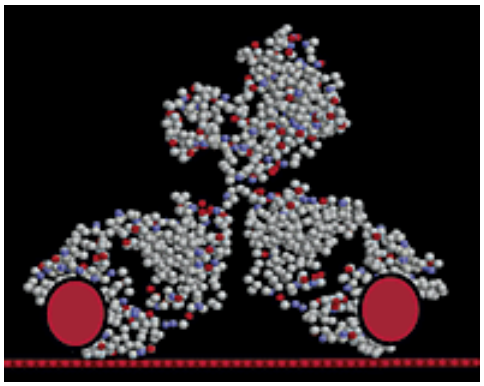


Figure 5.24: Distribution of positive residues in IgG1 (1IGY). (a) The dark red region corresponds to the maximum positive residue concentration of 14 for which the orientation angle (Φ, θ) is (80º, 137º). (b) The 'more positive' topography map of 1IGY corresponding to orientation angle (Φ, θ) is (80º, 137º) however, indicates that the adsorption to be unfavorable as the number of repulsive amino acids dominates (-3).

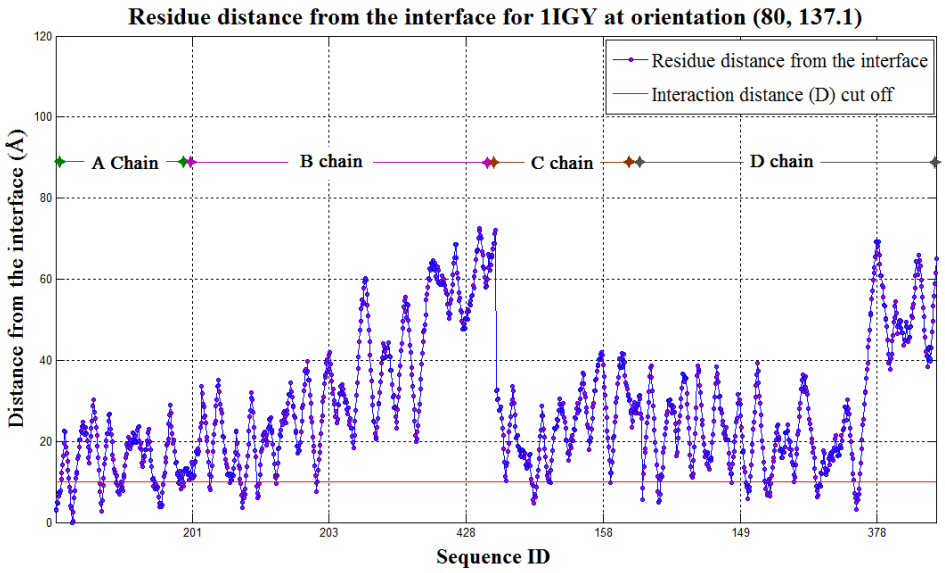
(a)



(b)



(c)



(d)

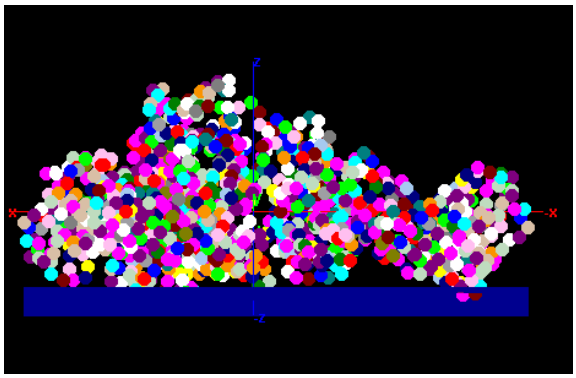
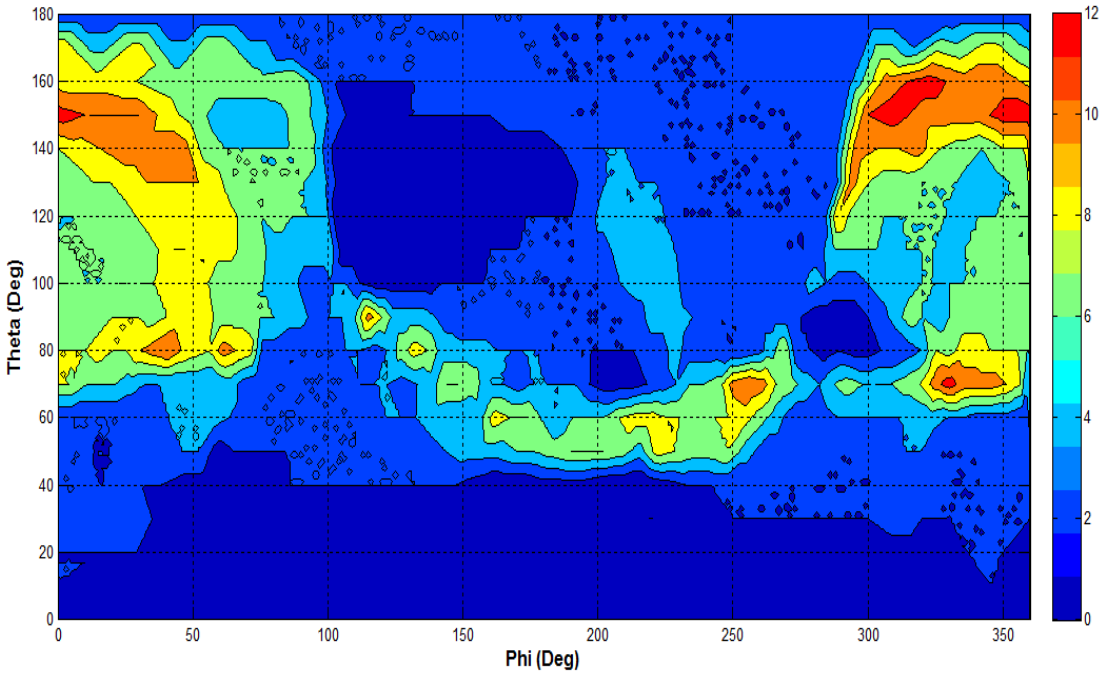


Figure 5.25: Orientation of IgG1 (1IGY) on negatively charged SAM for the orientation angle (80º, 137º) (a) Prediction by the current study. Dark blue indicates adsorbent surface plane. Residues are colored according to RASMOL[90] color scheme. (b) Prediction by united atom coarse-grained model[85]. Large red circles indicate bioactive site regions (c) Residue height from the interface for IgG1 (1IGY) for Φ, θ orientation angle set (80 º, 137º) on a negatively charged surface. The antigen binding site of IgG1 lies within the first 100 residues of each chain. (d) Orientation corresponding to maximum residue concentration of 6 based on the 'more positive' topography map of IgG1

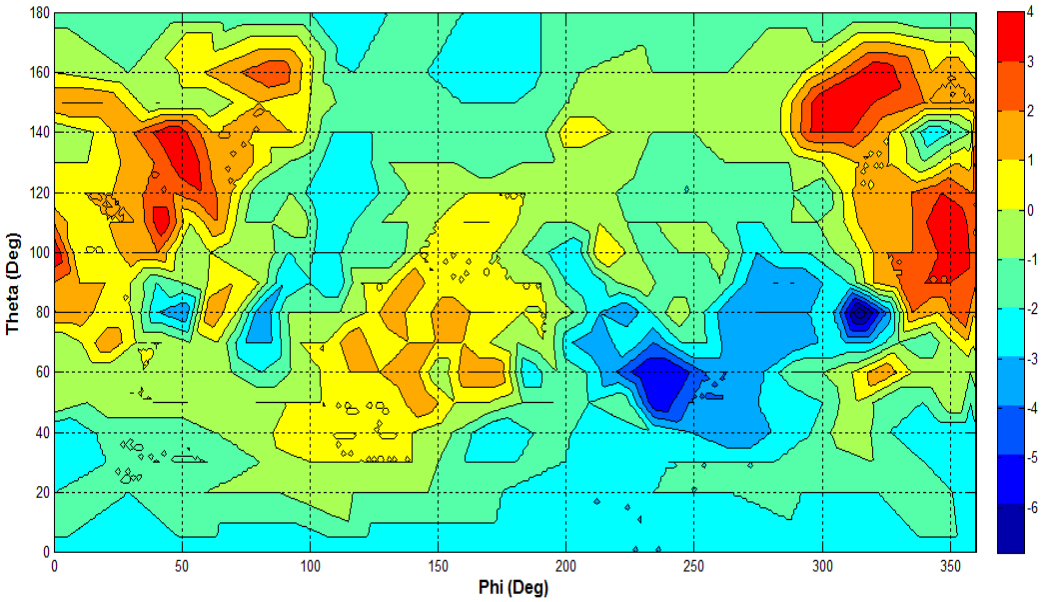
5.1.2.1.3.2 IgG2 (PDB Model: 1IGT)

The PDB model IGT is composed of four chains, namely the A chain, B chain, C chain, and D chain. The A and C chains are each 214 residues long and form the light-chain fragments of the antibody, while B and D chains are each 444 residues long and constitute the heavy-chain segments of the antibody. Within these chains the residue positions 1-108 (A and C) and 1-115 (B and D) form the antigen binding site, which contribute to the bioactivity of the antibody.

The topographical map of positive residues in 1IGT suggests multiple sites where the maximum number of positively charge residues is predicted to be located such to interact with a negatively charged surface (Fig 5.26).



(a)

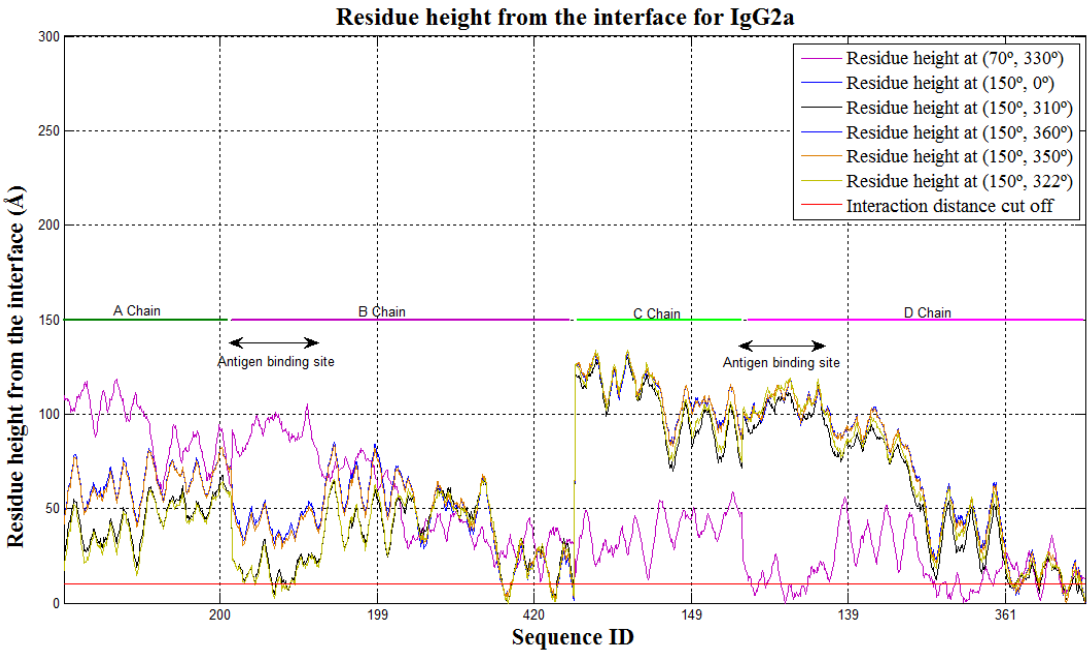


(b)

Figure 5.26: (a) Distribution of positive residues in IgG2a (1IGT) in the topography map of positively charged residues. The dark red regions corresponds to the positive residue concentration of 12 for which the orientation angles (Φ, θ) are (70º, 330º), (150º, 0º), (150º, 310º), (150º, 322º), (150º, 350º), and (150º, 360º). (b) The distribution of 'more positive' residues in the 'more positive' topography map of IgG2a. The maximum residue concentration of 4 is indicated by the red regions.

The orientation angles corresponding to maximum residue concentration sites shows at least one antigen binding site being adsorbed down on the interface, which is consistent with that observed on a negatively charged interface (Fig 5.27).

(a)



(b) (c) (d)

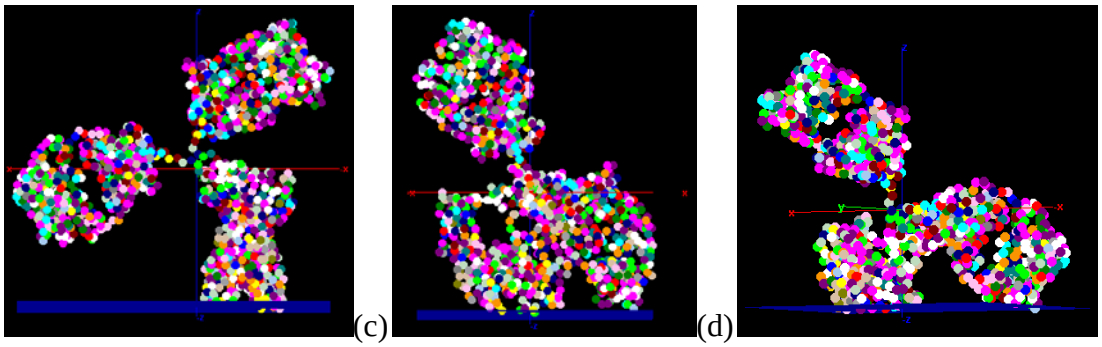


Figure 5.27: (a) Residue height from the interface for IgG2 (1IGT) on a negatively charged surface. The bioactive site of each chain lies within its first about 100 residues. Orientation of IgG2 (1IGT) on a negatively charged surface for Φ, θ orientation angles: (b) (70º, 330º), (c) (150º, 310º), and (d) (150º, 322º). Based on the combined analysis of Fig 5.26(a) and Fig 5.26(b), orientation angles (Φ, θ) around (150º, 310º) and (150º, 322º) are considered to be more preferred.

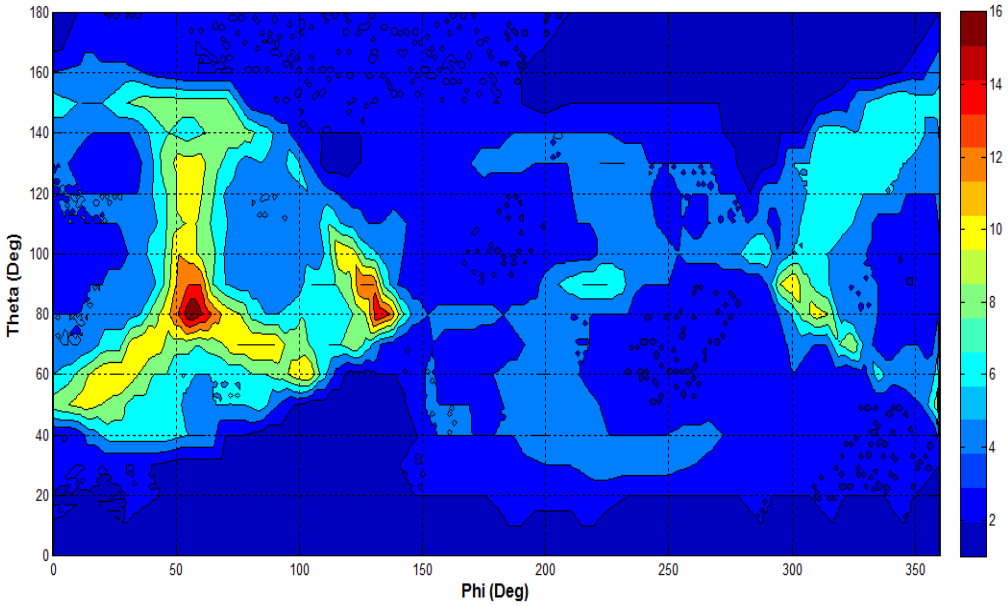
5.1.2.2 Prediction of orientation on a positively charged surface

The orientation of protein on a positively charged surface could be predicted by the topographic mapping of the negative residues on the protein surface. For the determination of the most probable site of adsorption a topographic map of the 'More negative regions' were performed to account for the presence of the counteracting repulsive effect of the positively charged residues with the positively charged surface.

5.1.2.2.1 Immunoglobulin G (IgG) antibody

Similar to the orientation of IgG on negatively charged surfaces; positively charged surfaces have also shown a preferential adsorption of IgG by a united-atom coarse-grained model[85], thereby suggesting a similar value in generating a topology map of these molecules for negatively charged residues.

