In Silico Peptide Selection for Biomining

by

Niloofar Nayebi

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics

University of Alberta

Abstract

Biomining is an efficient way for improving mineral extraction and remediation processes. Using specific/targeted separation is a promising method to increase purification yield of minerals and metals at low cost. In traditional methods extraction and purification of the desired materials from ores requires extensive processes. Through these processes, tailing streams coming out of a mine are usually discharged into tailing ponds. Presence of toxic and bioavailable elements in tailing ponds causes deleterious long-term consequences on the ecosystem. On the other hand, tailings ponds are usually mineral- and metal-rich environments. Our objective in this study was to consider tailings ponds as secondary sources for minerals and to design methods for the removal of toxic elements to help the environment.

To this end, we have introduced a new *in silico* method to select peptides with high affinity and specificity for a given target material to use as a recognition block in biomining applications. The selected peptides are proposed to be used as coating on magnetic nanoparticle core. Peptide-based engineered materials will have the ability to detect, bind and extract the target material by using magnetic fields.

In this project a library of 11,000 peptides has been created using AMBER software. These peptides were clustered and subjected to a recently developed Wigner-D function algorithm to select the conformations with higher shape similarity to the target. Selected conformations were docked to calcite 104 surface using the RosettaSurface package to find the strongest binders from among the peptides in the library.

As a result, eight different sequences, namely, p19 (TTNN), p266 (NTNS), p113 (SSYN), p509 (TQNY), p87 (QSTN), p37 (QSQN), p324 (TYSS) and nu67 (STTC) were identified as strong binders to the calcite 104 surface.

P19, p266, p113 and 4 control sequences (non-binder peptides) were experimentally examined to validate the model. Experiment results showed that the behavior of all the peptides except p113 was consistent with the computational predictions. The reason for this false positive can be explained by various strong intramolecular interactions within the peptides. These interactions reduce the energy of the system and result in incorrect predictions of the peptide-surface interaction binding energy.

Our results demonstrate that the major part of the binding energy is due to electrostatic and van der Waals interactions. The most important interacting amino acid found in these short peptides is glutamine (N).

Acknowledgements

I would like to thank the following people who are always here to help and support me.

First of all, my greatest appreciation goes to my supervisors, Professor Jack Tuszynski and Professor Carlo Montemagno for their wise guidance and invaluable assistance. Without their continuous support accomplishing this stage was not possible.

I want to give my gratitude to my great friends at the Tuszynski's Lab and Ingenuity Lab, Sara Ibrahim Omar, Salva Salmani-Rezaie, Francesco Gentille, Philip Winter, and Hallel Freedman for their support and great company.

I am also thankful of Mike Pacella, a member of RosettaSurface developer group, for his help and contributions in this study; and Sibel Cetinel for doing all the wet lab experiments.

This research could not be done without the computing resources provided by WestGrid, BlueGene, Compute/Calcul Canada and PharmaMatrix clusters.

The last but not least, my warmest appreciation goes to my beloved husband, Hossein, my parents, Shahrokh and Pouran, and my brother, Siavash, who were not physically close to me but I could feel their support and love every single day.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, the importance of biotechnology in the mining industry has grown rapidly. It plays a major role in processing minerals and recovery of metals, as well as pollution control. Biomining is a term used to describe utilization of biological systems, mostly microorganisms, to ease the extraction and recovery of target metals from mineral ores and concentrates [1, 2].

Up until now, biomining has been referred to two processes, namely bioleaching and biooxidation. Sulfide-forms of metals like copper, zinc, iron and molybdenum are the minerals that are the majority of metal ores.

Metal sulfide is insoluble in water, and it needs to degrade to sulfate in order to form a soluble compound. Metal sulfides play an electron donor role for aerobic sulfur-oxidizing microbes which convert them to metal sulfates. In this process the metal is leached into water and can be extracted. This procedure is known as bioleaching [3, 4, 5].

In biooxidation processes, the metal remains in its insoluble form. In the pretreatment step, microbes decompose the mineral to open up the structure and make the target metal more accessible for other chemicals to penetrate the mineral and solubilize it. This procedure is mostly applied to gold recovery from arsenopyrite ores. Gold is solubilized by cyanide and extracted in a subsequent step. Biooxidation process provides a better access of gold to cyanide which ends up with 95% of gold recovery [3, 6]. Bioleaching and biooxidation are based on the same principles and use similar consortia of microorganisms [7]. Specific bacteria and archaea can grow continuously by gaining energy from reduced sulfur or oxidation of ferrous

iron [8]. These microorganisms use the energy derived from a degraded sulfide mineral from the host ore site to fix CO_2 in the atmosphere to be able to use it for their life processes [9, 10]. This process releases the metal (metals as in sulfide minerals).

The ability of these bacteria to oxidize ferrous iron to ferric iron plays the most important role in catalyzing the dissolution reaction of the sulfide mineral. The largest portion of oxidant sulfide minerals is in the form of ferric iron. Some acidophilic prokaryotes oxidize the ferrous iron generated in previous process (equation 1.1) [11, 12]:

$$4Fe^{2+} + O_2 + 4H^+ \rightarrow 4Fe^{3+} + 2H_2O \tag{1.1}$$

In the case of copper mines, based on equation 1.2, copper ore is oxidized by ferric irons which are formed by microbial oxidation of ferrous irons [11]:

$$CuFeS_2 + 8Fe^{3+} + 3H_2O \rightarrow Cu^{2+} + 9Fe^{2+} + S_2O_3^{2-} + 6H^+ \tag{1.2}$$

Although, there is still full-scale adoption problem in practical world in some recent advances in biomining, some areas in biotechnology have been well established in this industry and are applied in many countries. Biomining has contributed an estimated global production rate of 5% gold, 15% copper, and smaller portions of extraction for other metals, such as zinc and nickel [13].

1.1 Biomining History

Although, the first operation known as biomining was initiated in the 1960s [11], biomining has a long history before that. Table 1.1 shows some of the most important achievements in biomining development.

Early miners have been exploiting the capability of microorganisms to decompose minerals and recover metals since the Roman times without being aware of the involvement of microbes [3, 11].