5.1.2.2.1.1 IgG1 (PDB Model: 1IGY)



(a)

(b)

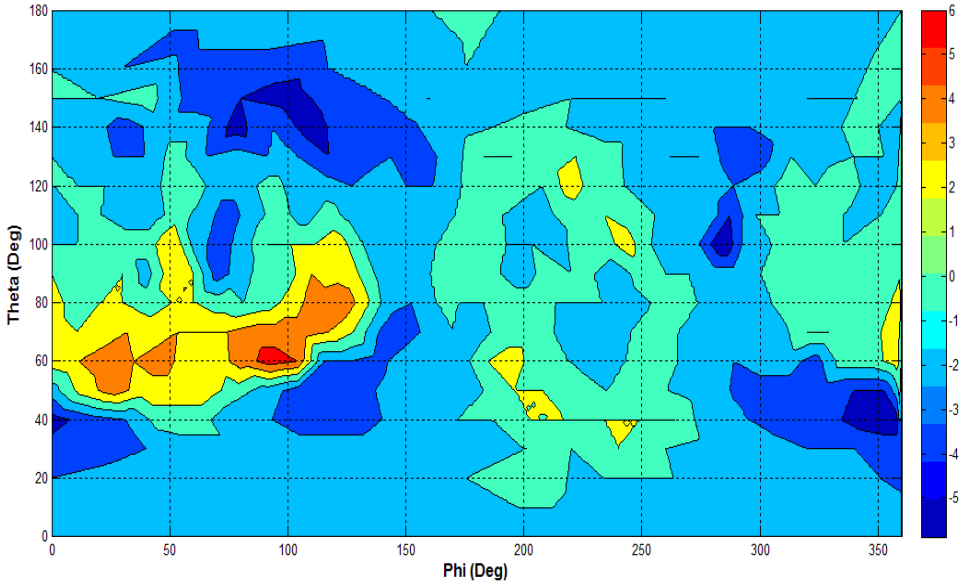
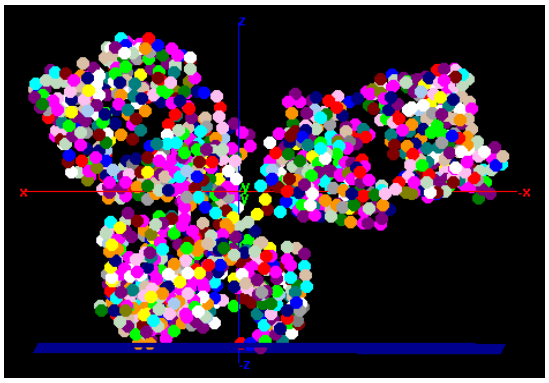


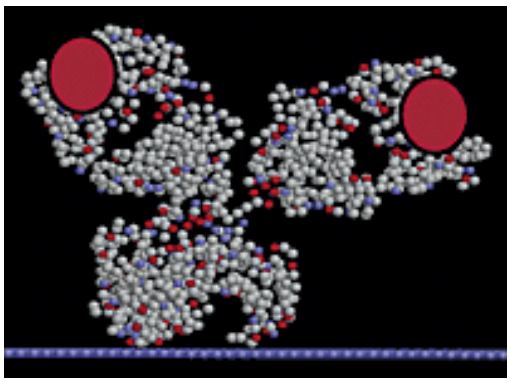
Figure 5.28: Distribution of negative residues in IgG1 (1IGY). (a) The dark red region corresponds to the negative residue concentration of 16 for which the orientation angle (Φ, θ) is (80º, 56º). (b) The red region corresponds to the 'more negative residue' concentration of 6 for which the orientation angle (Φ, θ) is (65º, 90º)

The topographical map of negative residues in IgG1 (Fig 5.28) suggests a unique site (80º, 56º) wherein the antibodies are oriented with the antigen binding site oriented away from the interface. (Fig 5.29)

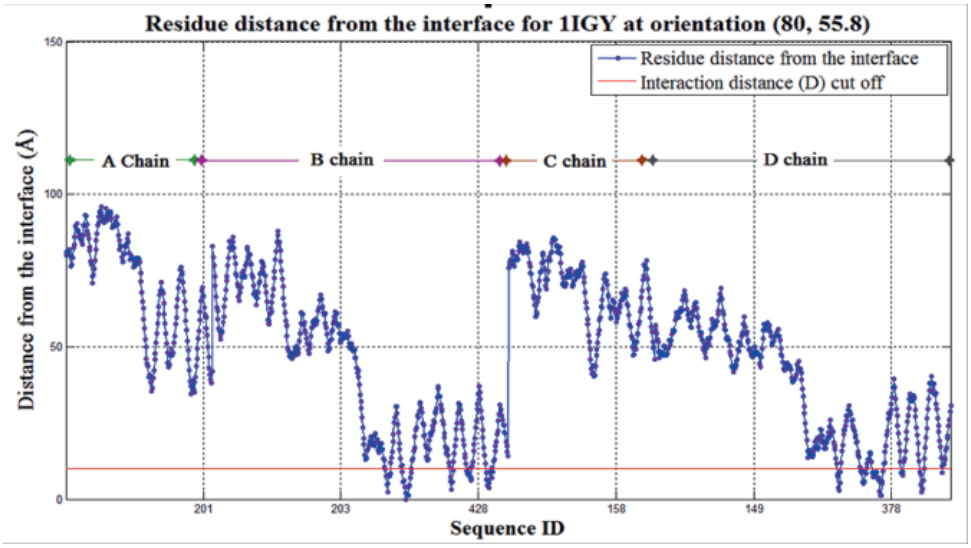
(a)



(b)



(c)



(d)

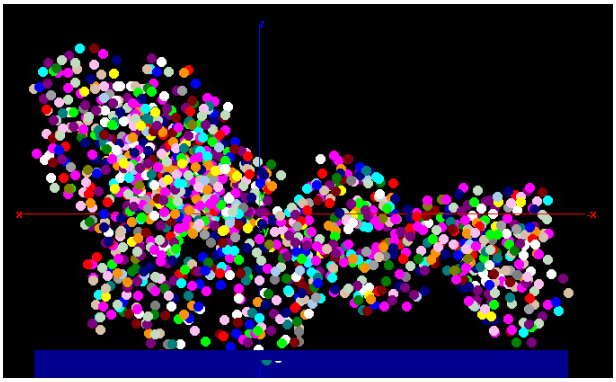
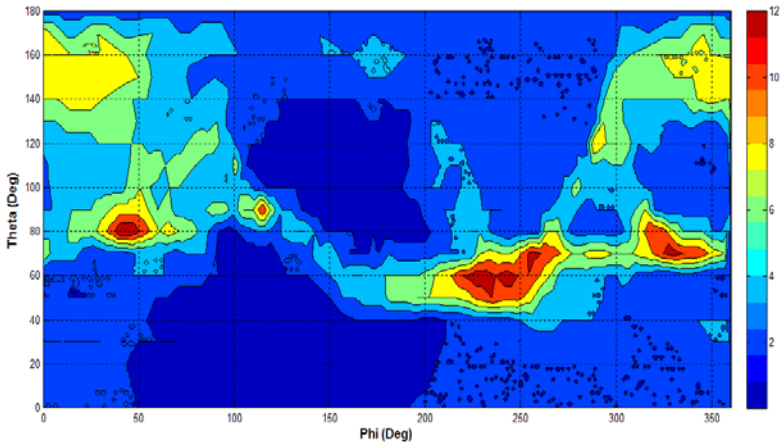


Figure 5.29: Orientation of IgG1 (1IGY) on positively charged surface (a) Prediction by the current study for the Φ, θ orientation angle set (80º, 56º). Dark blue indicates adsorbent surface plane. Residues are colored according to RASMOL[90] color. (b) Prediction by united-atom coarse-grained model[85]. Blue horizontal line at the bottom of the figure indicates the adsorbent surface plane. (c) Residue height from the interface for IgG1 (1IGY) for Φ, θ orientation angle set (80 º, 56º) on a positively charged surface. The bioactive site of each chain lies within its first about 100 residues. (d) Prediction by the current study for the Φ, θ orientation angle set (65º, 90º).

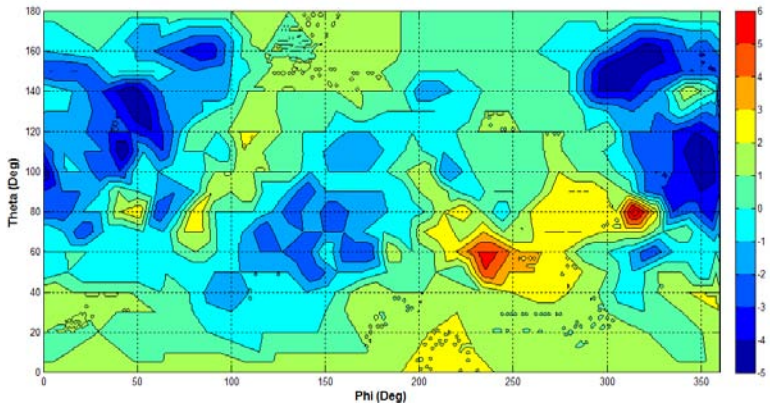
5.1.2.2.1.1 IgG2 (PDB Model: 1IGT)

The topographical map of negative residues in IgG2 (Fig 5.30) suggests multiple sites wherein the antibody is oriented with at least one antigen binding site being oriented away from the surface (Fig 5.30).

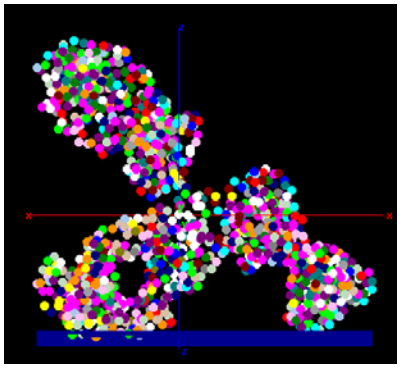
(a)



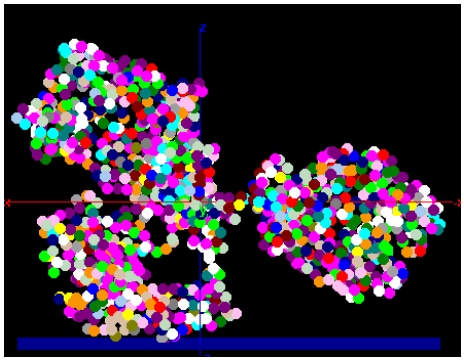
(b)



(c)



(d)



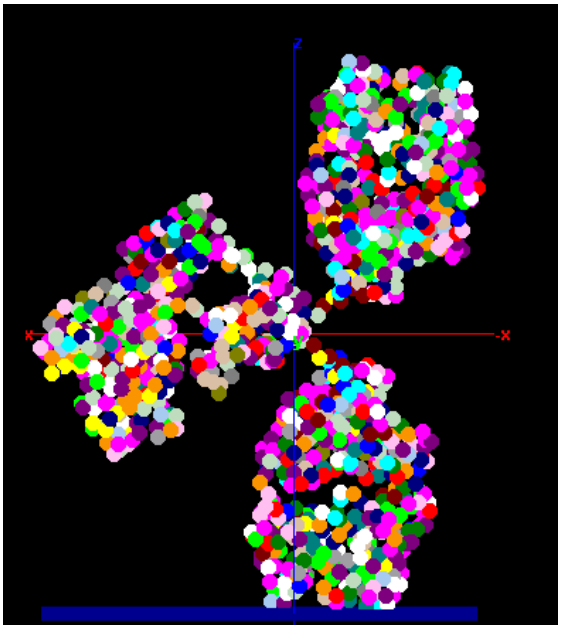
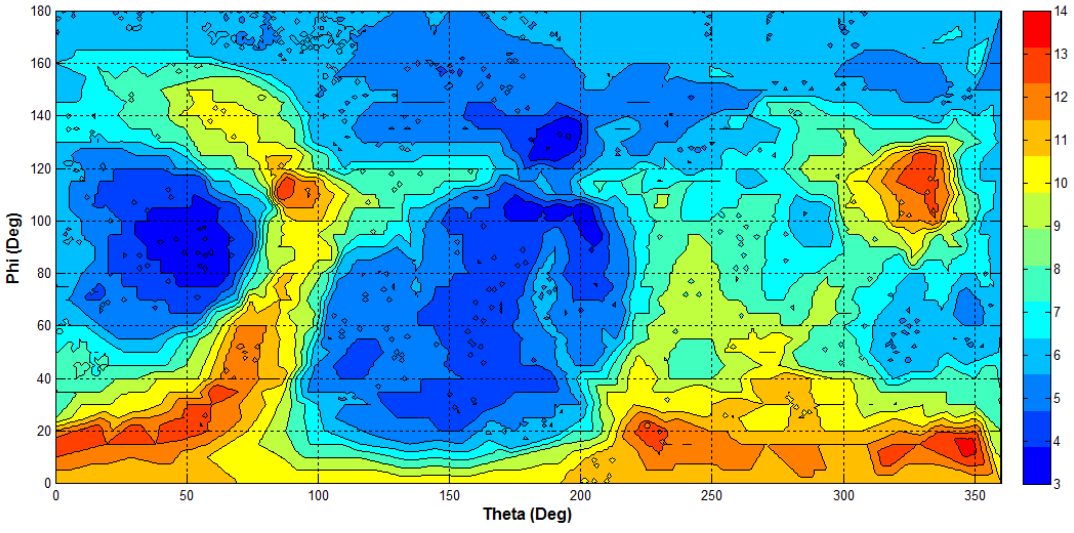
(e) 

Figure 5.30: (a) Distribution of negative residues in IgG2 (1IGT). The dark red regions correspond to the residue concentration of 13 for which the preferred orientation angles (Φ, θ) are (80º, 41º) and (80º, 46º). (b) The maximum residue concentration in 'more negative' residue topography map indicates the residue concentration of 5 for the orientation angles (Φ, θ) are (80º, 315º). (c) Orientation of IgG2 (1IGY) on positively charged surface for the orientation angle (80º, 41º) corresponding to the maximum residue concentration in Fig. 5.30(a). (d) Orientation of IgG2 (1IGY) on positively charged surface for the orientation angle (80º, 315º) corresponding to the maximum residue concentration in Fig. 5.30(b). (e) Orientation of IgG2 (1IGY) on positively charged surface for the orientation angle (60º, 240º) corresponding to the maximum concentration of residue by the combined analysis of Fig 5.30(a) and (b). Dark blue indicates adsorbent surface plane. Residues are colored according to RASMOL[90] color. Lower horizontal line indicates adsorbent surface plane.

5.1.2.2.2 Bovine β-Lactoglobulin (BLG PDB Model: 3BLG)

BLG is one of the whey proteins in cow's milk that has been known to adsorb rapidly onto metallic surfaces such as stainless steel, gold, and chromium. In dairy industries the operational costs incurred due to the adsorption of these proteins is substantial, hence the adsorption behavior of these proteins on metallic surfaces at acidic pH have gained considerable interest as metallic surfaces are considered to be positively charged at these pH.[39]

Since the primary driving force of the interaction between metallic surfaces and protein is believe to be due to electrostatic interactions, the orientation of these proteins on metallic surfaces at acidic pH could be predicted by the topographic mapping of the negatively charged residues constituting BLG. The topography map of the negative residues in the BLG indicates multiple sites of preferred orientation (Fig 5.31), which are listed in Table 5.2.



(a)

(b)

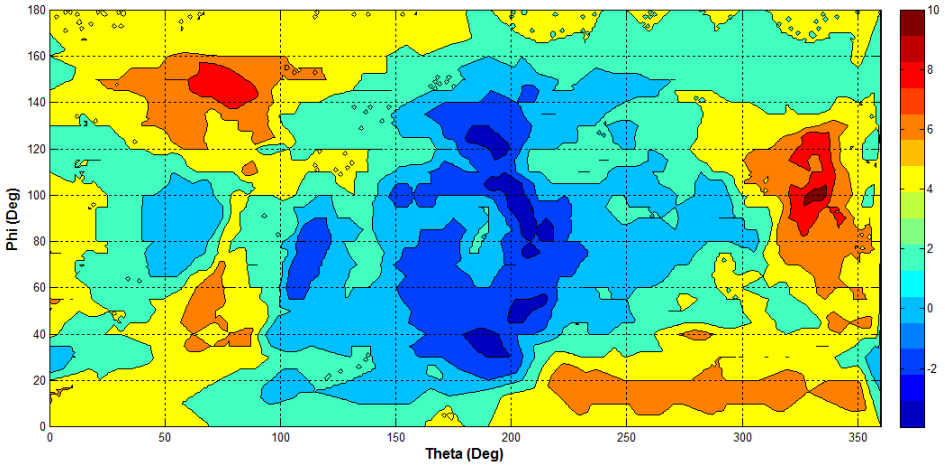


Figure 5.31: Distribution of negative residues in BLG (3BLG). (a) Topography map of negatively charged residues. The dark red regions correspond to the negative residue concentration of 14 for which the preferred Φ, θ orientation angles are listed in Table 5.2. (b) Topography map of the 'more negatively' charged residues for which the maximum residue concentration corresponds to (Φ, θ) orientation angle (100 º, 331º)

Table 5.2: Orientation angles corresponding to maximum residue concentration on a positively charged interface

Theta (degrees)	Phi(degrees)	Residue closest to interface
10	317	17
	345	
15	348	157
20	29	157
	44	17
110	90	63
120	329	34

Experimental and simulation studies have been conducted to understand the molecular mechanism mediating the adhesion of BLG to metallic surfaces, such as gold and stainless steel, and the orientations in which these protein adhere to these surfaces were found to be different. [39] These dissimilarities are expected especially since the surface charge density and packing of metallic atoms in each of these interfaces are different. In addition, it should be recognized that the force field parameters utilized for the simulation studies may also be a substantial contributory factor influencing these results.[39] However, using the current developed technique, the θ, Φ orientation angle set (10º, 317º) was found to be similar to the orientation prediction on a gold surface and the orientation angle set (100º, 331º) and (120 º, 329º) was found to be similar to the adsorbed orientation on stainless steel. The residue distributions on the interface are depicted in the Figure 5.33 and were found to be in excellent agreement with the predicted results.[39]

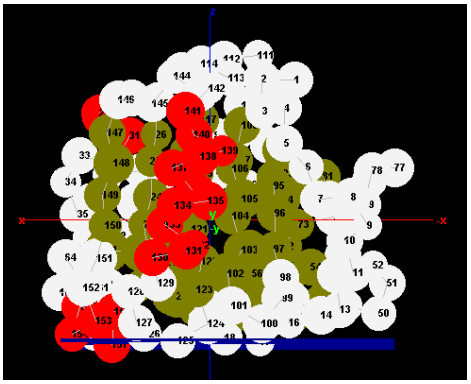
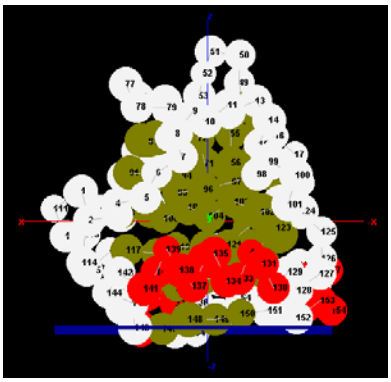
(a)  (b) 

Figure 5.32: Preferred orientation for BLG on positively charged surface at Φ, θ orientation angles (a) (15º, 347º), (b) (120 º, 329º). Blue plane indicates adsorbent surface plane.

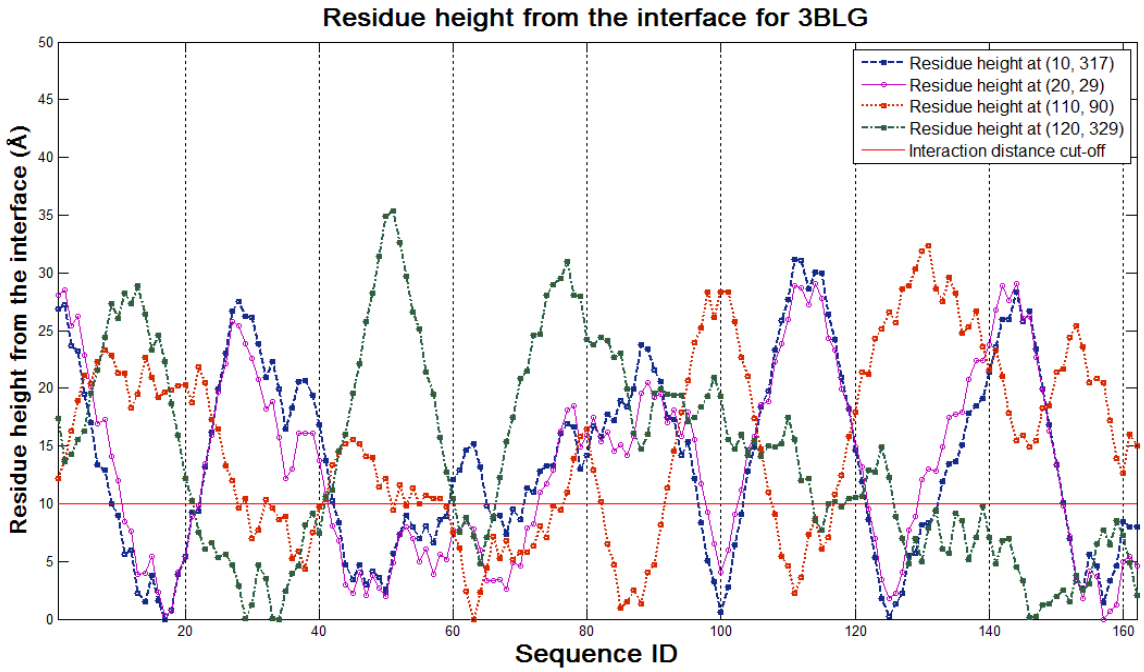


Figure 5.33: Residue height from the surface for BLG (3BLG) on a positively charged surface. The blue line corresponds to the distribution at (Φ = 10º, θ = 317º), magenta corresponds to (Φ = 20º, θ = 29º), red corresponds to (Φ = 110º, θ = 90º), and green corresponds to (Φ = 120º, θ = 329º).

5.2 Prediction of Orientation on Mixed Surface Chemistries

Most of the natural and synthetic materials are not uniformly charged or uncharged, but present surface chemistries that are a mixture involving many different types of functional groups. The accurate prediction of protein orientation in such topologies could be useful in controlling the orientation of the protein, which in turn could lead to design surface for controlling the bioactivity of proteins on any given surface chemistry on interface.

5.2.1 Prediction of protein orientation in a biological membrane

Biological membranes are natural examples of interfaces which present a mixed surface chemistry for interacting with protein through electrostatic and hydrophobic interactions.[35] Studies involving the prediction of the orientation of membrane proteins have determined the hydrophobic thickness or the depth of interaction corresponding to the protein orientation on the membrane surface to be about 10 Å.[80, 81] Hence using the default settings, the above method could also aid in predicting the orientation of proteins on biological interfaces. Prediction of protein orientation on biological membranes was limited to membrane bound proteins.

5.2.1.1 Sarcomeric mitochondrial creatine kinase (MtCk PDB Model: 2GL6)

MtCK is a membrane-bound protein that transports high energy phosphate from mitochondria to the cytosol creatine. MtCk occurs in two different oligomeric forms: dimeric and octameric. The orientation of this protein in the membrane through cardiolipin is important for the channeling function of the metabolite thereby regulating the mitochondrial respiration.[87]

The PDB model 2GL6 is an octameric structure but each of these chains contains many missing residues especially at the N-terminus, and hence may not accurately predict the orientation of the protein in cell membrane. However residues Lys369, Lys379, and Lys380 flanking the C-terminus are considered to make up the site of attachment for association of MtCk with the phospholipid membrane, the missing

residues are not considered to affect the predicted orientation.[35] The topographical map for positive and hydrophobic residues in 2GL6 suggests two Φ, θ, sites, (110º, 53º) and (150º, 80º), where the maximum number of residues is seen to interact with the surface (Fig.5.34). The orientation and residue distribution for (110º, 53º) is depicted in Fig 5.35.

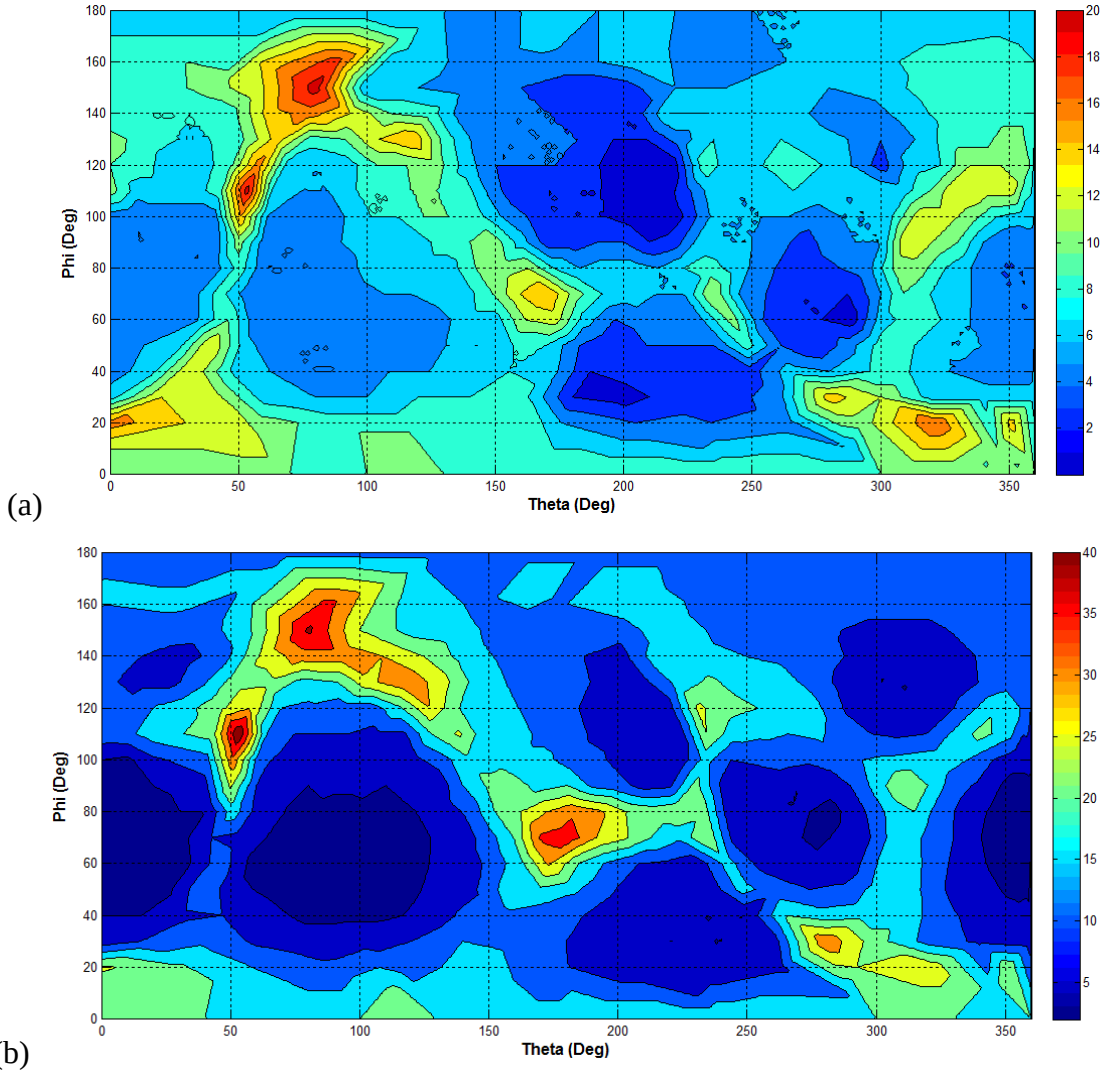


Figure 5.34: Topography map of MtCk for (a) Positively charged residues, and (b) hydrophobic residues.

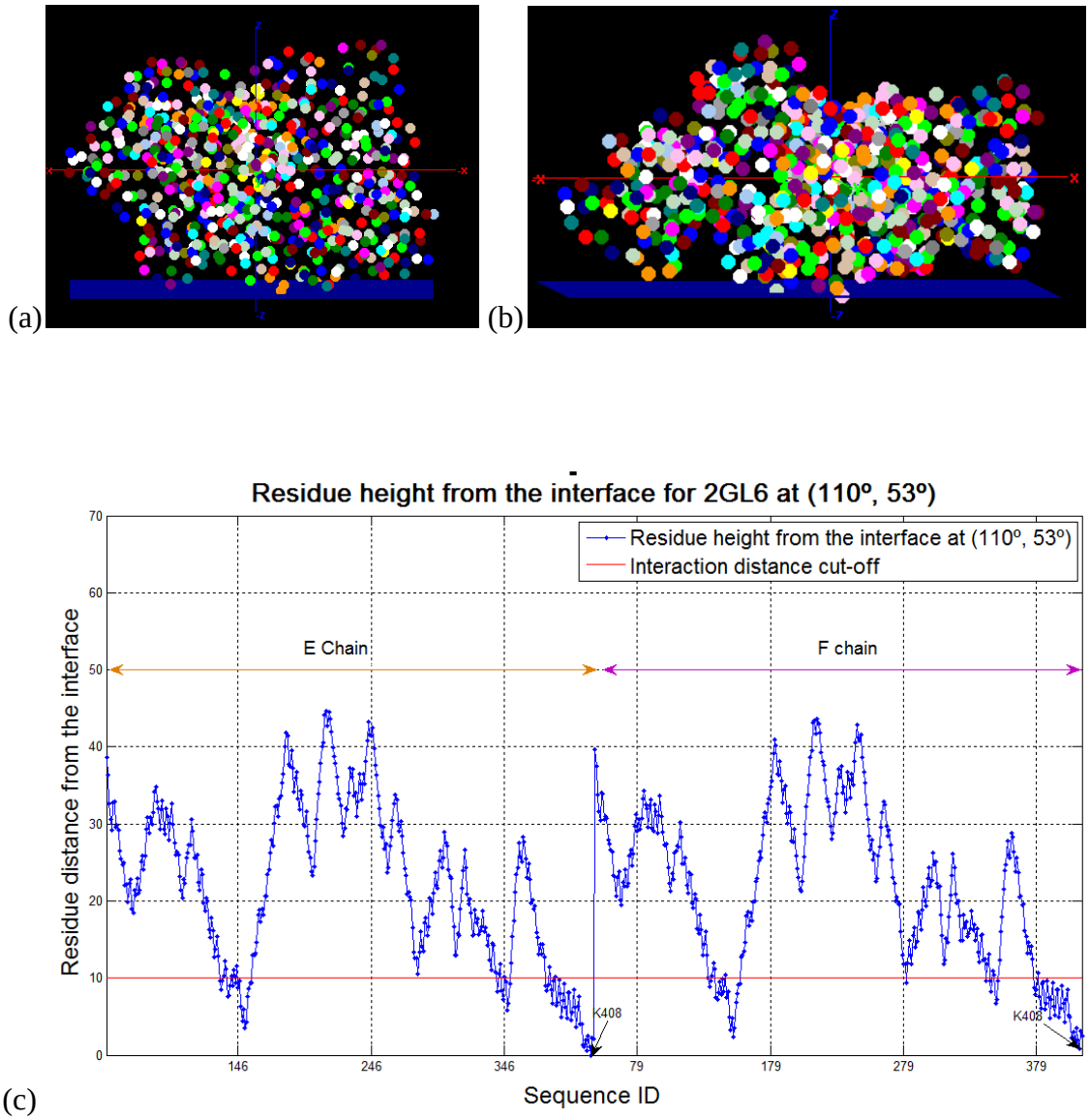
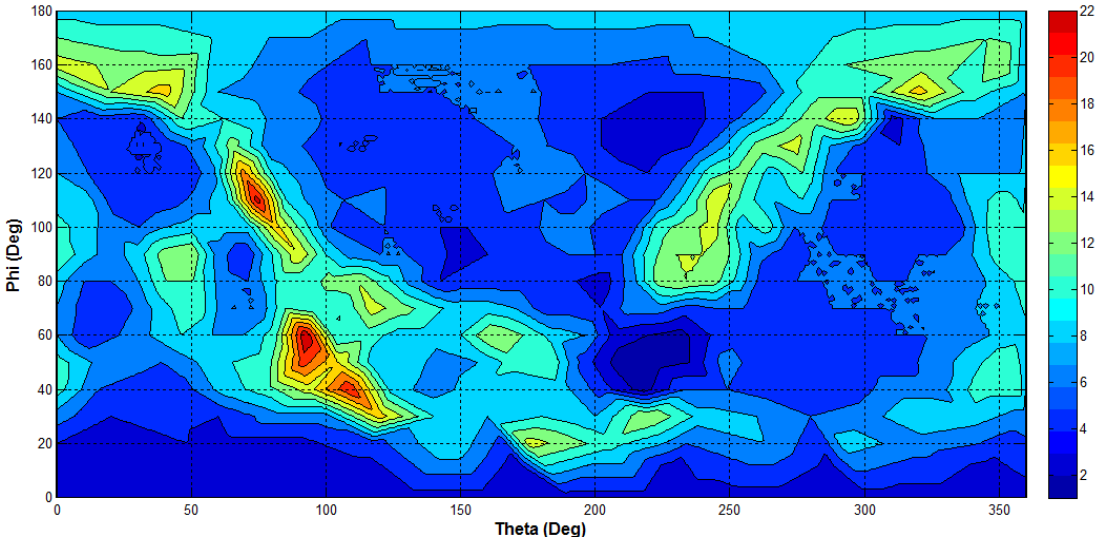


Figure 5.35: Predicted orientation of E and F chain of MtCk in a biological membrane for Φ, θ orientation angles (a) (110º, 53º) and (b) (150º, 80º). Dark blue indicate adsorbent surface plane. (c) Residue distribution on the biological membrane for MtCk (2GL6) on a biological membrane at (Φ = 110º, θ = 53º).

5.2.1.2 Type 1 human hexokinase (HXK PDB Model: 1HKC)

HXK is a key enzyme in the glycolysis process and is found to be associated with the outer mitochondrial membrane. Attachment of the protein to the membrane is postulated to be through the N-terminus. The PDB model 1HKC contains a single chain of 917 residues for which the topographical map of positive (Fig 5.36a) and hydrophobic (Fig 5.36b) residues suggests Φ, θ orientation (110º, 74º) to be the orientation in which the HXK is oriented in a biological membrane (Fig 5.37a). The residue heights corresponding to this orientation indicates the N-terminus to be oriented towards the membrane surface (Fig 5.37b). The predicted orientations are consistent with those predicted for HXK on a model biological membrane.[35]

(a)



(b)

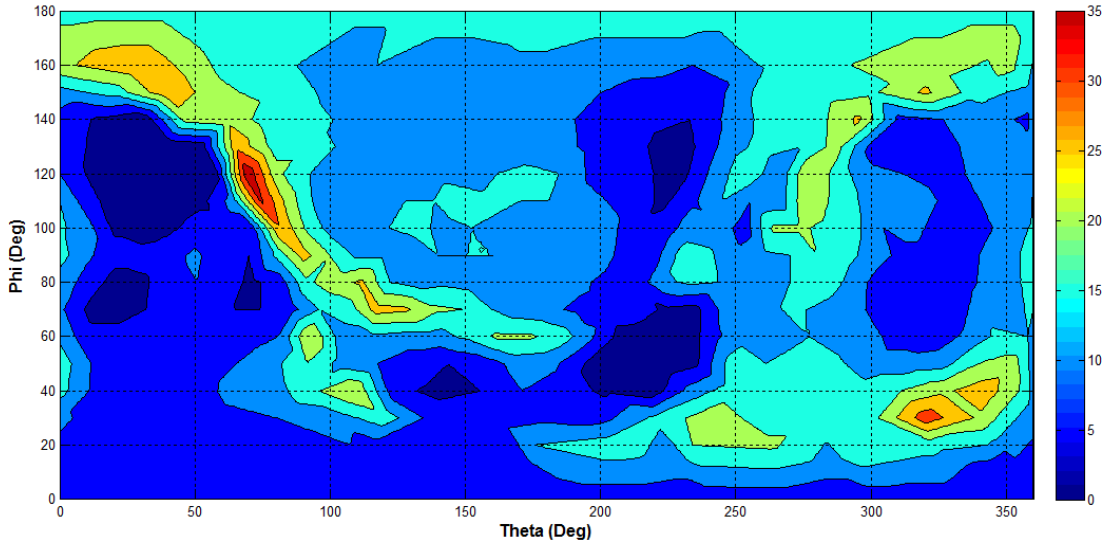
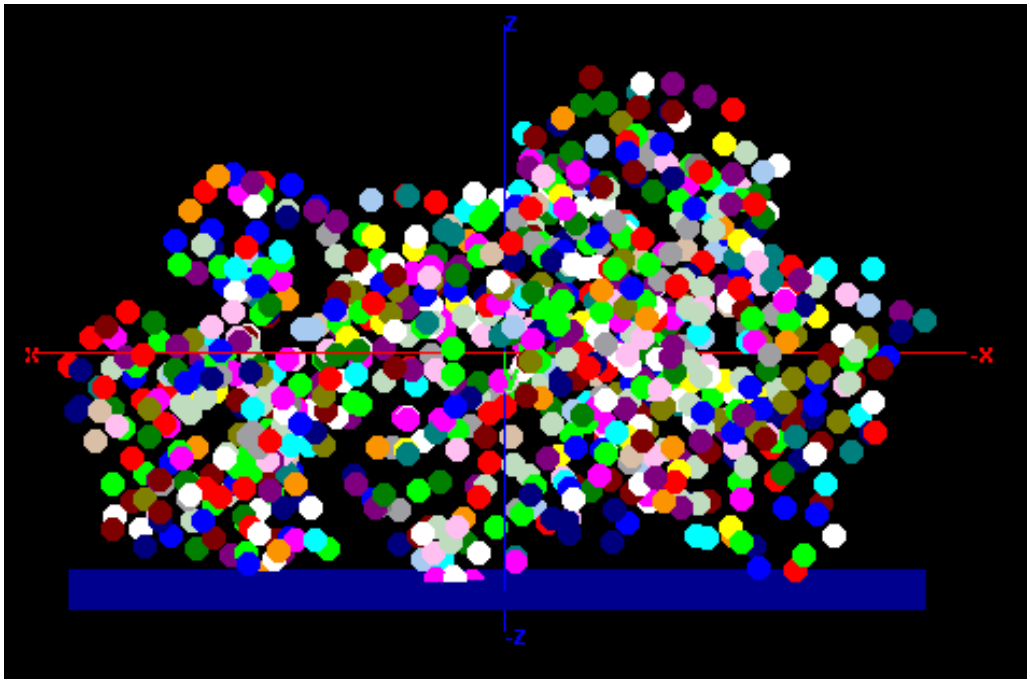


Figure 5.36: Topography map of HXK (1HKC) for (a) Positively charged, and (b) hydrophobic residues

(a)



(b)

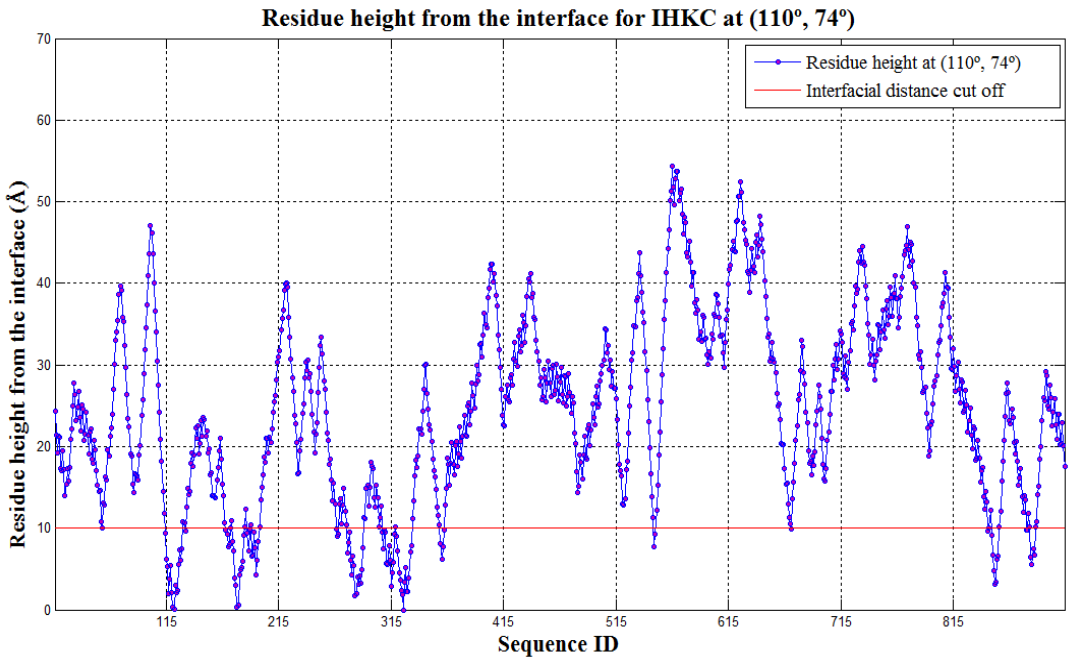
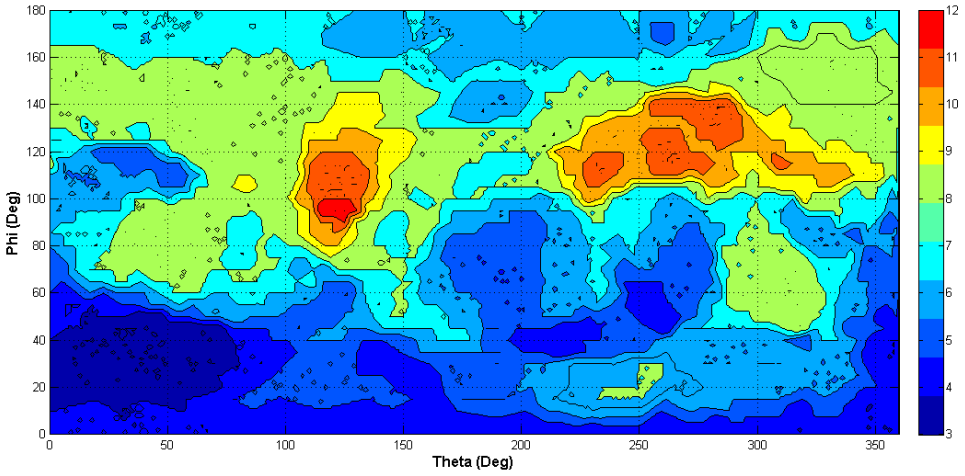