It was understood that water running out of a mine contains some soluble metals, and adding iron into that, one could extract copper. For the first time, as seen in

Time	Places	Target material and method
Middle ages	Spain, China, UK	Copper leaching
1960s	USA	Copper dump leaching
1960-1980	Canada	Uranium mining
1980-present	Chile	Copper heap leaching
1986-present	South Africa	Gold biooxidation (stirred tanks)
1999-present	Uganda	Cobaltiferouspyrit bioleaching (stirred tanks)
2004-2005	Chile	Chalocopyrite concentrate bioleaching
2008-present	Finland	Polymetallic black schist leaching

Table 1.1: Biomining history. this table has been extracted from *Development and application of biotechnologies in the metal mining industry.*

the Rio-Tinto mine, a villager in the area reported that iron would disappear when introduced into the river. The region's priest discovered an electrochemical reaction which led to the precipitation of copper from that water as iron dissolved [5]. This phenomenon was thought of, at the time, as an example for alchemy (converting one metal to another) [5].

The modern era of mining began by understanding the role of bacterium, Acidithiobacillus (called Thiobacillus, at the time) ferrooxidans, in copper extraction [2]; after about a decade from the discovery of first species which were catalyzing the dissolution metal-containing sulfide minerals in an acidic environment [14, 15].

The first commercial biomining operation was established by Kennecott Copper Corporation for copper extraction from their waste rock dumps in the Bingham Canyon Mine, in Utah, USA [2, 11, 16].

1.2 Applied Biomining Methods in Modern Technology

Three different engineering approaches represent commercial biomining; (a) dump bioleaching, (b) heap bioleaching and biooxidation, and (c) stirred tank bioleaching. These methods have been used globally to ease the extraction of precious metals (such as gold) from ores of sulfide minerals, while the metal is locked in a sulfide

mineral to leach base metals and to obtain copper from secondary copper ores [2, 7].

1.2.1 Dump Bioleaching

Large rock pieces containing less than 0.5% of copper from open pit operation are dumped, because the cost of flotation and smelting exceeds the value of the extracted materials.

Millions of tons of run-of-mine ores are stored in dumps (often more than 60 m deep), and it dilute sulfuric acid is applied. As the liquid permeates through the dump, stimulates naturally occurring mineral-oxidizing bacteria. The bacteria play a catalyst role for the copper sulfide minerals oxidation reaction. Leached copper is collected and subjected to solvent extraction [2].

The size of the fractions in the dump makes the transfer of the solution inefficient. In addition, air shortage puts huge limitations on the activities of microbes. Nevertheless, dump bioleaching remains one of the most economic methods for copper recovery from low-grade ores [2].

1.2.2 Heap Bioleaching and Heap Biooxidation

In this method, ores are crashed to the size of 20mm in size or smaller pieces and gathered in rotating barrels. Dilute sulfuric acid is added to make suitable environment for microorganisms. It glues small pieces to make bigger rock particles.

The ore is transferred with engineered pads, which are positioned at intervals along with high-density polyethylene (HDPE). The pads are covered with leaky plastic drains. The bigger rock particles remain above the drain lines and a network of air lines is arranged to provide air between these pieces.

External blowers are used to provide sufficient air for microorganisms. The ore collected in a 610m-height stacker and irrigated by acidic raffinate, is directed to solvent extraction unit. The presence of plentiful sulfide minerals and iron in acidic condition develops microorganisms and catalyzes the copper extraction [2].

Heap bioleaching is used around the world in secondary ores containing mineral chalcocite (Cu_2S) to extract copper [2]. 7% of the annual copper production is achieved by heap bioleaching [2].

Heap biooxidation applies to the ores with gold particles locked in mineral sulfides, usually pyrite or arsenopyrite minerals. In this case, heap biooxidation oxidizes sulfide minerals to open up the structure for the gold dissolution process. The process for heap biooxidation is very similar to heap bioleaching.

The first step is crushing the ores. Crushed ores are inoculated with three spices of microorganisms: mesophilic bacteria, extremely thermophilic archaea, and moderately termophilic bacteria [2].

1.2.3 Stirred Tank Bioleaching

This technology is usually used for refractory of gold concentrate as it has high operating costs [2]. The reactions in this technology are carried out in a series of bio-reactors (stainless steel tanks). The volume of the bio-reactors is greater than 1300 m^3. Each tank contains about 109 microorganisms per milliliter of solution and is equipped with agitators, which keep the grounded sulfidic gold concentration in suspension and provide sufficient amount of oxygen and carbon dioxide to the solution for microorganism activities [2].

In the case of sulfidic-refractory gold, the valuable metal is in solid residue form which exits the last reactor [2]. This solid residue is subjected to cyanide leach to recover the gold [2]. This method is working with high efficiency; the range of gold recovery is 95% to 98%. Meanwhile, if the concentrate is a base metal, leach solution will be used to dissolve the metal [2]; therefore, unlike the gold concentrate the solution is subject to a metal recovery process [2].

1.3 New Approaches to Biomining

Biomining is implemented in today's mining industry, as easy-to-process and high-grades ores are being depleted, this stem is capable of processing low metal concentrate and refractory. [13].

Metal mining has significant environmental and economical challenges in this century and biotechnology has the answer to part of these problems [11].

Since the beginning of civilization, around 1150 million tons of heavy metals

have been extracted and caused the production of an estimated 5 to 7 billion tons of tailings per year [17].

After mining and extracting the desired minerals, residuals and rejects of mills form the tailings streams [18]. The contents of these streams are mostly the fine-grained rock from the mined ore associated with processing water, reagents and a significant amount of dissolved metals from the host ores [18, 19]. Mining wastes often show a greater amount of metals compared to accessible ore bodies [11].

Tailings streams are usually discharged into tailings ponds; this is used globally by the mineral industry as an inexpensive and accepted technology [18, 20]. The presence of heavy metals, toxic and bioavailable elements in tailings ponds can eventually lead to serious ecosystems damages and risks to human health [21, 22]. These damages can have negative long-term consequences which can remain for thousands of years. One example is the presence of water contamination after 4500 years at the site of historical mining in Rio-Tinto estuary in Spain [23].

On the other hand, tailings ponds usually have a lot of valuable minerals, depending on geochemical nature of the host ore, that are usually considered as waste [24].

Moreover, mine operators have to face today's main problems such as carbon dioxide emission and energy consumption, in addition to an increasing need for mineral and metal resources. Therefore, reprocessing mining waste to recover some of the valuable resources we already wasted is a necessary step.