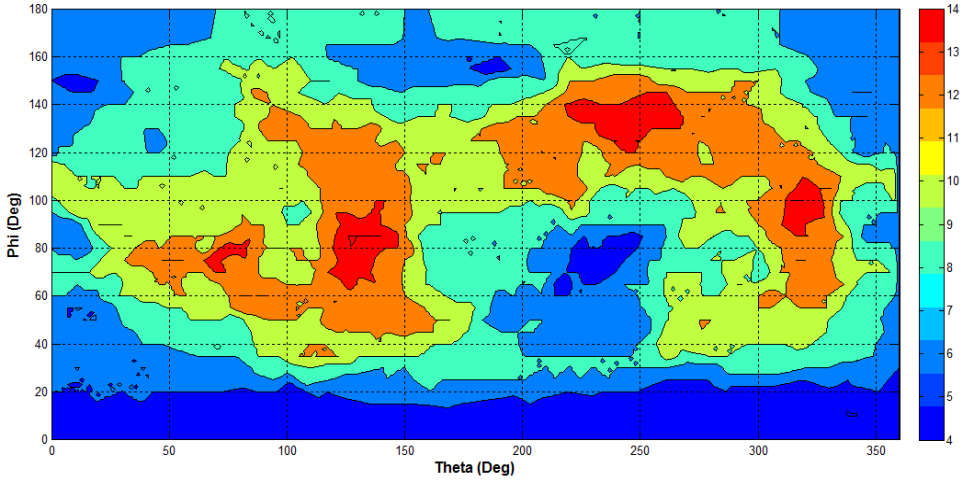
Figure 5.37: (a) Predicted orientation (dark blue indicates membrane surface), and (b) residue distribution of the type 1 hexokinase on the model biological membrane surface for HKK (1HKC) at (Φ, θ) orientation angle set (110º, 74º).

5.2.1.3 Mitochondrial cytochrome c (CYTC, PDB Model: 1HRC)

Prediction of CYTC orientation on a biological membrane corresponds to Φ, θ orientation angle set (90º, 125º) based on the topographic map of CYTC for hydrophobic (Fig 5.38a) and positively charged residues (Fig 5.38b). These figures also correspond to Fig 5.16, which is in accord with the prediction by some research groups[80] but is in discord with others.[35]



(a)



(b)

Figure 5.38: Topography map of CYTC (1HRC) for (a) positively charged, and (b) hydrophobic residues.

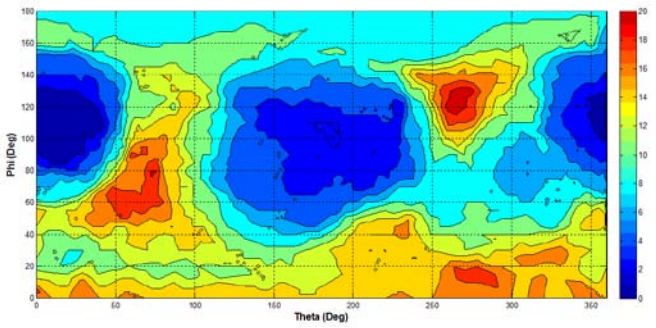
5.2.2 Prediction of orientation on non-biological interface

The orientation predicted for adsorbed proteins using biological membranes and CH_3-, OH-, COOH-, and NH_2-SAMs are model surfaces for a variety of polymers and metallic surface but do not necessarily relate to all type of polymers such as polystyrene

or nylon. Surface functionalities influence the interaction distance (D) and hence could affect the initial orientation in which the protein adsorbs onto the surface, thereby affecting the predicted adsorbed protein orientation. In order to validate this assumption, the orientation of bovine α-lactalbumin (BLA) adsorbed on polystyrene was predicted by varying the parameter 'D'.

5.2.2.1 Bovine α-lactalbumin (BLA PDB Model: 1F6S)

Using the default setting of 'D' as 10 Å, the topographic map of hydrophobic residues in (Fig 5.39a) predicts a preferential adsorption site with Φ ranging between 120º-130º and θ ranging from 260º-280º, for which the predicted orientation is similar to Fig 5.39b.

(a) 

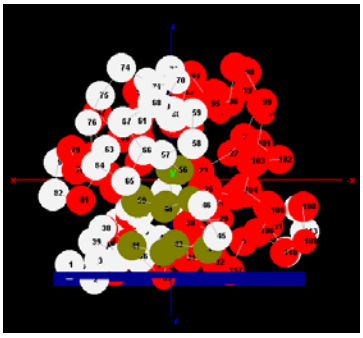
(b) 

Figure 5.39: (a) Topography map of hydrophobic residues in BLA. (b) Orientation of BLA on hydrophobic surface corresponding to θ, Φ orientation angles (130º, 268º).

However, simulation results of peptide interactions on an aromatic ring functionalized substrate have revealed the minimum energy interaction to be lesser than

10 Å .[73]. An interaction distance cut-off (D) of 7.5 Å could thus best describe the most favorable interaction of hydrophobic residues on interfaces with phenyl functional groups such as polystyrene.[73] Thus the 'D' was set as 7.5 Å for current new set of predictions. The topography map of hydrophobic residues in BLA under such conditions (Fig 5.40a) reveals two potential sites for favorable interaction. The site with Φ ranging between 120º-130º and θ ranging from 260º-280º is similar to Fig 5.40. The other potential site of protein adsorption corresponds to the Φ, θ orientation angle set of (60 º, 75º) for which the orientation is as depicted in Fig 5.40b.

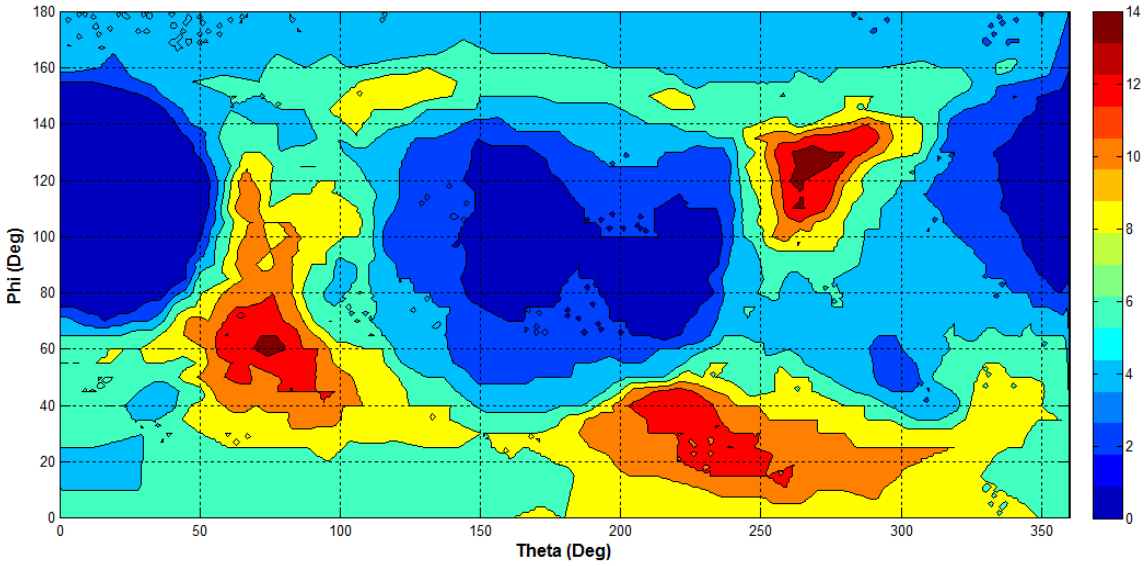
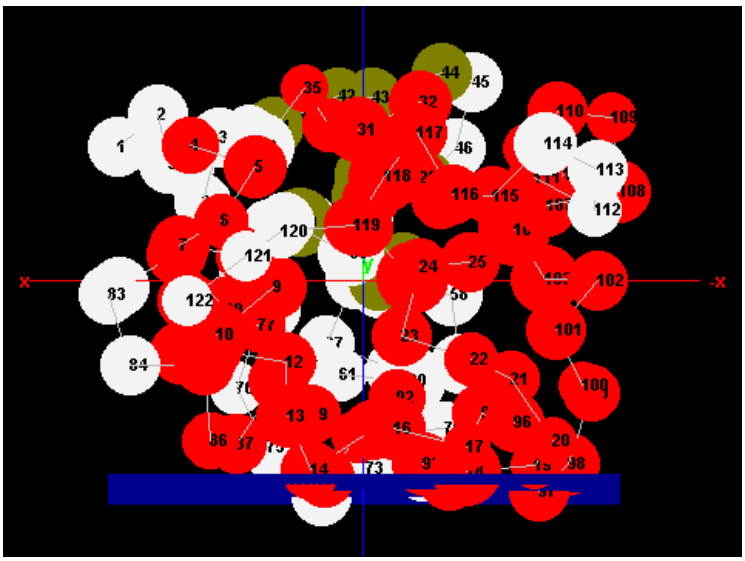


Figure 5.40: Topography map of hydrophobic residues in BLA.

The residue height from the interface corresponding to this orientation is shown in Fig 5.41, which corresponds to the experimentally determined protein orientation on polystyrene nanospheres (60nm) .[88] Although on a macroscopic scale the surface

topography of the nanospheres may not be flat, relative to the much smaller size of the protein, the surface can be considered to closely approximate a flat surface.

(a)



(b)

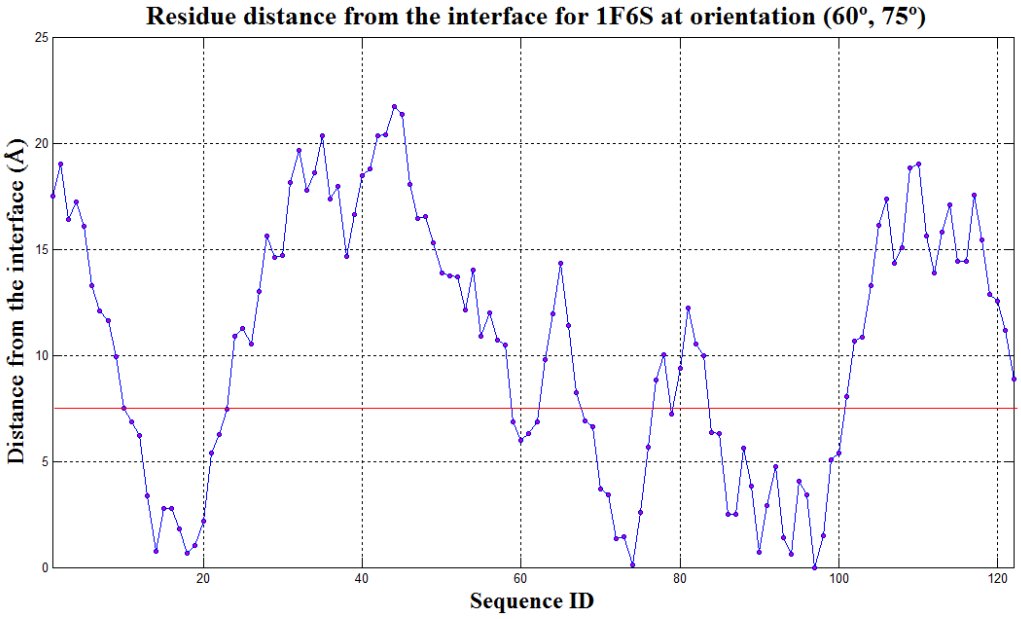


Figure 5.41: (a) Orientation and (b) Residue distribution for BLA (1F6S) on the polystyrene surface for Φ, θ orientation angles (60º, 75º).

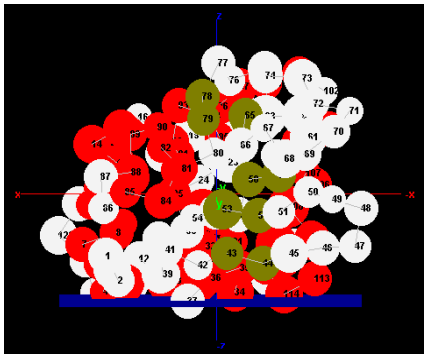
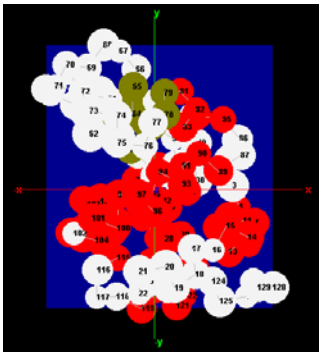
5.2.2.1 Hen egg white lysozyme (HEWL PDB Model: 7LYZ)

The importance of 'D' for a given adsorption-system is further evident on interfaces which are mixed SAM surfaces with varying densities of charge, polar and non-polar functional groups. For example, using the default settings for generating the topography map for HEWL, prediction of orientation of HEWL on mixed surface chemistries yields orientations on different surface chemistries as given in Table 5.3.

Table 5.3: Predicted orientation of HEWL on mixed surface chemistries.

Surface chemistry (1:1 mixture ratio)	**Orientation angle (degrees)**	**Number of residues**	**Orientation**	**Figure**
Hydrophobic/Hydrophilic	(110, 229)	46	Side on active site down	Fig 5.42 (a), (b)
Positive/Negative	(110, 85)	14	Side on active site up	Fig 5.42 (c), (d)
	(110, 229)		Side-on active site down	Fig 5.42 (a), (b)
	(110, 239)		Side on active site down	-
Hydrophobic/Positive	(110, 229)	28	Side on active site down	Fig 5.42 (a), (b)
Hydrophobic/Negative	(110, 255)	32	Side on active site down	Fig 5.42 (e), (f)

Hydrophilic/Positive	(110, 229)	32	Side on active site down	Fig 5.42 (a), (b)
Hydrophilic/Negative	(110, 229)	30	Side on active site down	Fig 5.42 (a), (b)
Hydrophobic/Hydrophilic /Positive	(110, 229)	52	Side on active site down	Fig 5.42 (a), (b)
Hydrophobic/Hydrophilic /Negative	(110, 229)	54	Side on active site down	Fig 5.42 (a), (b)
Hydrophobic/Positive /Negative	(110, 229)	36	Side on active site down	Fig 5.42 (a), (b)
	(110, 255)		Side on active site down	Fig 5.42 (e), (f)
Hydrophilic/Positive /Negative	(110, 229)	38	Side on active site down	Fig 5.42 (a), (b)
Hydrophobic/Hydrophilic /Negative/Positive	(110, 229)	60	Side on active site down	Fig 5.42 (a), (b)

(a)  (b) 

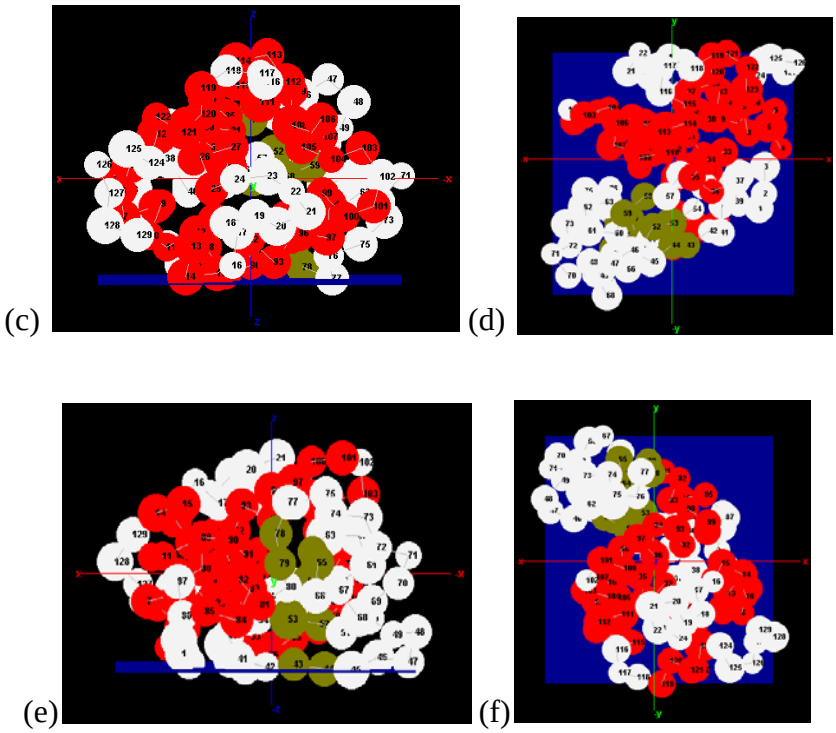


Figure 5.42: Predicted orientations on mixed surfaces. Dark blue indicates adsorbent surface plane. (a) Side view of the (Φ, θ) orientation angles (110º, 229º). (b) Top view of the (Φ, θ) orientation angles (110º, 229º). (c) Side view of the (Φ, θ) orientation angles (110º, 85º) (d) Top view of the (Φ, θ) orientation angles (110º, 85º). (e) Side view of the (Φ, θ) orientation angles (110º, 255º). (f) Top view of the (Φ, θ) orientation angles (110º, 255º).

The orientation angle predicted for mixed surfaces represents the minimum energy positions. Though previous studies on the orientation of HEWL on mixed surfaces predict side-on orientations, these studies have also predicted end-on orientations especially for hydrophilic/positive and hydrophilic/negative surfaces. The end-on

orientation, however, is not predicted in any of the maximum residue concentration sites, thereby indicating that the interaction distance cut-off of 10Å could be an over-estimation of the distance which the surface-induced force is considered to be acting on the protein.

Similar erroneous prediction could affect the predicted protein orientation when the charge density of a charged surface is increased. NH_2-SAM and COOH-SAM have been measured to have charge densities of 11% and 50% respectively at bulk pH of 7.4[91], and the default setting of 'D' is based on this approximation. When these charge densities are decreased or increased a corresponding decrease or increase in the 'D' setting may be expected. Changes in the ionic concentration of the protein solution can be expected to have an even greater effect on the appropriate 'D' parameters because the ion concentration in solution will greatly influence the electrostatic shielding of the charged groups of both the material surface and the protein. These effects could potentially explain the under-estimation of the orientations predicted for IgG1 and IgG2 on charged surfaces especially at high charge densities.

A further variability in the number of the potential orientations is due to the method used to predict the potential interaction. The current method is based on the site which corresponds to the maximum number of residues and not on affinity of the residue to the surface. The affinity of the residue to the surface is influenced by (a) the packing of functional groups presented by the interface and (b) the attractive/opposing force due to the neighboring residues. These two factors are not completely accounted for in the current form of this program and could influence the predicted orientations with which

the protein would interact most favorably with the interface, thereby necessitating a ranking/parameterization based approach. This is very specifically illustrated by the orientation prediction on charged interfaces. By accounting for the opposing residue interaction with the surface, the predicted orientations are different from those predicted by the maximum number of residues alone. In fact by combining the analysis of the topography map predicted by the maximum residue concentration and opposing residue concentration, the predicted orientations are in excellent coherence with those observed in literature, further underlining the importance of a ranking/parameterization based approach.

Despite these shortcomings in the current program design, the accuracy and the computation speed with which the minimum energy orientations are predicted makes the current approach a valuable technique in the selection of the protein orientation for simulation studies involving quantification of the thermodynamic parameters, conformational shifts, and bioactivity prediction.

CHAPTER 6

CONCLUSION

Molecular modeling is currently the best approach to theoretically understand the molecular mechanisms mediating protein adsorption to material surfaces. A critical component in most molecular simulations to investigate protein adsorption behavior is to select the initial orientation(s) with which to begin the simulation. The typical time that it takes for predicting protein orientation on a selected material surface with molecular weight 14-67 kDa is 6-200 hours. This current study was therefore aimed at providing a rapid approach in predicting a preferred orientation on a wide range of surface types by approximating the free energy of adsorption as a function of the number and type of residues that would interact with the surface. These residues were identified based on the criterion that they lie within a defined interaction distance cut-off (D) of the surface when the protein approached the surface within a defined orientation. A topography map was thus generated over all possible rotations of a protein to identify the preferred low-energy orientations of the protein on a given type of surface.

When applied to characterize the applicability of the current approach to a wide range of proteins (11 kDa-150 kDa in size) on a wide range of surface types, the predicted orientations on homogenous interfaces were in excellent agreement with both experimental and molecular simulation data for these proteins with average computation time being in minutes as opposed to hours or even days.

Although the current approach for predicting protein adsorbed orientation on an surface predicts the potential minimum energy orientation, this may not necessarily

correspond to the global minima because of the method being based on rigid protein models and may have to be further verified for a given simulation environment based on the particular force field parameters that are used in the simulation. However the current approach will aid in the rational selection of the initial orientations to be considered as starting points for further studies to investigate the conformation and bioactivity of adsorbed proteins.

The applicability of topography maps could be further expanded to predict the free energy of adsorption by parameterization of the affinity of each type of amino acid residue to specific types of surface functional groups. Other potential applications include rational material selection and design which could help in controlling the inherent bioactivity of adsorbed protein that could be further harnessed for a wide range of applications.

APPENDICES

Appendix A: Expanded methods

Modeling the solvent accessibility

As described in chapter 4, the solvent accessibility was modeled according to Wodak and Janin's analytical approximation[59]. The parameters are as described below:

A.1 Calculation of the radii, surface area and the volume of amino acids

r = Radius of the amino acid computed from van der Waal radii of C,N,O,S atoms.

= (Sum of the van der Waal radii of the CNOS constituting the amino acid)/1.2325

r_w = Radius of the solvent

S = Total surface area of the solvated amino acid. = $4\pi(r + rw)^2$

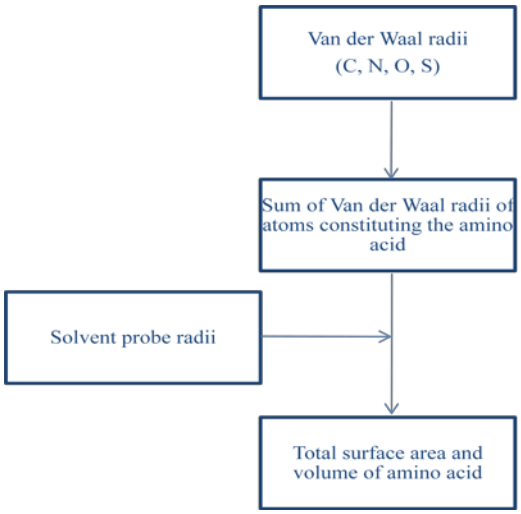


Figure A-1: Flow chart for determining the radii, surface area and volume of the amino acid

A.2 Calculation of the solvent accessibility of individual amino acid and the protein molecule

s = intermolecular penetration depth

d = distance between two amino acid. $= \sqrt{(x1-x2)^2 + (y1-y2)^2 + (z1-z2)^2}$

b_i = maximum buried surface area by molecule 1 on molecule 2

$$= \prod (r_1 + r_w)(r_1 + r_2 + 2r_w - d)(1 + \frac{r_2 - r_1}{d})$$

b_i' = minimum buried surface area by molecule 1 on molecule 2.

$$= \prod (r_1 + r_w)(r_1 + r_2 + 2r_w - s - d)(1 + \frac{r_2 - s - r_1}{d})$$

B' = buried surface area of the molecule 1 $= \sum_{i=1}^{n} bi$

A_i = accessible surface area of the molecule 1 due to B'

$$= S\prod_{i=2}^{n}(1 - \frac{b_i - b_i'}{S})$$

A_c = Effective accessible surface area of the molecule = $\sum_{i=1}^{n} Ai$

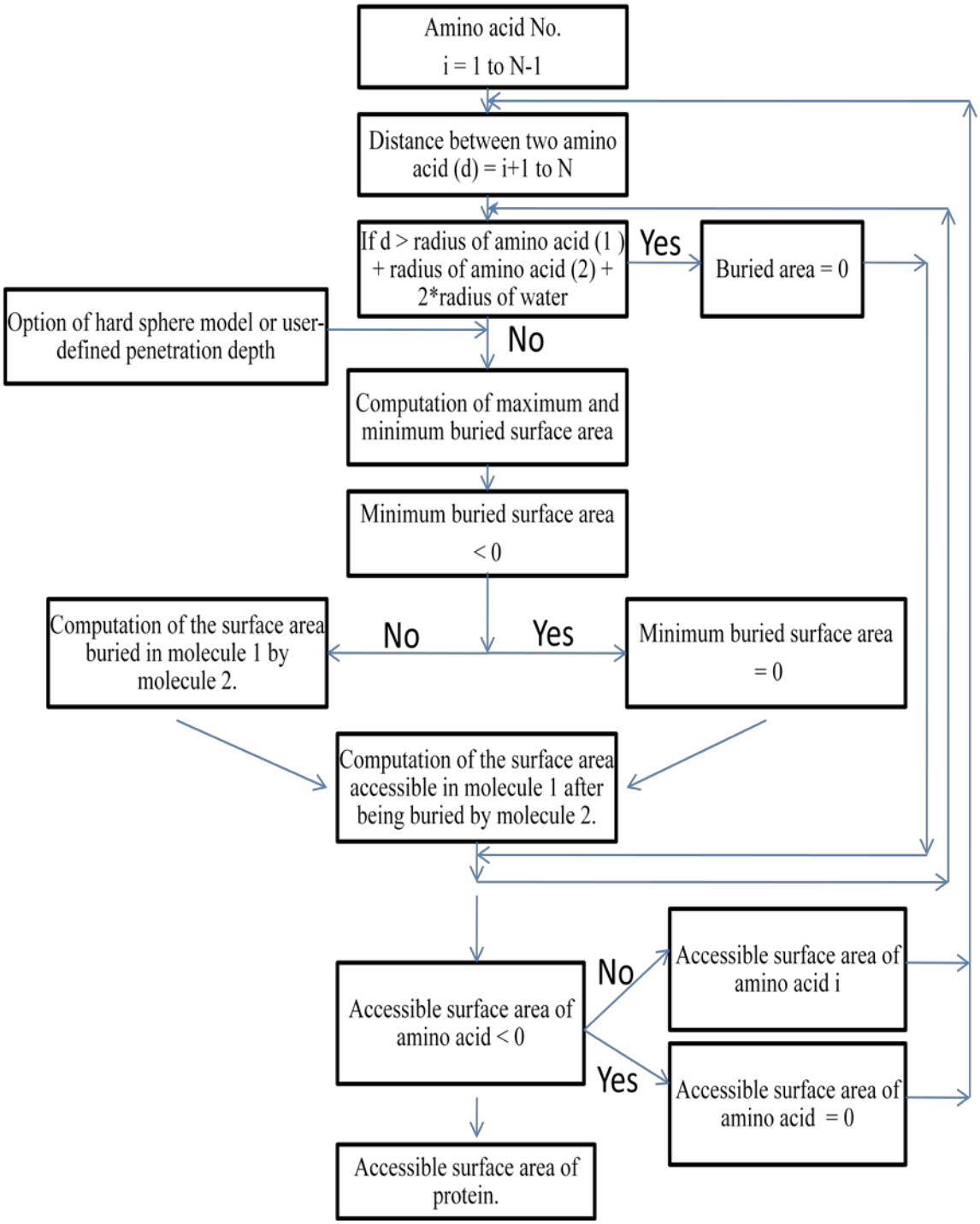


Figure A-2: Flow chart for determining the individual solvent accessible surface area of the amino acid and protein

Table A.1: ASA of the HEWL molecule determined by the algorithm described in Fig. A-1. Reference structure is 1AKI based on which the RMSD of the models have been determined. Resolution and pH are specific to the PDB model.

PDB ID	**RMSD (Å)**	**Resolution (Å)**	**pH**	**ASA ($Å^2$)**
1AKI	0	1.50	4.48	6523
1VDQ	0.1	1.50	-	6518
1GXV	1.3	-	2.00	6092
1HEL	0.3	1.70	-	6662
2EPE	0.4	2.50	4.70	6370
7LYZ	0.4	2.50	-	6623
1VDS	0.4	1.60	-	6571
1BVX	0.4	1.80	4.50	6513
1BWH	0.4	1.80	4.70	6641
1YKZ	0.4	1.80	4.60	6591
1JIS	0.4	1.90	4.60	6646
193L	0.4	1.33	4.30	6337
2LYM	0.4	2.00	-	6804
1HEM	0.4	1.80	-	6576
1LSM	0.4	1.70	-	6482
1RCM A	0.7	1.90	-	6710
1RCM B	0.5	1.90	-	6805
1YL0	0.4	1.90	4.60	6867
1FLU	0.3	1.78	4.70	6563
6LYT	0.4	1.90	-	6602
194L	0.4	1.40	4.30	6363
3A93	0.4	1.55	4.50	6395

Excluding residues with unusually high variation in atomic positioning

XYZ coordinates represented in the structural file merely represent the maximal probability a given atom could be found at the given coordinate rather than the actual position of the atom. An estimate of the uncertainty in the atomic position is provided by the B-factor or the 'atomic thermal factor'. The B-factor from a refinement process is the combined result of the disorder and thermal motions occurring in the protein molecule. While the disorder pertains to atom occupying different position in the crystal, thermal motion refers to vibration of an atom about its equilibrium position. In either case, the higher the B-factor of a residue, the lesser is the probability of finding the residue at the specified XYZ coordinate. Hence residues with unusually high thermal factors should not be included in the analysis. The standard method of removing such discrepancies involves removing segments with a Z score greater than three from the normalized B-factor.

The B-factor of the alpha carbon of an amino acid within a protein was extracted from the PDB coordinates and was computed according to the following formula:

$$B_{zi} = (B_i - \mu)/\sigma, \text{ where,}$$

B_{zi} = Z score of B factor of the alpha carbon at position i.

B_i = B-factor of the alpha carbon at position i

μ = mean of the B-factor of alpha carbon atoms in the chain.

σ = Standard deviation of the B-factor of alpha carbon atoms in the protein chain.

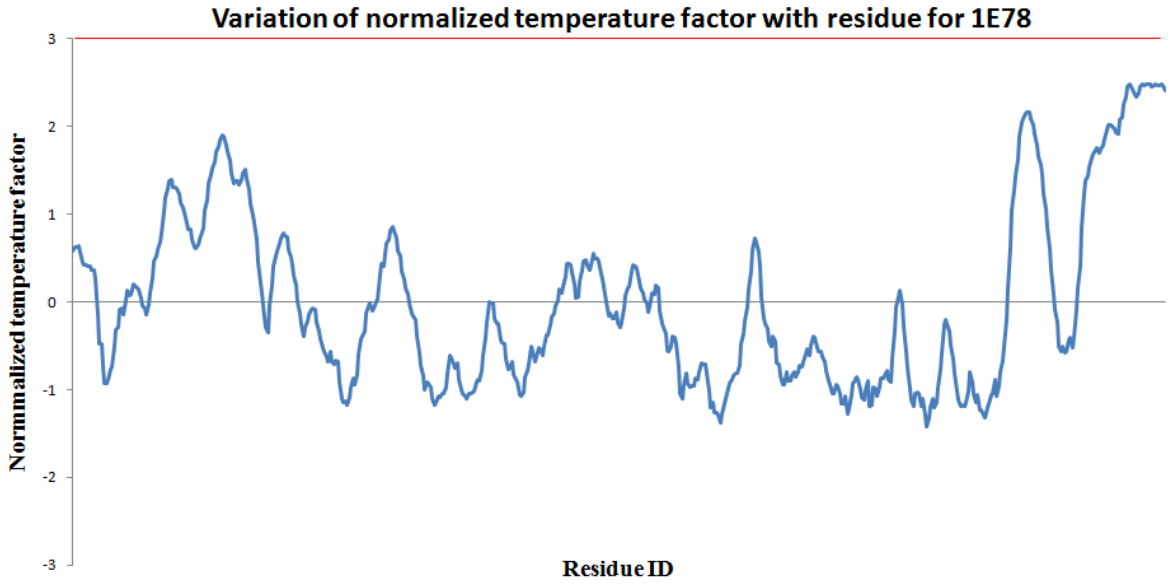
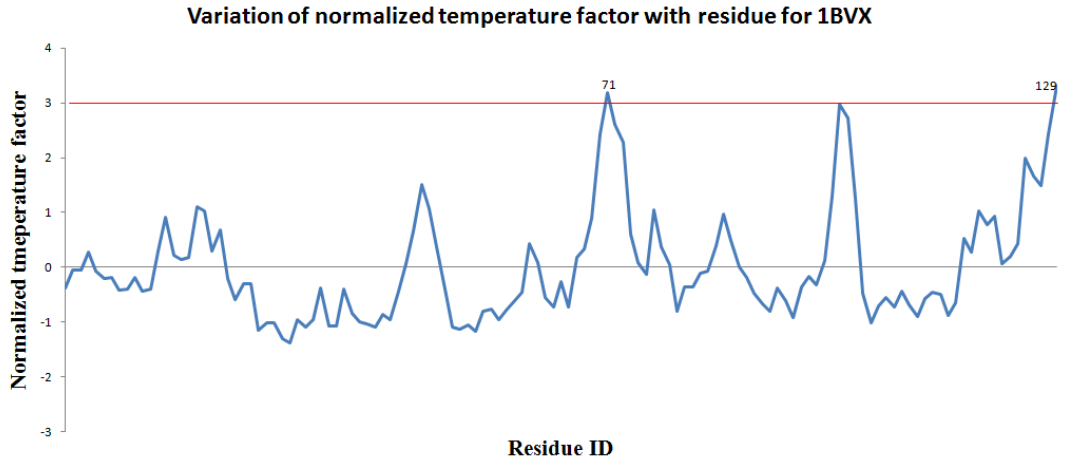


Figure A-3: Normalized temperature factor for HSA. None of the residues in this model have highly uncertain XYZ locations as indicted by the fact that the thermal factor distribution for all atoms is within a Z-score cut-offs at ±3, which are marked in red.



(a)

(b)

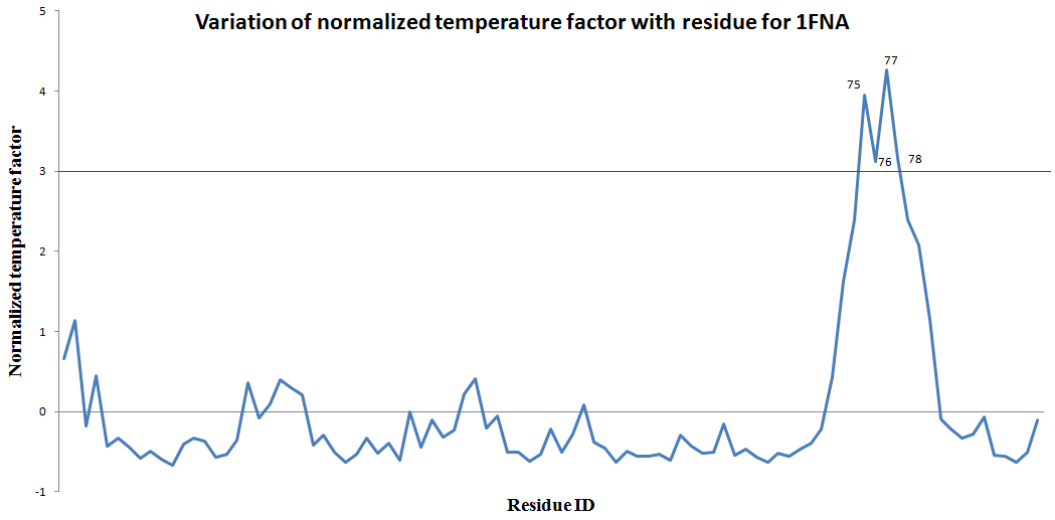


Figure A-4: Normalized temperature factor for (a) HEWL and (b) Fibronectin. In HEWL, two residues (71 and 129), while in the fibronectin model four residues have highly uncertain XYZ locations as indicted by the fact that the thermal factor distribution is higher than the Z-score cut-off at 3, which is marked in red.

Although such a screening approach is important to the predicted orientation of protein and is enabled in the current feature of MATLAB® script, flexibility analysis has not been applied to any of the results described in the present manuscript due to the lack of a comparable data.

Appendix B: Expanded results

Comparison of Topographical Maps for Similar Protein Structures

This section is incorporated in order to demonstrate the application of local positioning and the ability of the topography maps in identifying even the slightest change in the protein structure. HEWL was chosen as the protein of interest as it is one the most highly solved protein structures for which the structural coordinate files are available.

Table B.1: The root mean square deviation (RMSD) of similar HEWL models resolved at varying environmental conditions. Reference structure is 1BVX.