Our ultimate goal is to use tailings ponds as a secondary source of mineral resources to extract the valuable materials, as well as to remove harmful substrates from them. We aim to achieve this goal using biomining technology.

Current commercial-scale biomining processes are limited to sulfidic ores where the desired metal particles are presented in sulfide-form of minerals or reduced oxidation state in a mineral [25]. In both cases, mineral dissolution takes place in the presence of the prokaryotic acidophiles in acidic lixiviant to produce chemical oxidant, sulfuric acid, and ferric iron [8, 25, 26]. In addition, many of the valuable metals are in oxidized or reduced ores, as well as other forms.

There are two methods mostly used by the mining industry to remediate mine

waters at mining sites.

The first one is an active chemical approach which consists in raising water pH by adding alkaline (mostly CaO) to precipitate metal in carbonate and hydroxide forms which generates a suspension. Dewatering and flocculation of the suspension will produce sludge, a mixture of ferrihydrite, gypsum, and different metals [11].

Another method is a passive biological treatment. In this approach, organic composts usually contain degradable materials like animal manures and recalcitrant e.g., sawdust and sometimes zero-valent iron (ZVI) will be added [11]. Composts will flow through the system in mine waters or a place used as an intercept in the underground for contaminated underground waters [11]. These organic materials can degrade under anaerobic conditions and leave small organic compounds which are electron donors for sulfate and iron reducing bacteria.

However, this method's performance usually changes seasonally and does not show acceptable performance under extremely acidic condition [27]. The compost should be replaced every 5-10 years as the metals accumulated in the compost are gradually exhausted [11]. Both the compost and sludge from the above-mentioned method are classified as toxic wastes, so, neither of them is allowed to be used for recovery of the metals and materials [11].

An alternative approach is using different species of bacteria which participate in metal biomineralization and sometimes the ones generating alkalinity. This approach targets specific metals and removes them from contaminated waters. The water coming out of mines usually contain noticeable concentration of several metals and metalloids [11]. So, different methods are needed to clean up the water.

Over the past 50 years, biomining has been limited to the use of microorganisms [28]. Now by advancing modern technology, biomining can be extended to use proteins and peptides.

In nature, proteins have highly specific molecular recognition properties. This allows them to recognize individual target molecules among all other competitors.

In living systems, proteins and peptides are associated with inorganic material and are thought to have a key role in controlling the nucleation, growth, and organization of biocomposite materials [29, 30]. Proteins operate in this way, because

of their specific noncovalent interactions with inorganic materials [29, 30, 31].

Almost all proteins associated with inorganic materials show several repeating peptide domains [32]. Peptides are less complex compared to proteins, and they can have the same level of recognition for a specific material [33]. Manipulation of peptides is easier as they provide chemical diversity (aromatic, basic, hydrophobic and etc.) and can show different secondary structures [30]. These factors make peptides more preferential than proteins as recognition blocks for inorganic materials.

The above-mentioned properties of peptides inspired us to use these small recognition blocks for biomining research fields.

These molecular recognition elements are used as a coating on magnetic nanoparticles. By placing these smart particles in the stream coming out of mines, we can target specific metals, rare earth elements, or minerals at low concentration.

In other words, our peptides-based engineering materials have the ability to detect and bind with specific materials, even in low concentrations, and extract them by using magnetic fields. Figure 1.1 shows a schematic of this process.

1.4 Biomining Advantages

The advantages of biomining versus conventional mining can be reviewed by economical and environmental points of view.

Biomining consumes less energy compared to other processing technologies (compared to the energy required for roasting and smelting which is the major investment of mining companies [11]), and it operates at low temperatures (usually 30C- 50C) and pressures [11]. Biomining happens in an atmospheric pressure and usually does not need extra heat sources, as the oxidation process of sulfide minerals is an exothermal reaction.

Biomining is perceived as a green method, because the bacteria involved in biomining fix the CO_2 in the atmosphere like green plants and do not produce environmentally harmful gases [5, 8, 18].

Furthermore, conventional mining tailings ponds may leach in the presence of water and air and cause generation of acid and metal pollution [18]. It is worth

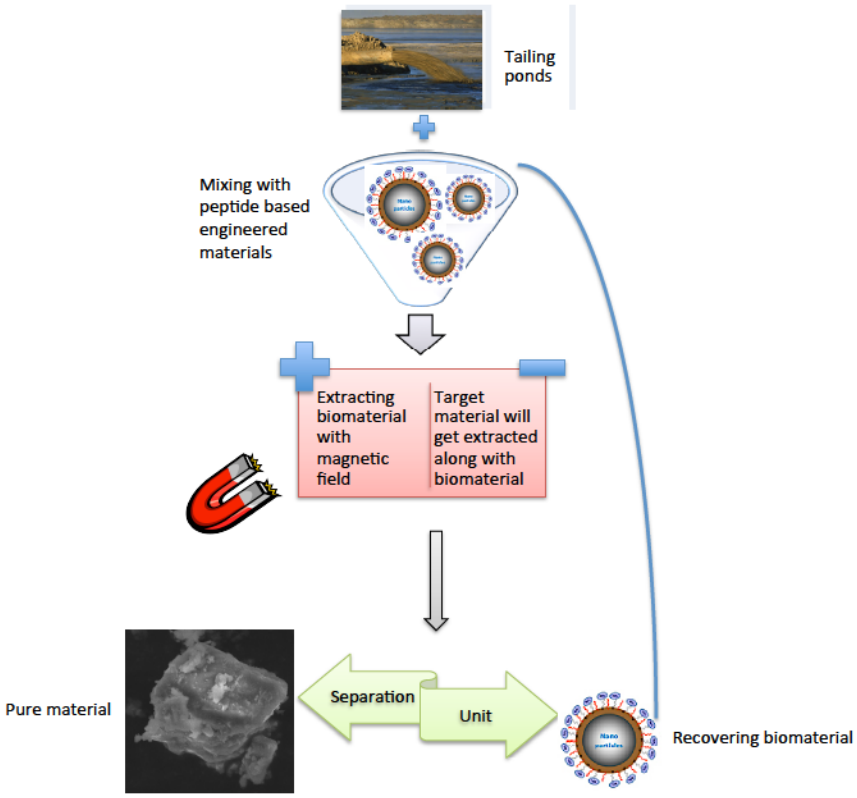


Figure 1.1: Schematic of the new approach in biomining.

mentioning that tailings from biomining are not chemically active, as this method is physiochemical.