PDB ID	pH	Temperature (K)	Resolution (Å)	RMSD (Å)
1bvx	4.5	-	1.8	0.0
1bwh	4.7	-	1.8	0.1
1ykz	4.6	298	1.8	0.1
1bwj	4.7	293	1.8	0.1
1jis	4.6	293	1.9	0.1
1vds	-	-	1.6	0.1
1z55	4.6	298	1.9	0.1
1lza	-	-	1.6	0.1
1jit	4.6	293	1.9	0.1
1yl0	5.5	298	1.9	0.1
1ykx	4.6	298	1.9	0.1
1yky	4.6	298	1.9	0.1
1jj0	4.6	293	1.9	0.2
1yl1	4.6	298	1.9	0.2

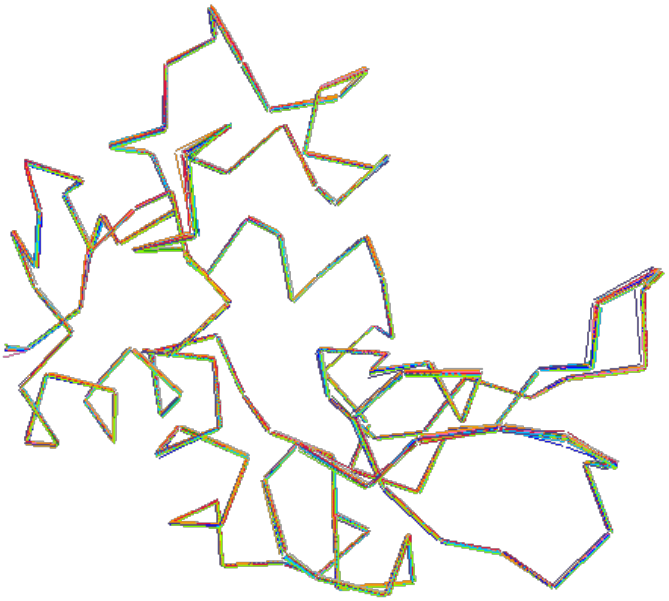
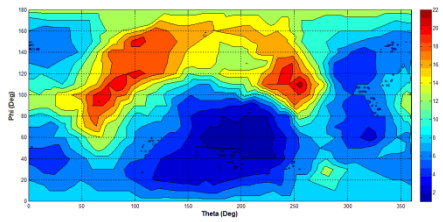
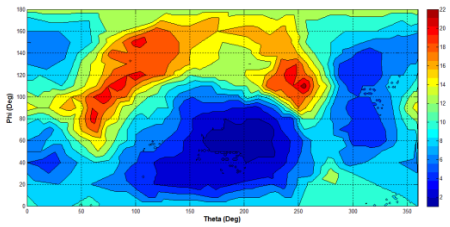
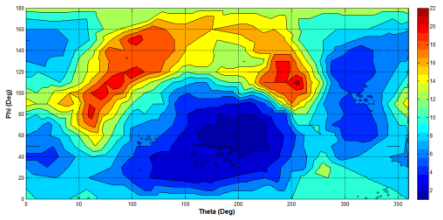
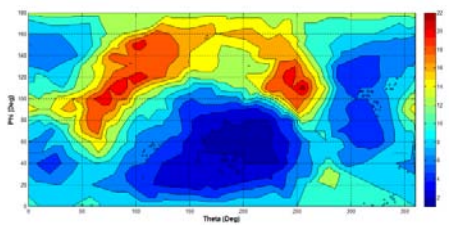
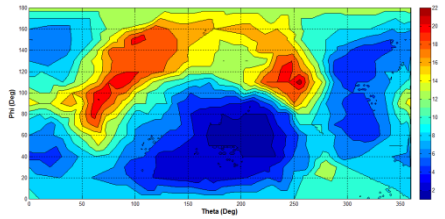
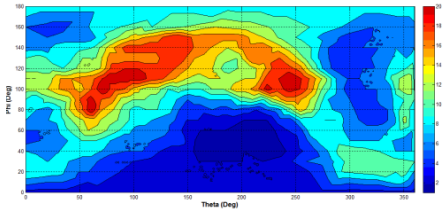
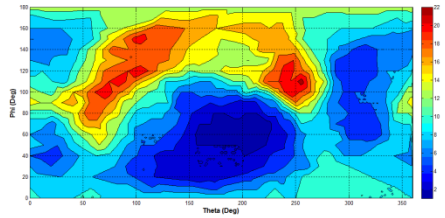
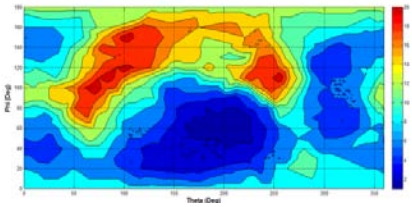
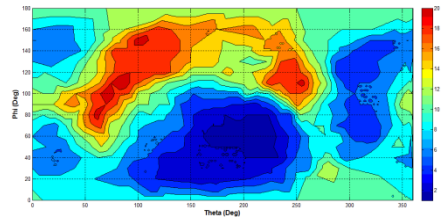
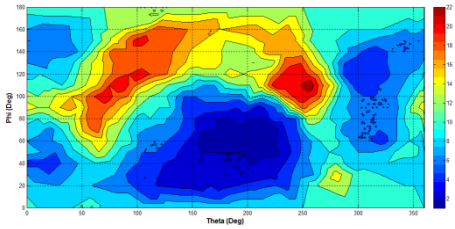
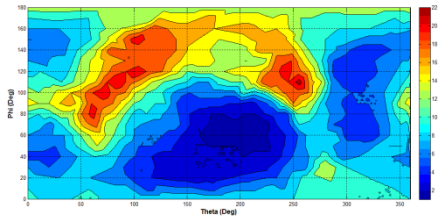
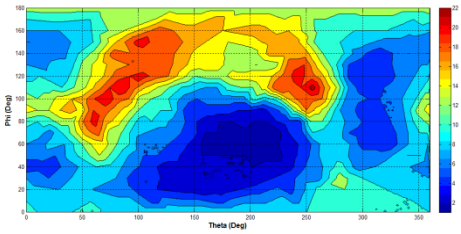
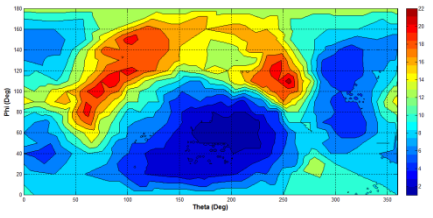
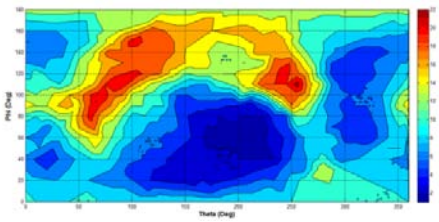


Figure B-1: Superimposed structure of HEWL models described in Table B-1. Image generated in DALI.

Table B.2: Topography map of hydrophobic residues in HEWL models generated at angle of rotation (α) = 10º described in Table B.1

 **1BVX**	 **1BWH**
 **1YKZ**	 **1BWJ**

 **1JIS**	 **1VDS**
 **1Z55**	 **1LZA**
 **1JIT**	 **1YL0**
 **1YKX**	 **1YKY**
 **1JJ0**	 **1YL1**

Comparison of Topographical Maps for Dissimilar Protein Models

Table B.3: The root mean square deviation (RMSD) of dissimilar HEWL models resolved at varying environmental conditions. Reference structure is 1BVX.

PDB ID	RMSD (Å)	pH	Temperature (K)	Resolution (Å)
1bvx	0.0	4.5	-	1.8
1rcm (A)	0.4	-	-	1.9
1rcm (B)	0.6	-	-	1.9
1sq2 (L)	0.9	7	295	1.45
1v7t (B)	0.8	4.5	263	1.13

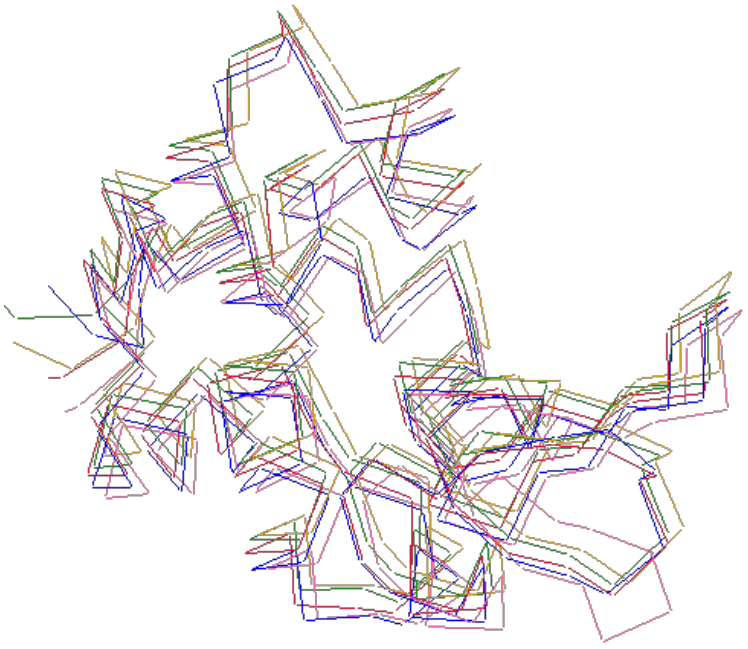
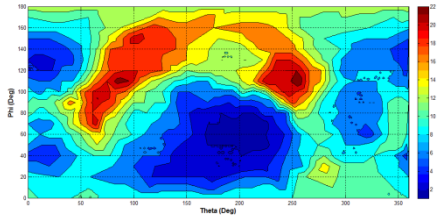
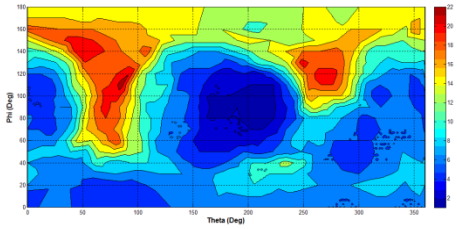
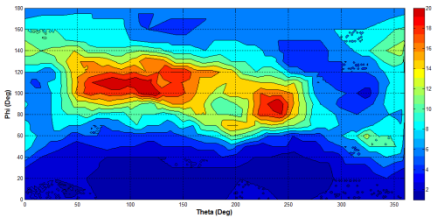
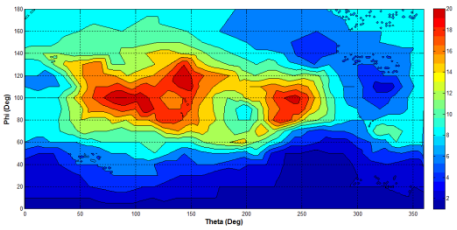


Figure B-2: Superimposed structure of HEWL models described in Table B-2. Image generated in DALI[49].

Table B.4: Topography map of hydrophobic residues in HEWL models described in Table B.3

 **1RCM (A)**	 **1RCM (B)**
 **1SQ2 (L)**	 **1V7T (B)**

Effect of Angle of Rotation (α) on Topography Map and Predicted Orientation

The angle of rotation (α) determines the grid points within the topography maps that correspond to the distribution of residues interacting with the surface. The closer that the grid points are, the better and more accurate is the interpolation for generating the topographical map. Hence smaller 'α' is expected to affect the quality of the topography map generated for the current study, but extremely small 'α' is prohibitory as the run time involved in the configurational space search would be very large. The typical value of α is 10º with values such as 15º being used for extremely small proteins such as CYTC.

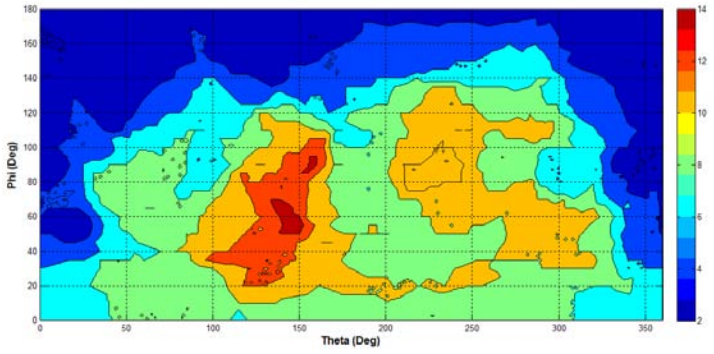
In order to demonstrate the effect of α on protein, a comparison was done on (a) cytochrome c (1HRC), (b) hen egg white lysozyme (1BVX), (c) human serum albumin (1E78), and (d) *Aspergillus niger* glucose oxidase (1CF3) for α values of 5º, 10º, 15º (Table B-5). Based on these results it could be noticed that the residue distributions are significantly different at higher and lower angles of rotation. Since the current method employs a rigid model with no other mode of interactions involved, the topography maps at lower 'α' are very important for the accurate prediction of the preferred orientation, resulting in the selection of α = 5º for the subsequent studies.

Table B.5: The effect of 'α' on different proteins with different molecular sizes.

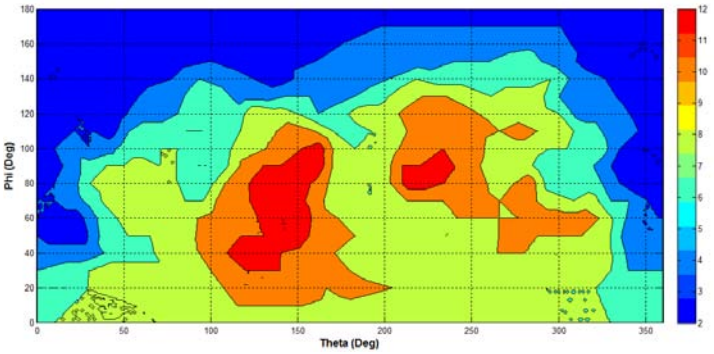
Protein type	**PDB model**	**Resolution**	**Molecular weight (kDa)**	**Approx. radii (nm)**	**Figure**
CYTC	1HRC	1.90	11.8	1.67	B-3
HEWL	1BVX	1.80	14.3	1.78	B-4
HSA	1E78	2.60	66.5	2.97	B-5
GOx	1CF3	1.90	63.3	2.93	B-6

Cytochrome C (1HRC)

(a)



(b)



(c)

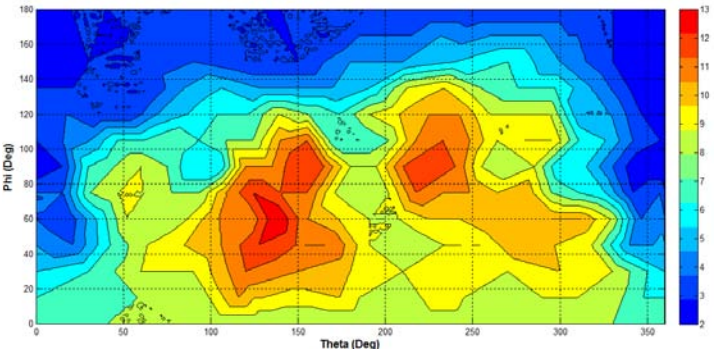
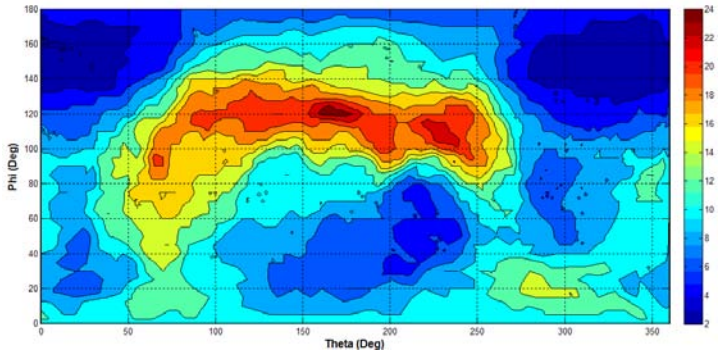


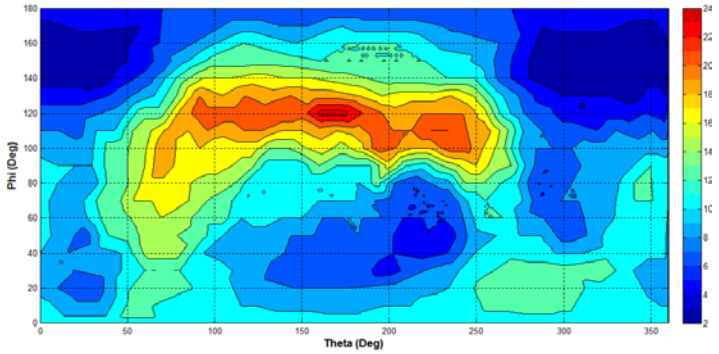
Figure B-3: Distribution of hydrophilic residues in CYTC at different angles of rotation: (a) 5º, (b) 10º, and (c) 15º at 'D' of 10 Å and residue exclusion of relative ASA less than 10%.

Hen egg white lysozyme (HEWL)

(a)



(b)



(c)

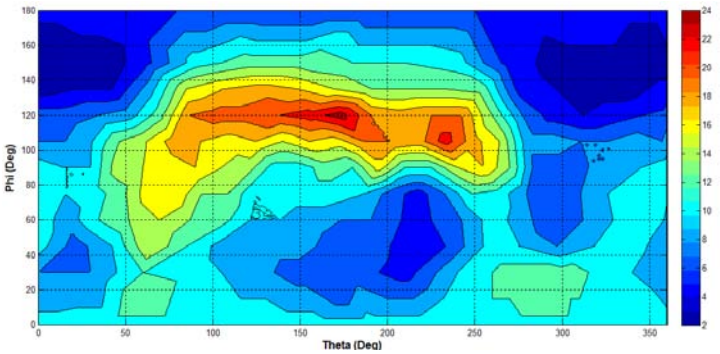


Figure B-4: Distribution of hydrophilic residues in HEWL at different angle of rotations: (a) 5º, (b) 10º, and (c) 15º at 'D' of 10 Å and residue exclusion of relative ASA less than 10%.