The other distinct advantage of biomining is in recovery of materials from low-grade ores, ponds and dumps which were abundant from previous mining operations; so, this method is capable of treating the material that would be considered as waste [3, 13, 18]. Many of these materials do not show economic advantages for recovering with non-biological methods [18].

Chapter 2

Organic-Inorganic Interactions, Principles and Related Issues

2.1 Introduction

Understanding the principles of peptide-inorganic interactions is very important in various research fields, such as biomedicine, biochemistry and biotechnology.

In nature, peptide/protein-inorganic interactions are important for the formation of highly ordered materials [31, 34]. One example is the formation of bone structure through the interaction between protein collagen and mineral hydroxyapatite [35, 36]. Different ratios of protein-mineral concentrations lead to different mechanical strength in bones.

Organic-inorganic interactions play a vital role in the immune response to an implant or biomaterial. Therefore, successful performance of therapeutic devices is defined by their biocompatibility, which depends on organic-inorganic interactions [37, 38].

Moreover, this kind of interaction is the basis for composition material formation in nature [36]. Recent advances in material science make use of peptides for inorganic surface modifications [39].

Our aim is to make use of this interaction for biomining purposes. In our study, as an example of this phenomenon, we are selecting peptides that have high affinity and specificity to calcite ($CaCO_3$) to extract this mineral from the waste of mining.

The implications of organic-inorganic interactions with respect to our understanding of the behavior of peptides on inorganic interfaces are addressed in this

chapter.

2.2 Physical Chemistry

Two important factors involved in any interfacial process are their kinetic and thermodynamic properties. In organic-inorganic interactions, reaction kinetics represents the rate of peptide adsorption and desorption on inorganic surfaces. It also takes the changes in peptide conformations in adsorbed state into account. On the other hand, thermodynamics focuses on the state of the system at chemical equilibrium at a specific temperature and pressure [30]. For simplicity, we assume the interaction is as follows [40]:

$$P + S \rightleftharpoons PS \tag{2.1}$$

where P represents the peptide in solution, S represents the site surface and PS represents the peptide-surface adsorbed state. Assuming the reaction is reversible and elementary on both sides, its kinetics can be described by these equations [30]:

$$\frac{d[P]}{dt} = -k_{ads}[P][S] \qquad \frac{d[PS]}{dt} = -k_{des}[PS] \tag{2.2}$$

where k_{ads} is the rate coefficient for the adsorption reaction and k_{des} is the rate coefficient for the desorption reaction.

The equilibrium constant can be simply written as [41]:

$$K_{ads} = \frac{[PS]}{[P][S]} \tag{2.3}$$

In equilibrium, the concentrations of P, S, and PS remain constant. Therefore, one finds the following relationship between the rates of reactions [30]:

$$\frac{k_{ads}}{k_{des}} = K_{ads} \tag{2.4}$$

For an organic-inorganic interaction, the most important reported thermodynamic property is the change in the free energy upon adsorption [30]. Free energy is a thermodynamic state function such that, at equilibrium, its changes depend on

its quantity in the initial and final states, irrespective of how the system arrived to the new state [42]. The changes in the Gibbs free energy for peptide-inorganic interaction can be illustrated by changes between the free energy of the solution state (where the surface and the peptide are far from each other, and there is no interaction between them) and in the adsorbed state (where the peptide and the surface are in contact with each other)[30].

2.3 Driving Force

At present, the nature of peptide-inorganic interaction is not completely understood. To understand the mechanism of peptide adsorption into an inorganic surface, the enthalpic and entropic components of the free energy of the system should be considered [43].

Generally, adsorption is an exothermic reaction, because there is a tendency to minimize the enthalpy of the system. Moreover, losing mobility due to adsorption decreases the entropy of the system. This decrease in entropy is less reported for long polypeptides compared to several short ones, while the total adsorbed mass is equally considered [43].

Electrostatic interactions have been shown to be an important component of the driving force for this type of interactions [44, 45]. Other forces, such as hydrophobic or hydrophilic interactions, hydrogen bonding, and van der Waals forces have significant contribution to the driving forces of the interaction, especially where peptides are neutral or have repulsive charge conditions [44, 46].

The organic-inorganic interaction can be influenced by different factors such as: intrinsic properties of peptides, inorganic surface physiochemical properties, and the media where the complex is present [47].

Amino acid differences are the reason for the diversity among peptides. The inherent properties of peptides, which are known to have influence on peptide-inorganic binding, are charge, hydrophobicity and peptide conformation. The net charge of a peptide also plays a key role in electrostatic interactions with charged surfaces [43, 47].

The properties of the media in which the adsorption process occurs, also have a strong effect on peptide-surface interaction [43, 47]. These properties include the ionic strength, pH, solution fraction and the temperature of the media [43].

Peptides and surfaces can have different electrostatic states at different pH [48]. For instance, the charges of metal oxides, hydroxides or peptides depend on the net gain or loss of protons in the media.

Ionic strength represents the concentration of dissolved ions in the media. Electrostatic interactions between charged elements, peptide-peptide and peptide-surface couplings decrease when ionic strength increases [43]. Different studies have shown that the type of salt ions present in the media also has an effect on adsorption processes. Peptide precipitation may increase with the presence of high concentrations of ions, like SO_4^{2-} and Mg^{2+} or reduced by ions, such as NH_4^+ [49, 50]. Ca^{2+} is reported to increase the peptide-surface binding [51] by acting as a bridge between the peptide and the surface.

The increase in temperature causes the acceleration in diffusivity and mobility of peptides towards the surface and a net gain of entropy by releasing adsorbed salt ions and water molecules [43].

2.4 Surfaces and Interfaces

To gain insight into the adsorption behaviors of materials-binding peptides, characterization of the surface of the target material is as important as the conformations of the peptides themselves.

Characterization of the target inorganic material surface has a huge impact on gaining insight into organic-inorganic interactions. The shape of a solvated material interface, the presence of steps, kinks, and defects on the planar material surface, and its charge state in solution with different pH are some examples of the important factors, which have effects on the adsorption of organic molecules on surfaces [52].

Moreover, the tendency of materials to oxidize adds some challenges to the understanding of how peptides recognize these materials; as some materials, such as semiconductors (e.g. GaAs and CdSe), represent complex oxide structure and are

not very well understood at the atomistic level [30]. Lastly, the above-mentioned factors have an influence on the structure of the solvent at the interface in addition to the material surface structure [30].

Another level of complication arises in the surface reconstruction. Crystals have a repetitive pattern in three-dimensional space for the arrangement of their atoms, ions or molecules. Different crystalline surfaces are generated by cutting the bulk crystalline material along a plane, which is defined by three points (which are not all collinear) in crystal lattice. Each crystalline surface can represent a different arrangement of atoms on the surface (see Fig 2.1).

Therefore, a favorable surface reconstruction affects the peptide adsorption on the surface via a change in the arrangement of atoms and solvent structures on the surface at the interface [30].

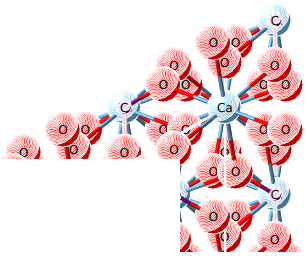
The structure of the solvent at the interface is thought to have a huge effect on the peptide structure in its adsorbed state [30]. One possible way for this effect can be the changes in the solvent's hydrogen bonds in the interfacial region. This can possibly alter the probability for peptide-interfacial solvent hydrogen bonds.

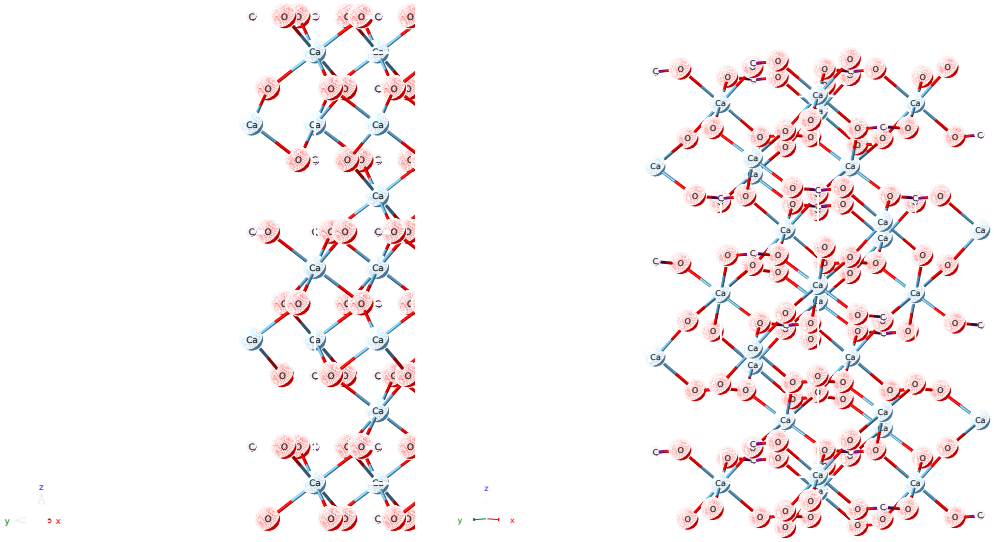
The thermodynamics of the peptide-inorganic surface complex depends on how tight and strong the solvent and surface interaction is. If the peptide is making direct contact with the surface, it will replace the solvent molecules on the surface [53, 54]. This can be analyzed in terms of free energy.

The changes in free energy can be attributed to a loss of enthalpy for the released solvent molecules, a gain of enthalpy for peptide adsorption, a gain in entropy for joining water molecule which are replaced by peptide to the bulk, and a loss in peptide entropy because of its binding to the surface.

2.5 Secondary Structure and Function

Peptides are made up of distinct sequences of amino acids, which form polypeptide chains. A peptide secondary structure is defined by the pattern of hydrogen bonds formed between carbonyl oxygen and amine hydrogen atoms in the peptide backbone. The other way to define a secondary structure is based on a regular pat-





(c) N(

(d) N(111)

Figure 2.1: Planes with different Miller indices in calcite.

tern of dihedral angles in the backbone in a particular part of the corresponding Ramachandran plot (2.2) [55]. Secondary structure provides the scaffold for the peptides.

An alpha-helix, a beta-sheet, a polyproline II helix (PPII), and a beta-turn are some examples of secondary structures. In this area, a random coil is not referring to a specific secondary structure. It has been defined as the situation under which dihedral angels of each residue backbone do not depend on neighboring residues. It can be concluded that all the possible conformations should be sampled in these cases [56].

Up until now, experimental studies showed a lack of well-defined secondary structures for peptides in solution and in adsorbed states on surfaces [57, 58, 59, 60].

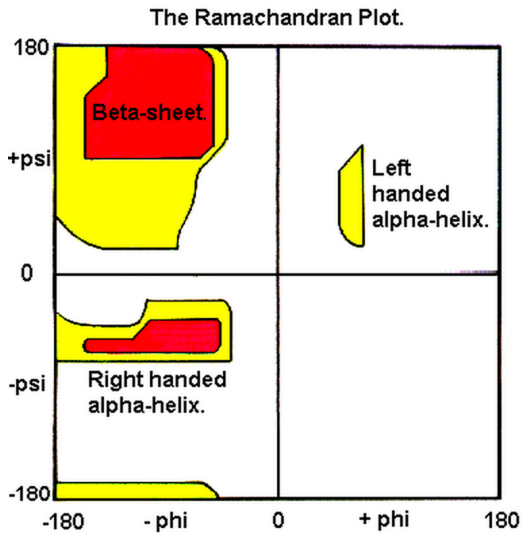


Figure 2.2: Ramachandran plot. The values of and angles (or conformation) which are possible for an amino-acid residue in a protein.

The relation between a protein's three-dimensional structure and its function is still under investigation. Researches show that a 3D structure is not necessary for proteins function [30].

Within our body, more than 40% of proteins are partially unfolded or without any well-defined structure in their active form [61]. These proteins, which lack a three-dimensional structure, are known as Intrinsically Disordered Proteins/Peptides (IDPs) [62, 63]. IDPs can get different conformations upon external stimulus, such

as binding and are showing target dependent conformations [64, 65].