Aspergillus niger glucose oxidase (GOx)

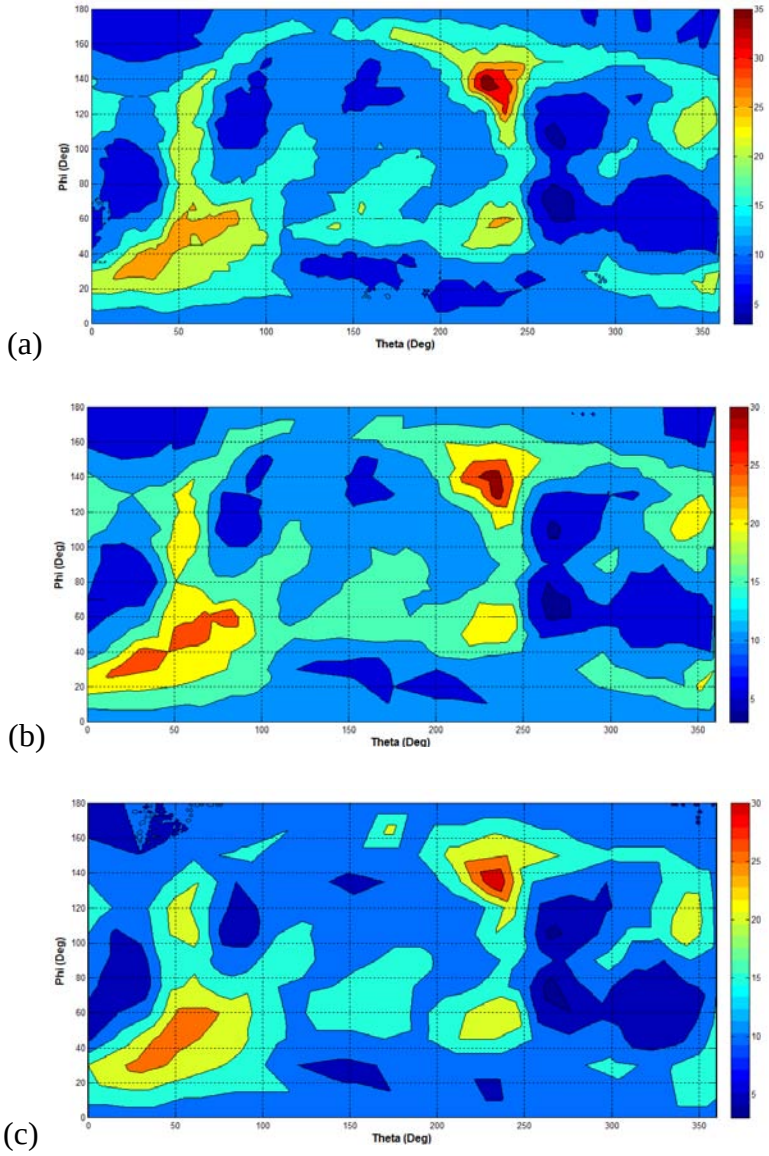
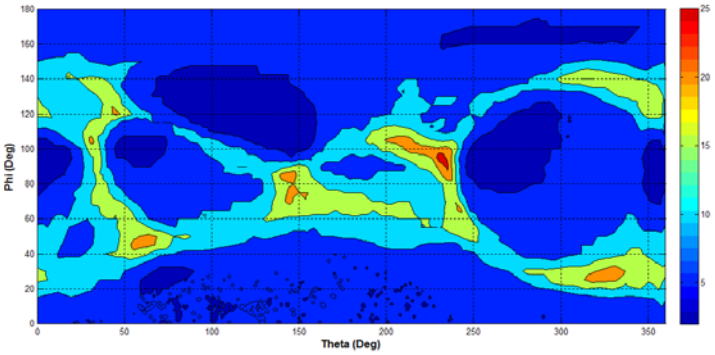


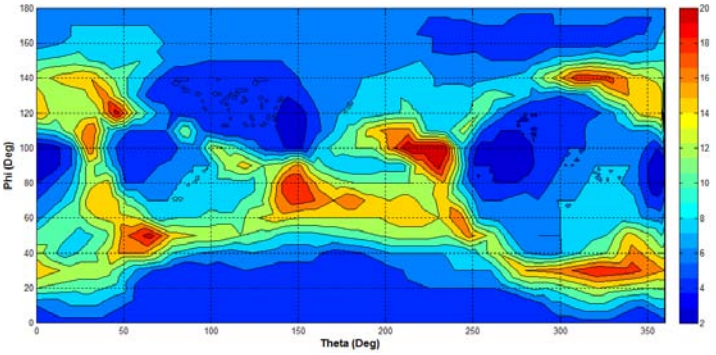
Figure B-5: Distribution of hydrophilic residues in GOx at different angle of rotations: (a) 5º, (b) 10º, and (c) 15º at 'D' of 10 Å and residue exclusion of relative ASA less than 10%.

Human serum albumin (HSA)

(a)



(b)



(c)

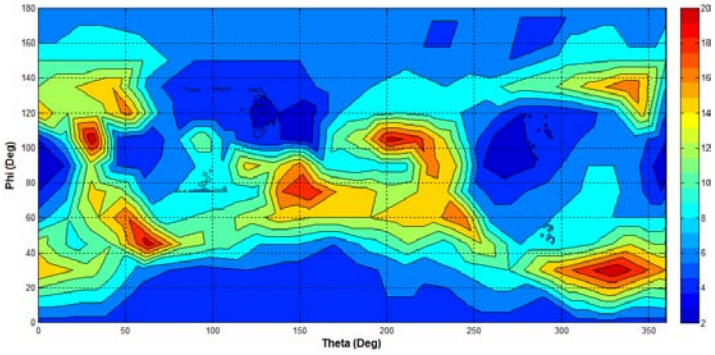
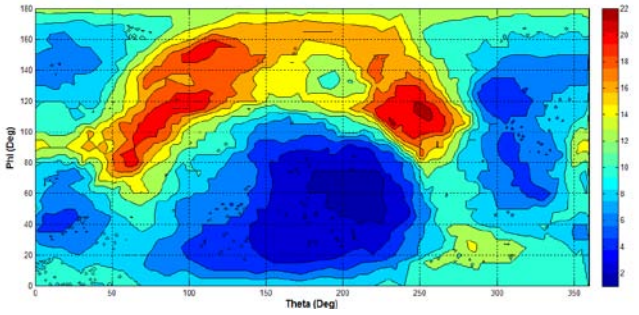


Figure B-6: Distribution of hydrophilic residues in HSA at different angle of rotations: (a) 5º, (b) 10º, and (c) 15º at 'D' of 10 Å and residue exclusion of relative ASA less than 10%.

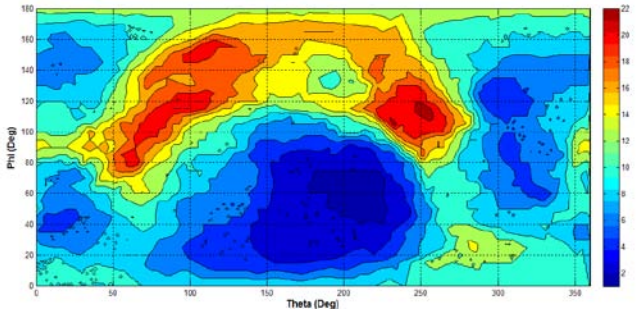
Effect of Residue Exclusion on Topography Map and Predicted Orientation

As discussed in chapter 4 of the main manuscript, solvent accessibility of the amino acid is a critical parameter in mediating the interactions of an amino acid with the surface. Exclusion of buried residues from being quantified is thus important for the accurate prediction of the amino acid and would thus affect the topography map generated for a given protein generated for a given interaction distance cut-off. The default setting for defining a particular residue as buried is the relative accessibility of less than 10%, but the program is flexible to accept any user-defined input, as there is a no-defined criterion for the selection of such cut-off (Fig B-7). However, the best result was obtained with a relative accessibility cut-off of 10% and this value was used for all subsequent studies.

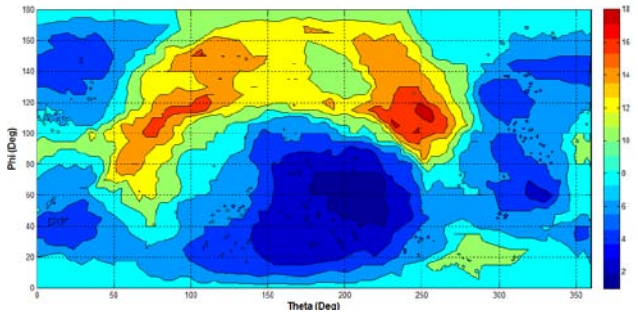
(a)



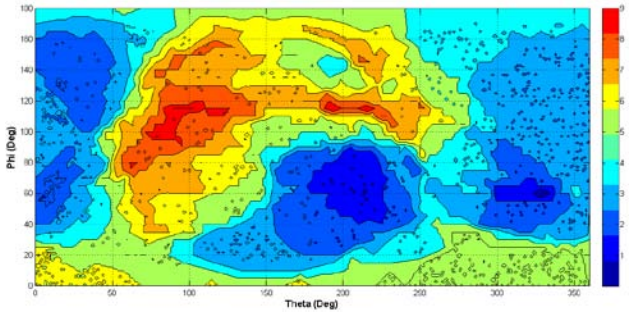
(b)



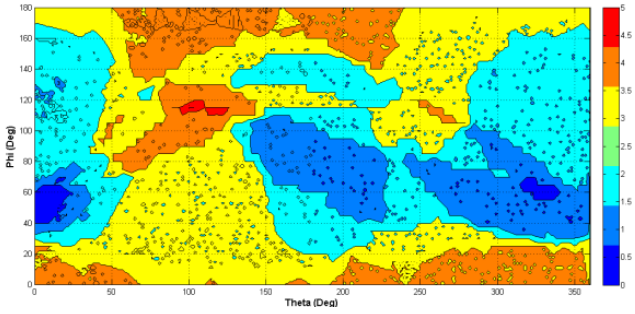
(c)



(d)



(e)



(f)

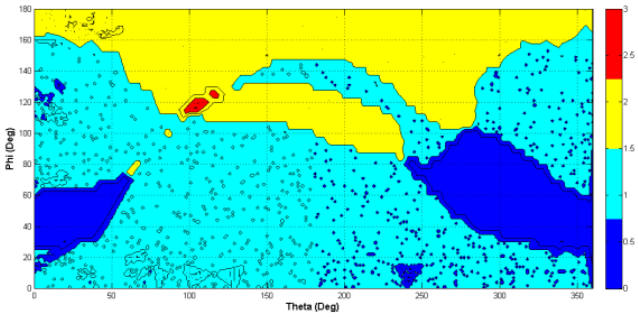
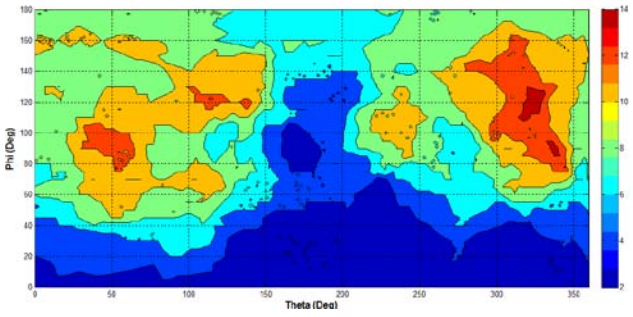


Figure B-7: Distribution of hydrophobic residues at different relative ASA in HEWL (a) no exclusion, (b) 10%, (c) 15%, (d) 20%, (e) 25%, and (f) 35% at 'D' of 10 Å and angle of rotation = 5º.

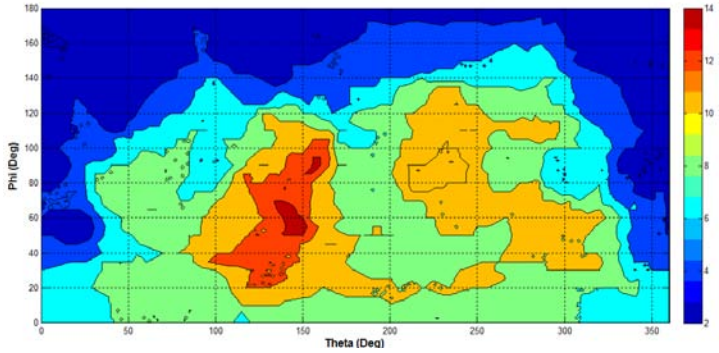
Effect of Interaction Distance Cut-Off (D) on Topography Map and Predicted Orientation

The interaction distance cut-off is an approximation of the potential of mean force profile acting on the amino acid constituting the protein where the interaction with the surface is considered to reach a minimum of energy. The quantification of the residues at a user-defined 'D' followed by the selection of the face of the protein corresponding to the maximum residue concentration could thus indicate the most favorable orientation with which the protein would adsorb to the surface.

'D' is found to vary with the functional groups presented by the surface to the peptide such as SAMs with CH_3, C_6H_5OH, and $NHCOCH_3$ functional groups as indicated by Vellore et al.[73] Hence, for the accurate prediction of the preferred orientation, a proper approximation of 'D' would thus be required. Theoretically it is expected that the residues closer to the interface would interact more favorably than the residues which are positioned far from the protein surface, hence in the event of multiple sites presenting residues with maximum concentration, sites corresponding to residues closest to the surface which have maximum relative solvent accessibility are ideal for the identification of preferred orientations than those far from the surface.

(a) 

(b)



(c)

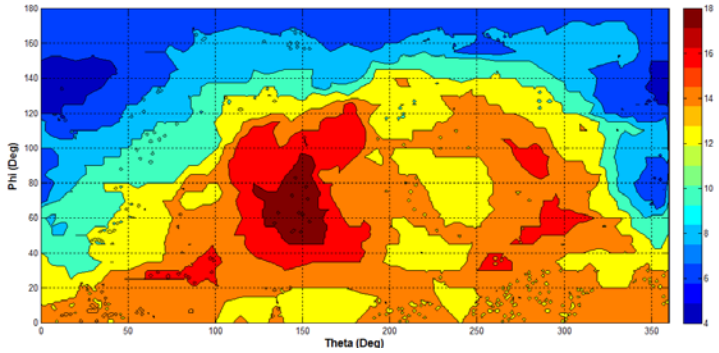


Figure B-8: Distribution of hydrophilic residues at different relative 'D' in CYTC: (a) 5Å, (b) 10 Å, and (c) 15 Å at an angle of rotation of 5° and relative accessibility cut-off of 10%.

In ideal surfaces mimicking CH_3, OH, NH_2, and COOH SAMs, the minimum energy of interaction was found at an average 'D' of 10 Å and this value of D was thus chosen as the default setting. However, the value of D can be varied according to user-input.

References

1. RATNER, B.D. BIOMATERIALS SCIENCE: AN INTERDISCIPLINARY ENDEAVOR. BIOMATERIALS SCIENCE: AN INTRODUCTION TO MATERIALS IN MEDICINE, 1-8 (1996).

2. BRASH, J.L. & HORBETT, T.A. ACS AMERICAN CHEMICAL SOCITEY SYMPOSIUM SERIES 343. PROTEINS AT INTERFACES PHYSICOCHEMICAL AND BIOCHEMICAL STUDIES 192ND MEETING OF THE AMERICAN CHEMICAL SOCIETY ANAHEIM CALIFORNIA USA (1986).

3. GORBET, M.B. & SEFTON, M.V. BIOMATERIAL-ASSOCIATED THROMBOSIS: ROLES OF COAGULATION FACTORS, COMPLEMENT, PLATELETS AND LEUKOCYTES. BIOMATERIALS **25**, 5681-5703 (2004).

4. LATOUR, R.A. BIOMATERIALS: PROTEIN–SURFACE INTERACTIONS. IN THE ENCYCLOPEDIA OF BIOMATERIALS AND BIOENGINEERING, 2ND EDITION, G.E. WNEK AND G.L. BOWLIN (EDS.), INFORMA HEALTHCARE, VOL. 1, PP. 270-284, 2008.

5. TZANOV, T., ANDREAUS, J., GUEBITZ, G. & CAVACO-PAULO, A. PROTEIN INTERACTIONS IN ENZYMATIC PROCESSES IN TEXTILES. ELECTRONIC JOURNAL OF BIOTECHNOLOGY **6**, 313-321 (2003).

6. NAKANISHI, K., SAKIYAMA, T. & IMAMURA, K. ON THE ADSORPTION OF PROTEINS ON SOLID SURFACES, A COMMON BUT VERY COMPLICATED PHENOMENON. JOURNAL OF BIOSCIENCE AND BIOENGINEERING **91**, 233-244 (2001).

7. KIM, D.C. & KANG, D.J. MOLECULAR RECOGNITION AND SPECIFIC INTERACTIONS FOR BIOSENSING APPLICATIONS. SENSORS **8**, 6605-6641 (2008).

8. COSNIER, S. BIOMOLECULE IMMOBILIZATION ON ELECTRODE SURFACES BY ENTRAPMENT OR ATTACHMENT TO ELECTROCHEMICALLY POLYMERIZED FILMS. A REVIEW. BIOSENSORS & BIOELECTRONICS **14**, 443-456 (1999).

9. OSCARSSON, S. FACTORS AFFECTING PROTEIN INTERACTION AT SORBENT INTERFACES. JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL SCIENCES AND APPLICATIONS **699**, 117-131 (1997).

10. TZANNIS, S.T., HRUSHESKY, W.J.M., WOOD, P.A. & PRZYBYCIEN, T.M. ADSORPTION OF A FORMULATED PROTEIN ON A DRUG DELIVERY DEVICE SURFACE. JOURNAL OF COLLOID AND INTERFACE SCIENCE **189**, 216-228 (1997).

11. RAMACHANDRAN, N., LARSON, D.N., STARK, P.R.H., HAINSWORTH, E. & LABAER, J. EMERGING TOOLS FOR REAL-TIME LABEL-FREE DETECTION OF INTERACTIONS ON FUNCTIONAL PROTEIN MICROARRAYS. FEBS JOURNAL **272**, 5412-5425 (2005).

12. TAYLOR, S. ET AL. SYNTHESIS AND CHARACTERIZATION OF PEPTIDE-FUNCTIONALIZED POLYMERIC NANOPARTICLES. BIOMACROMOLECULES **5**, 245-248 (2004).

13. LATOUR, R.A. MOLECULAR SIMULATION OF PROTEIN-SURFACE INTERACTIONS: BENEFITS, PROBLEMS, SOLUTIONS, AND FUTURE DIRECTIONS (REVIEW). BIOINTERPHASES **3**, FC2-FC12 (2008).

14. DALY, S.M., PRZYBYCIEN, T.M. & TILTON, R.D. COVERAGE-DEPENDENT ORIENTATION OF LYSOZYME ADSORBED ON SILICA. LANGMUIR **19**, 3848-3857 (2003).

15. TILTON, R.D., ROBERTSON, C.R. & GAST, A.P. MANIPULATION OF HYDROPHOBIC INTERACTIONS IN PROTEIN ADSORPTION. LANGMUIR **7**, 2710-2718 (1991).

16. SIVARAMAN, B. & LATOUR, R.A. THE RELATIONSHIP BETWEEN PLATELET ADHESION ON SURFACES AND THE STRUCTURE VERSUS THE AMOUNT OF ADSORBED FIBRINOGEN. BIOMATERIALS **31**, 832-839 (2010).

17. VOROS, J. THE DENSITY AND REFRACTIVE INDEX OF ADSORBING PROTEIN LAYERS. BIOPHYSICAL JOURNAL **87**, 553-561 (2004).

18. LATOUR, R.A. MOLECULAR MODELING OF BIOMATERIAL SURFACES. CURRENT OPINION IN SOLID STATE & MATERIALS SCIENCE **4**, 413-417 (1999).

19. LATOUR, R.A., JR. & HENCH, L.L. A THEORETICAL ANALYSIS OF THE THERMODYNAMIC CONTRIBUTIONS FOR THE ADSORPTION OF INDIVIDUAL PROTEIN RESIDUES ON FUNCTIONALIZED SURFACES. BIOMATERIALS **23**, 4633-4648 (2002).

20. RAUT, V.P., AGASHE, M.A., STUART, S.J. & LATOUR, R.A. MOLECULAR DYNAMICS SIMULATIONS OF PEPTIDE-SURFACE INTERACTIONS. LANGMUIR **21**, 1629-1639 (2005).

21. SUN, Y., WELSH, W.J. & LATOUR, R.A. PREDICTION OF THE ORIENTATIONS OF ADSORBED PROTEIN USING AN EMPIRICAL ENERGY FUNCTION WITH IMPLICIT SOLVATION. LANGMUIR : THE ACS JOURNAL OF SURFACES AND COLLOIDS **21**, 5616-5626 (2005).