The inorganic-binding peptides are thought to change their conformations upon adsorption [30]. This has led to the idea that inorganic-binding peptides can be classified as IDPs [57].

Experimentally determining a peptide's conformation on the material interface in solution is a complicated task. Interconversion among different conformations in the state of adsorption and structural changes caused by peptide-peptide interactions can be examples of this complexity [66, 67]. Therefore, obtaining structural information at the atomistic level for a peptide adsorbed on inorganic surfaces remains a challenge [30].

Understanding the relation between a peptide sequence and all conformations supported by its sequence and its tendency to selectively bind to a surface plays a key role in future inorganic-binding peptide design.

There are many examples that show a particular sequence of the amino acids in a given peptide controls binding selectivity and affinity between the peptide and the material substrate [30, 68]. As an example AQNPSDNNTHTH binds to GaAs with high affinity and to silica with low affinity [69]. By changing the order of amino acids to TNHDHSNAPTNQ the selectivity will be destroyed, and its binding to both materials with the same strength is found [69].

Therefore, it is clear that a sequence of the peptide is an important factor for selectivity and affinity of this peptide to a given inorganic material; and that having the residues, which have strong individual binding properties for the material, is not enough [30, 58].

In addition, each peptide-inorganic binding interaction is unique, and one cannot see obvious patterns among strong binders for a given material or noticeable differences between strong and weak binders [30, 60, 70, 71, 72, 73].

Today's computational technology can help us study the interactions between peptides and inorganic surfaces at an atomistic level. The computational modeling can show us interactions between atoms and lead to a better understanding of the link between sequences of peptides and their affinity and selectivity for a given inorganic material.

2.6 Experimental Methods

There are two approaches to select peptides for an inorganic target; one is deriving them from naturally occurring protein-inorganic complexes; for example R5 peptides extracted from diatoms [74]. The second approach is a combinatorial based approach such as phage display [75, 76] and cell-surface display [33, 39].

Phage display is the most common approach among combinatorial techniques and is used for selecting strong binding peptides through a screening process [77]. Several kinds of phage can be used in this technique. T4, T7 and M13 phages are some examples.

M13 is the best-characterized and the most commonly used library display. M13 bacteriophage (phage) is a bacterial virus comprised of a single-strand DNA (ssDNA) contained in minor and major coat proteins. Gp3 and gp8 coat proteins are suitable choices in M13 for making a display library [78, 79].

The libraries are generated by insertion of random DNA sequences (nucleic acid sequences) into a certain location of the phage genomes or bacteria plasmids. The sequences are expressing specific peptide sequences on the surface of a bacterium or a phage.

These peptides are usually 5-38 amino acids long, longer sequences decrease the phage viability and efficiency [78]. Millions of different peptides are generated on the surface of different phages or cells. Diversity of peptide libraries can go up to 10^{11} different sequences [76, 80].

In the first step, the entire library content is exposed to a target substrate. The non-bound phages are removed in washing steps, and bound ones are collected and amplified in the Escherichia coli bacterial host. The elute phages or cells are exposed to the target material. These steps are repeated with more restricted washes until the phages with strong binding to the target material remain.

After identifying the phages with high affinity, their DNAs are sequenced to obtain the peptide sequences with high affinity to the specific material. These peptides are called genetically engineered polypetides for inorganics (GEPI) [71].

The pioneering work for selecting peptides for inorganic material using the

combinatorial based approach has been done by Stanley Brown by surface display to select gold-binding peptides [81]. In the last decade, peptide sequences have been selected for different targets by using these methods, namely: metals and their oxides, Au [81, 58], Ag [72], Pt [82, 83], SiO_2 [68, 84], Cu_2O [85], ZnO [86, 87], TiO_2 [88, 89, 90], and ZrO_2 [91], minerals including hydroxyapatite, calcite, sapphire [92, 93, 94], semiconductors including GaN, ZnS, and CdS [70, 95, 96], graphite [97, 98], and diamond-like carbon [99]. This information has been obtained from *Bio-Inspired Nanotechnology: From Surface Analysis to Applications* [30].

Phage or cell display methods have been widely used during the past decade but have not significantly improved [100].

The main limitation of these methods is the heavy bias in the combination of the library sequences [100]. To illustrate, there are several amino acids which are very rare or even not present in the whole library (e.g. cysteine) [100, 101].

Other limitations are included in not having dominant sequences or background interference caused by nonspecific binding peptides [74].

2.7 Computational Studies

Molecular simulation of organic-inorganic interactions can provide a detailed atomistic insight and reveal remarkable principles of the interactions. Modeling can be performed at different levels of detail. The factors involved in the selection of the methods used, are the type of information required, the process time scale, and the size of the system.

Among all multi-scale simulation methods, there is always a give-and-take regarding the accessible time and the desired accuracy [102].

The methods that have been used in this field can be narrowed down to Quantum Mechanics (QM), Monte Carlo (MC), and Molecular Dynamics (MD) simulations.

2.7.1 Quantum Mechanics Simulation

Quantum mechanics (QM) represents the highest level of theory and is treating electron transfer, bond forming or breaking, etc., i.e. generally speaking the inclusion

of electronic effects. This method reveals an accurate binding energy and conformational properties of a peptide in the adsorbed state on the inorganic material such as, interatomic distances and angles.

This method is computationally very expensive and is limited to a small number of atoms. Such calculations on even a small system can be used for improving molecular mechanics force fields [103, 104].

QM calculations have parameterized single amino acids in contact with hydroxyapatite (HAp) [105], gold [103, 106, 107], copper [108, 109], nickel [110], TiO_2 [111].

2.7.2 Molecular Dynamics Simulation

Molecular dynamics (MD) is a deterministic method. The time evolution of the state of each atom is determined by solving its equations of motion (Newton's laws) for a specific thermodynamic ensemble. Based on theory, thermodynamic ensemble averages should be obtained if an MD simulation runs involved a sufficiently long time. In such a state, the system reaches its equilibrium and thermodynamic properties can be calculated [112, 113].

This classical simulation method uses a force field (FF) which is an interatomic potential energy, whose parameters are obtained from experimental data or quantum calculations. This allows to study interfaces and dynamics of much larger and more complex systems. Several FFs for a wide range of organic molecules have been developed [114, 115, 116].

There are some FFs which have been developed for use with minerals and their interfaces in solution, ranging from a simple non-bonded point of view [117], to bonded layers [118, 119], and reactive methods [120].

The main problem in simulating organic-inorganic interactions is no force field has been developed to consider organic-inorganic interface mineral computationally. Therefore, in MD simulation of these systems, combination of general FF and mineral FF have been used. MD simulation of adsorption of small alcohol onto $Al(OH)_3$ using the combined FF showed the results with agreement with experimental results [121].

MD simulation has been used to model the adsorption of different types of organic molecules onto mineral surfaces. Some examples can be small organic molecules [122, 123, 124, 125], bioorganic acids [126, 127, 128], polymers [129, 130, 131], and entire nucleic acid strands [132, 133, 134, 135, 136].

One of the advantages of MD simulation is its ability to calculate kinetic properties.

2.7.3 Monte Carlo Simulations

The Monte Carlo (MC) method is a technique to obtain possible configurations of the system. In principle, MD should be able to sample all these configurations, but in some cases, MD is not efficient, especially when the rate of change in the system's dynamics is extremely slow.

The most common used implementation of the MC method is the Metropolis algorithm. In each step, MC randomly changes the configuration and calculates the corresponding energy. If the change in energy is negative the move will be accepted and if it is positive, the new configuration will be accepted with a probability of exp(-E/k_BT). Like an MD simulation, in MC simulations the average will converge to a thermodynamic average.

The simplest MC model for studying organic-inorganic interactions is on a homopolymer [137]. From this model, a complete folding phase was obtained [138]. Another study focused on pH-dependence of the adsorption of a charged polymer on the surface [139]. Some other studies showed the conformations of proteins at a particular surface along with their interaction properties [140, 141, 142]. More complex studies apply explicit solvation in the simulations of a small peptide on a surface [143, 144].

Recently, the MC-plus-minimization method has been used to determine the minimum energy and a complementary structure between peptide and surface [145].

Most of the computational methods that have been implemented to study the organic-inorganic complexes used experimentally determined sequences and focused on the understanding of the interactions and driving forces within the system. However, few studies have identified binding peptide sequences for a given material

in silico.

Oren et al. categorized experimentally selected peptides based on a sequence similarity scoring matrix and predicted new sequences with known affinity based on this matrix [68].

In another computational study by Masica et al., 16-mer peptide sequences have been designed involving 7 amino acids of the 20 naturally occurring ones [145].

Chapter 3

Methods

3.1 Library

The twenty standard amino acids have been grouped in seven different categories as acidic (Aspartic acid, Glutamic acid), basic (Arginine, Histidine, and Lysine), polar (Asparagine, Glutamine, Threonine, Tyrosine and Serine), non-polar (Methionine, Phenylalanine, Proline, Tryptophan, Valine, Alanine, Cysteine, Glycine, Isoleucine, and Leucine), aromatic (Tryptophan, Phenylalanine, and Tyrosine), nucleophilic (Threonine, Serine, and Cysteine) and special cases (Cysteine, Glycine and Proline) to cover all their distinct chemical properties. All possible permutations of the amino acids in each group have been considered. This classification produced a set of with 10,965 different 4-mer sequences; these peptides were built by the XLEaP module of AMBER [146] using their linear sequences.

3.2 Molecular Dynamics Simulations

In order to obtain energetically equilibrated peptide structures we performed molecular dynamics (MD) simulations for each of the peptide structures generated in our virtual library. All simulations were carried out within the canonical ensemble (at a constant temperature) using the SANDER program in AmberTools12 [146]. The AMBER99SB force field [147] was employed and implicit solvation effects were introduced by using the Generalized Born module [148] in AMBER12 [146].

The peptide structures were minimized for 500 steps of steepest descent followed by 500 steps of conjugate gradient algorithms.

In the next step the system was heated up from 0 to 300K. The system was simulated over 50ps in six incremental stages. As we could not start with an experimental crystal structure (which was obviously not available) and the system was generated manually, it is not stable enough therefore it is heated up in different stages to be equilibrated in each stage and by doing so the chances of exploding the system, were significantly reduced. A very short time step was used as a heating step to be able to control the relaxation of the system.

The fully unrestrained system was simulated for 50 ns at 300K generating 50,000 three-dimentional conformational structures for each peptide.

3.3 RMSD-based Structure Clustering

The generated peptide structures were clustered to obtain manageable numbers of representative structures. The average-linkage algorithm [149] in the PTRAJ utility of AmberTools12[146] was used for this purpose. To remove structural differences due to rotations or translations the conformations were RMSD-fitted to the structure at the start of the production step. 5000 peptide conformations, at an interval of 20 ps of the entire MD simulation, were clustered based on mass-weighted RMSD of the backbone to 1-100 clusters.

To find an optimum number of clusters for each peptide two clustering metrics, the Davies-Bouldin index (DBI) [150] and the percentage of variance (sum of squares of the regression (SSR) total sum of squares (SST) metrics)) [149] were calculated. The best number of clusters would ideally occur at a local minimum of the DBI values and a corresponding plateau in the percentage of variance [149]. The highest populated clusters for each peptide whose total was more than 70% of the total conformations in the representative trajectory were used for analysis using the recently developed Wigner-D function algorithm [151] that allows for rapid and accurate 3D geometrical shape analysis.

3.4 The Wigner-D Function Algorithm

The Wigner-D function algorithm is a *de novo* algorithm for comparing arbitrary 3D structures with different number of atoms, which was recently developed by Saberi et al. [151]. It was used for determining the level of shape similarity between a peptide and the target calcite surface.