22. LATOUR, R.A. MOLECULAR SIMULATION OF PROTEIN-SURFACE INTERACTIONS. BIOLOGICAL INTERACTIONS ON MATERIALS SURFACES: UNDERSTANDING AND CONTROLLING PROTEIN, CELL, AND TISSUE RESPONSES, 69-95 (2009).

23. AGASHE, M., RAUT, V., STUART, S.J. & LATOUR, R.A. MOLECULAR SIMULATION TO CHARACTERIZE THE ADSORPTION BEHAVIOR OF A FIBRINOGEN GAMMA-CHAIN FRAGMENT. LANGMUIR : THE ACS JOURNAL OF SURFACES AND COLLOIDS **21**, 1103-1117 (2005).

24. MASICA, D., MAKRODIMITRIS, K., GLIFORT, E. & GRAY, J. DESIGN OF A PROTEIN-SOLID SURFACE INTERFACE. BIOPHYSICAL JOURNAL, 218A-218A (2007).

25. HAYNES, C.A. & NORDE, W. STRUCTURES AND STABILITIES OF ADSORBED PROTEINS. JOURNAL OF COLLOID AND INTERFACE SCIENCE **169**, 313-328 (1995).

26. MALMSTEN, M. FORMATION OF ADSORBED PROTEIN LAYERS. JOURNAL OF COLLOID AND INTERFACE SCIENCE **207**, 186-199 (1998).

27. MACRITCHIE, F. PROTEINS AT INTERFACES. ANFINSEN, C. B., JOHN T. EDSALL AND FREDERIC M. RICHARDS (ED.). ADVANCES IN PROTEIN CHEMISTRY, VOL. 32. VII+347P. ILLUS. ACADEMIC PRESS, INC.: NEW YORK, N.Y., USA; LONDON, ENGLAND. ISBN 0-12-034232-4, 283-326 (1978).

28. CHOTHIA, C. PRINCIPLES THAT DETERMINE THE STRUCTURE OF PROTEINS. ANNUAL REVIEW OF BIOCHEMISTRY **53**, 537-572 (1984).

29. DILL, K.A. DOMINANT FORCES IN PROTEIN FOLDING. BIOCHEMISTRY **29**, 7133-7155 (1990).

30. RICHARDS, F.M. AREAS, VOLUMES, PACKING, AND PROTEIN-STRUCTURE. ANNUAL REVIEW OF BIOPHYSICS AND BIOENGINEERING **6**, 151-176 (1977).

31. WEI, Y. & LATOUR, R.A. DETERMINATION OF THE ADSORPTION FREE ENERGY FOR PEPTIDE−SURFACE INTERACTIONS BY SPR SPECTROSCOPY. LANGMUIR **24**, 6721-6729 (2008).

32 DZUBIELLA, J., SWANSON, J.M.J. & MCCAMMON, J.A. COUPLING HYDROPHOBICITY, DISPERSION, AND ELECTROSTATICS IN CONTINUUM SOLVENT MODELS. PHYSICAL REVIEW LETTERS **96** (2006).

33. GIOVAMBATTISTA, N., LOPEZ, C.F., ROSSKY, P.J. & DEBENEDETTI, P.G. HYDROPHOBICITY OF PROTEIN SURFACES: SEPARATING GEOMETRY FROM CHEMISTRY. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA **105**, 2274-2279 (2008).

34. HORINEK, D. ET AL. MOLECULAR HYDROPHOBIC ATTRACTION AND ION-SPECIFIC EFFECTS STUDIED BY MOLECULAR DYNAMICS. LANGMUIR : THE ACS JOURNAL OF SURFACES AND COLLOIDS **24**, 1271-1283 (2008).

35. TALASAZ, A.H. ET AL. PREDICTION OF PROTEIN ORIENTATION UPON IMMOBILIZATION ON BIOLOGICAL AND NONBIOLOGICAL SURFACES. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA **103**, 14773-14778 (2006).

36. SPEROTTO, M.M., MAY, S. & BAUMGAERTNER, A. MODELLING OF PROTEINS IN MEMBRANES. CHEMISTRY AND PHYSICS OF LIPIDS **141**, 2-29 (2006).

37. HAO-JEN, H., SHEH-YI, S. & RUEY-YUG, T. PREFERRED ORIENTATION OF ALBUMIN ADSORPTION ON A HYDROPHILIC SURFACE FROM MOLECULAR SIMULATION. COLLOIDS AND SURFACES B: BIOINTERFACES, 183-191 (2008).

38. JIAN, Z., JIE, Z. & SHAOYI, J. MOLECULAR SIMULATION STUDIES OF THE ORIENTATION AND CONFORMATION OF CYTOCHROME C ADSORBED ON SELF-ASSEMBLED MONOLAYERS. JOURNAL OF PHYSICAL CHEMISTRY B **108**, 17418-17424 (2004).

39. HAGIWARA, T., SAKIYAMA, T. & WATANABE, H. MOLECULAR SIMULATION OF BOVINE BETA-LACTOGLOBULIN ADSORBED ONTO A POSITIVELY CHARGED SOLID SURFACE. LANGMUIR **25**, 226-234 (2009).

40. BOURNE, P.E. & WEISSIG, H. STRUCTURAL BIOINFORMATICS. (WILEY-LISS, HOBOKEN, N.J.; 2003).

41. BERMAN, H.M. ET AL. THE PROTEIN DATA BANK. NUCLEIC ACIDS RESEARCH **28**, 235-242 (2000).

42. DESHPANDE, N. ET AL. THE RCSB PROTEIN DATA BANK: A REDESIGNED QUERY SYSTEM AND RELATIONAL DATABASE BASED ON THE MMCIF SCHEMA. NUCLEIC ACIDS RESEARCH **33**, D233-D237 (2005).

43. ALTMAN, R. & DUGAN, J. IN STRUCTURAL BIOINFORMATICS. (ED. P.E. BOURNE, AND GU, J.) (HOBOKEN, N.J. : WILEY-BLACKWELL, 2009).

44. WILKINS, M.R. ET AL. PROTEIN IDENTIFICATION AND ANALYSIS TOOLS IN THE EXPASY SERVER. 2-D PROTEOME ANALYSIS PROTOCOLS **112**, 531-552 (1999).

45. HOOFT, R.W.W., VRIEND, G., SANDER, C. & ABOLA, E.E. ERRORS IN PROTEIN STRUCTURES. NATURE **381**, 272-272 (1996).

46. VRIEND, G. WHAT IF - A MOLECULAR MODELING AND DRUG DESIGN PROGRAM. JOURNAL OF MOLECULAR GRAPHICS **8**, 52-& (1990).

47. LASKOWSKI, R.A., MACARTHUR, M.W., MOSS, D.S. & THORNTON, J.M. PROCHECK - A PROGRAM TO CHECK THE STEREOCHEMICAL

QUALITY OF PROTEIN STRUCTURES. JOURNAL OF APPLIED CRYSTALLOGRAPHY **26**, 283-291 (1993).

48. EISENBERG, D., LUTHY, R. & BOWIE, J.U. VERIFY3D: ASSESSMENT OF PROTEIN MODELS WITH THREE-DIMENSIONAL PROFILES. MACROMOLECULAR CRYSTALLOGRAPHY, PT B **277**, 396-404 (1997).

49. HOLM, L. & SANDER, C. DALI - A NETWORK TOOL FOR PROTEIN-STRUCTURE COMPARISON. TRENDS IN BIOCHEMICAL SCIENCES **20**, 478-480 (1995).

50. KOLBECK, B., MAY, P., SCHMIDT-GOENNER, T., STEINKE, T. & KNAPP, E.W. CONNECTIVITY INDEPENDENT PROTEIN-STRUCTURE ALIGNMENT: A HIERARCHICAL APPROACH. BMC BIOINFORMATICS **7** (2006).

51. ORTIZ, A.R., STRAUSS, C.E.M. & OLMEA, O. MAMMOTH (MATCHING MOLECULAR MODELS OBTAINED FROM THEORY): AN AUTOMATED METHOD FOR MODEL COMPARISON. PROTEIN SCIENCE **11**, 2606-2621 (2002).

52. CHOU, P.Y. & FASMAN, G.D. PREDICTION OF THE SECONDARY STRUCTURE OF PROTEINS FROM THEIR AMINO-ACID SEQUENCE. MEISTER, ALTON (ED.). ADVANCES IN ENZYMOLOGY AND RELATED AREAS OF MOLECULAR BIOLOGY, VOL. 47. V+499P. ILLUS. JOHN WILEY AND SONS, INC.: NEW YORK, N.Y., USA; CHICHESTER, ENGLAND. ISBN 0-471-04116-5, 45-148 (1978).

53. KABSCH, W. & SANDER, C. DICTIONARY OF PROTEIN SECONDARY STRUCTURE - PATTERN-RECOGNITION OF HYDROGEN-BONDED AND GEOMETRICAL FEATURES. BIOPOLYMERS **22**, 2577-2637 (1983).

54. GARNIER, J., GIBRAT, J.F. & ROBSON, B. GOR METHOD FOR PREDICTING PROTEIN SECONDARY STRUCTURE FROM AMINO ACID SEQUENCE. COMPUTER METHODS FOR MACROMOLECULAR SEQUENCE ANALYSIS **266**, 540-553 (1996).

55. PTITSYN, O.B. & FINKELSTEIN, A.V. PREDICTION OF PROTEIN SECONDARY STRUCTURE BASED ON PHYSICAL THEORY - HISTONES. PROTEIN ENGINEERING **2**, 443-447 (1989).

56. SCHNEIDER, G. & WREDE, P. PREDICTION OF THE SECONDARY STRUCTURE OF PROTEINS FROM THE AMINO-ACID-SEQUENCE WITH ARTIFICIAL NEURAL NETWORKS. ANGEWANDTE CHEMIE-INTERNATIONAL EDITION IN ENGLISH **32**, 1141-1143 (1993).

57. MCGUFFIN, L.J., BRYSON, K. & JONES, D.T. THE PSIPRED PROTEIN STRUCTURE PREDICTION SERVER. BIOINFORMATICS **16**, 404-405 (2000).

58. ROST, B., SANDER, C. & SCHNEIDER, R. PHD - AN AUTOMATIC MAIL SERVER FOR PROTEIN SECONDARY STRUCTURE PREDICTION. COMPUTER APPLICATIONS IN THE BIOSCIENCES **10**, 53-60 (1994).

59. WODAK, S.J. & JANIN, J. ANALYTICAL APPROXIMATION TO THE ACCESSIBLE SURFACE-AREA OF PROTEINS. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA-PHYSICAL SCIENCES **77**, 1736-1740 (1980).

60 LEE, B. & RICHARDS, F.M. THE INTERPRETATION OF PROTEIN STRUCTURES: ESTIMATION OF STATIC ACCESSIBILITY. JOURNAL OF MOLECULAR BIOLOGY **55**, 379-400, IN373-IN374 (1971).

61. AHMAD, S., GROMIHA, M., FAWAREH, H. & SARAI, A. ASAVIEW: DATABASE AND TOOL FOR SOLVENT ACCESSIBILITY REPRESENTATION IN PROTEINS. BMC BIOINFORMATICS **5** (2004).

62. FRACZKIEWICZ, R. & BRAUN, W. EXACT AND EFFICIENT ANALYTICAL CALCULATION OF THE ACCESSIBLE SURFACE AREAS AND THEIR GRADIENTS FOR MACROMOLECULES. JOURNAL OF COMPUTATIONAL CHEMISTRY **19**, 319-333 (1998).

63. HIGA, R.H. ET AL. STING MILLENNIUM SUITE: INTEGRATED SOFTWARE FOR EXTENSIVE ANALYSES OF 3D STRUCTURES OF PROTEINS AND THEIR COMPLEXES. BMC BIOINFORMATICS **5** (2004).

64. LI, H., ROBERTSON, A.D. & JENSEN, J.H. VERY FAST EMPIRICAL PREDICTION AND RATIONALIZATION OF PROTEIN PK(A) VALUES. PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS **61**, 704-721 (2005).

65. GUEX, N. & PEITSCH, M.C. SWISS-MODEL AND THE SWISS-PDBVIEWER: AN ENVIRONMENT FOR COMPARATIVE PROTEIN MODELING. ELECTROPHORESIS **18**, 2714-2723 (1997).

66. CASE, D.A. ET AL. THE AMBER BIOMOLECULAR SIMULATION PROGRAMS. JOURNAL OF COMPUTATIONAL CHEMISTRY **26**, 1668-1688 (2005).

67. BROOKS, B.R. ET AL. CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS. JOURNAL OF COMPUTATIONAL CHEMISTRY **4**, 187-217 (1983).

68. SCHWIETERS, C.D., KUSZEWSKI, J.J., TJANDRA, N. & CLORE, G.M. THE XPLOR-NIH NMR MOLECULAR STRUCTURE DETERMINATION PACKAGE. JOURNAL OF MAGNETIC RESONANCE **160**, 65-73 (2003).

69. GU, J. & BOURNE, P.E. STRUCTURAL BIOINFORMATICS. (WILEY-BLACKWELL, HOBOKEN, N.J.; (2009).

70. DAMODARAN, S. & RAZUMOVSKY, L. ROLE OF SURFACE AREA-TO-VOLUME RATIO IN PROTEIN ADSORPTION AT THE AIR-WATER INTERFACE. SURFACE SCIENCE **602**, 307-315 (2008).

71. TSAI, J., TAYLOR, R., CHOTHIA, C. & GERSTEIN, M. THE PACKING DENSITY IN PROTEINS: STANDARD RADII AND VOLUMES. JOURNAL OF MOLECULAR BIOLOGY **290**, 253-266 (1999).

72. NOVOTNY, M., SEIBERT, M. & KLEYWEGT, G.J. ON THE PRECISION OF CALCULATED SOLVENT-ACCESSIBLE SURFACE AREAS. ACTA CRYSTALLOGRAPHICA SECTION D-BIOLOGICAL CRYSTALLOGRAPHY **63**, 270-274 (2007).

73. KUBIAK, K. & MULHERAN, P.A. MOLECULAR DYNAMICS SIMULATIONS OF HEN EGG WHITE LYSOZYME ADSORPTION AT A CHARGED SOLID SURFACE. JOURNAL OF PHYSICAL CHEMISTRY B **113**, 12189-12200 (2009).

74. VELLORE, N.A., YANCEY, J.A., COLLIER, G., LATOUR, R.A. & STUART, S.J. ASSESSMENT OF THE TRANSFERABILITY OF A PROTEIN FORCE FIELD FOR THE SIMULATION OF PEPTIDE-SURFACE INTERACTIONS. LANGMUIR (2010).

75. SIVARAMAN, B. & LATOUR, R.A. THE ADHERENCE OF PLATELETS TO ADSORBED ALBUMIN BY RECEPTOR-MEDIATED RECOGNITION OF BINDING SITES EXPOSED BY ADSORPTION-INDUCED UNFOLDING. BIOMATERIALS **31**, 1036-1044 (2010).

76. FEARS, K.P., SIVARAMAN, B., POWELL, G.L., WU, Y. & LATOUR, R.A. PROBING THE CONFORMATION AND ORIENTATION OF ADSORBED ENZYMES USING SIDE-CHAIN MODIFICATION. LANGMUIR **25**, 9319-9327 (2009).

77. SUN, Y., WELSH, W.J. & LATOUR, R.A. PREDICTION OF THE ORIENTATIONS OF ADSORBED PROTEIN USING AN EMPIRICAL ENERGY FUNCTION WITH IMPLICIT SOLVATION. LANGMUIR **21**, 5616-5626 (2005).

78. WERTZ, C.F. & SANTORE, M.M. ADSORPTION AND REORIENTATION KINETICS OF LYSOZYME ON HYDROPHOBIC SURFACES. LANGMUIR **18**, 1190-1199 (2002).

79. TRZASKOWSKI, B., LEONARSKI, F., LES, A. & ADAMOWICZ, L. ALTERING THE ORIENTATION OF PROTEINS ON SELF-ASSEMBLED MONOLAYERS: A COMPUTATIONAL STUDY. BIOMACROMOLECULES **9**, 3239-3245 (2008).

80. LOMIZE, A.L., POGOZHEVA, I.D., LOMIZE, M.A. & MOSBERG, H.I. POSITIONING OF PROTEINS IN MEMBRANES: A COMPUTATIONAL APPROACH. PROTEIN SCIENCE **15**, 1318-1333 (2006).

81. LOMIZE, A.L., POGOZHEVA, I.D., LOMIZE, M.A. & MOSBERG, H.I. THE ROLE OF HYDROPHOBIC INTERACTIONS IN POSITIONING OF PERIPHERAL PROTEINS IN MEMBRANES. BMC STRUCTURAL BIOLOGY **7** (2007).

82. MULHERAN, P. & KUBIAK, K. PROTEIN ADSORPTION MECHANISMS ON SOLID SURFACES: LYSOZYME-ON-MICA. MOLECULAR SIMULATION **35**, 561-566 (2009).

83. JUNQUEIRA, L.C.U., CARNEIRO, J. & KELLEY, R.O. BASIC HISTOLOGY. (APPLETON & LANGE, STAMFORD, CONN.; (1998).

84. CHA, T., GUO, A. & ZHU, X.Y. ENZYMATIC ACTIVITY ON A CHIP: THE CRITICAL ROLE OF PROTEIN ORIENTATION. PROTEOMICS **5**, 416-419 (2005).

85. ZHOU, J., TSAO, H.K., SHENG, Y.J. & JIANG, S.Y. MONTE CARLO SIMULATIONS OF ANTIBODY ADSORPTION AND ORIENTATION ON CHARGED SURFACES. JOURNAL OF CHEMICAL PHYSICS **121**, 1050-1057 (2004).

86. ZHOU, J., CHEN, S.F. & JIANG, S.Y. ORIENTATION OF ADSORBED ANTIBODIES ON CHARGED SURFACES BY COMPUTER SIMULATION BASED ON A UNITED-RESIDUE MODEL. LANGMUIR **19**, 3472-3478 (2003).

87. SCHLATTNER, U. ET AL. MITOCHONDRIAL KINASES AND THEIR MOLECULAR INTERACTION WITH CARDIOLIPIN. BIOCHIMICA ET BIOPHYSICA ACTA-BIOMEMBRANES **1788**, 2032-2047 (2009).

88. ENGEL, M.F.M., VISSER, A. & VAN MIERLO, C.P.M. CONFORMATION AND ORIENTATION OF A PROTEIN FOLDING INTERMEDIATE TRAPPED BY ADSORPTION. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA **101**, 11316-11321 (2004).

89 OW, Y.-L.P., GREEN, D.R., HAO, Z. & MAK, T.W. CYTOCHROME C: FUNCTIONS BEYOND RESPIRATION. NAT REV MOL CELL BIOL **9**, 532-542 (2008).

90 SAYLE, R.A. & MILNER-WHITE, E.J. RASMOL: BIOMOLECULAR GRAPHICS FOR ALL. TRENDS IN BIOCHEMICAL SCIENCES 20, 374-376 (1995).

91 FEARS, K.P., CREAGER, S.E. & LATOUR, R.A. DETERMINATION OF THE SURFACE PK OF CARBOXYLIC- AND AMINE-TERMINATED ALKANETHIOLS USING SURFACE PLASMON RESONANCE SPECTROSCOPY. LANGMUIR 24, 837-843 (2008).