In this algorithm, the Cartesian coordinates, xyz, of all atoms in a molecule are converted to the corresponding Euler angels, $\alpha\beta\gamma$, relative to molecule's center of mass. Second step would be defining the molecule shape function *f*(α, β, γ). The peptide is defined as the convex hull of the structure composed all of the peptide's atoms. The presence of the molecule's atom in a voxel makes the *f* function equal to one and *f* is zero in the absence of an atom. This function is expanded in term of Wigner-D function (equation 3.1) [151].

$$f(\alpha,\beta,\gamma) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{n=-l}^{l} C_{lmn} D^{l}_{mn}(\alpha,\beta,\gamma) \tag{3.1}$$

D^{l}_{mn} is a Wigner-D function and it is defined as below (equation 3.2) and C_{lmn} are expansion coefficients [152, 151].

$$D^{l}_{mn}(\alpha,\beta,\gamma) = e^{-im\alpha} e^{-im\gamma} d^{l}_{mn}(cos\beta) \tag{3.2}$$

$d^{l}_{mn}(cos\beta)$ is defined as [151]:

$$d^{l}_{mn}(cos\beta) = \varepsilon \sqrt{\frac{(1-\frac{\nu+\sigma}{2})!(1+\frac{\nu+\sigma}{2})!}{(1-\frac{\nu-\sigma}{2})!(1+\frac{\nu-\sigma}{2})!}} 2^{\frac{\nu+\sigma}{2}} (1-x)^{\frac{\nu}{2}} (1+x)^{\frac{\sigma}{2}} P^{\nu\sigma}_{l-\frac{\nu+\sigma}{2}}(x) \tag{3.3}$$

where $P^{l}_{m}(x)$ is the associated Legendre polynomial, $\sigma = 1$ if $n \geq m$ and $\sigma = -1$ if $n < m$, $\nu = |n-m|$ and $\sigma = |n+m|$

The discrete Fourier transform in terms of the Wigner-D function on SO(3) can be written as below [151, 152, 153]:

$$f(\alpha,\beta,\gamma) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{n=-l}^{l} \hat{f}_{lmn} D^{l}_{mn}(\alpha,\beta,\gamma) \tag{3.4}$$

$\hat{f}$ is the Fourier transform of f. by comparing equation 3.4 with equation 3.1 it can be concluded that the C_{lmn} coefficients can be Fourier transforms of the f

function. It was understood that structure factor is representing the corresponding Fourier transform of a shape function [154]. Thus, the C_{lmn} coefficients are introduced as the structure factors. The novelty in this method is illustrating the relationship between the expansion coefficients and the structure factors. Considering the calculated shape function, C_{lmn} matrix elements can be obtained by equation 3.5 [151]:

$$C_{lmn} = \frac{(2l+1)}{8\pi^2} \int\int\int f(\alpha,\beta,\gamma) D^l_{mn}(\alpha,\beta,\gamma) sin\beta d\beta d\alpha d\gamma \tag{3.5}$$

Since a structure factor is a complex number, it can be translated to the 2D Euclidean space. Therefore, for any molecule a space with a dimension of two times of the calculated size of the structure factor can be defined. In the generated matrix, the distance between each pair of elements is calculated and results in an $n \times n$ matrix. The distances matrices of two molecules are used to indicate the correlation between two molecules as the equation 3.6 shows.

$$D^2 = 2\sum_{i<j}^{n}\sum_{j=2}^{n}(d_{ij} - d'_{ij})^2 = 2\sum_{i<j}^{n}\sum_{j=2}^{n}(d^2_{ij} + d'^2_{ij} - 2d_{ij}d'_{ij}) \tag{3.6}$$

where d_{ij} and d'_{ij} are the elements of the distance matrix for each of the two molecules.

The equation 3.6 is a RMSD relation which is used most commonly for evaluating the shape similarity between two proteins. In this equation, if there is a maximum similarity the D^2=0 and if there is no similarity $D^2 = d^2_{ij} + d'^2_{ij}$. To obtain a direct measure of shape complementarity between two molecules, Similarity Value (SV) was defined as [151]:

$$SV = \frac{1}{2}(1 - \frac{D^2}{d^2 + d'^2}) \tag{3.7}$$

SV and RMSD are not equivalent and the comparison between two molecules with different number of atoms by SV gives more accurate results [151] since SV is scale independent.

All the selected conformations from the clustering step were subjected to this calculation. For all the groups except non-polar top 10% of the most similar peptides were selected for further analysis. The corresponding number for the non-

polar group was 1% (due to their lower probability of binding to the surface). For each peptide's most similar conformation have been used.

3.5 Visual Molecular Dynamics

Visual Molecular Dynamics (VMD) [155] was used to place the peptide on top of the surface. The center of mass for the peptide and the surface was calculated and the peptide moved above the center of the surface.

3.6 RosettaSurface

RosettaSurface [156, 157] has been used to dock the peptides onto the calcite 104 surface. RosettaSurface is a MC-based docking method which incorporates minimization after each docking cycle.

RosettaSurface starts from a fully extended peptide chain that undergoes cycles of refinement to get the structure close to what is frequently occurring in the Protein Data Bank (pdb) database [158] In this study a modified version of RosettaSurface [156, 157] was used to start with the structure predicted by an MD simulation and selected by clustering and the use of the Wigner-D function algorithm.

The structure goes to n cycles (n between 1 and 5) of refinements in implicit solvation. During each cycle, the peptide underwent two different loops of optimization, an inner and an outer loop. The difference between these loops is in the method of minimization, which is followed by each applied movement. The outer loop used the conjugate-gradient minimization and the inner loop used a line minimization along the initial gradient.

In each cycle the outer loop was repeated five times and the inner loop was repeated five times for each outer loop.

At this point the surface coordinate was introduced and adsorbed-state complex formed by bringing these two structures in contact in a random orientation. 5-n refinement cycles (similar to the solution state refinement) were applied to the complex followed by a modified version of RosettaDock [159], a high-resolution docking procedure [158]. In this modified version of RosettaDock the backbone,

side-chain and rigid-body optimization are applied at the same time.

After completing the above-mentioned cycles two additional refinement cycles are applied to get the lowest energy conformation. The RosettaSurface calculations were run 100,000 times for each sequence to reach the thermodynamic ensemble average.

Hawng et al showed that the 104 surface of calcite is the most stable one [160]. Therefore, the 104 surface was selected for this study.

The Lennard-Jones parameter in the force field used by RosettaSurface was selected from Stockelmann et al. [145, 161].

The free energy landscape of water on the 104 calcite surface was calculated by Kerisit and Parker [162]. In RosettaSurface, the parameter for the solvation free energy evaluation was calculated from the water density profiles given in the Kerisit and Parker study [158].

Atomic charges were taken from a quantum mechanical study [158].

3.7 The Scoring Function

Almost all empirical all-atom molecular mechanics force fields (FFs) are made to represent biological systems and phenomena [163]. FFs are parameterized for a particular application and their parameters are tuned for that one application and usually are not suitable for a different kind of simulation [163].

Force field transferability is the main challenge in the simulations of protein adsorption behavior on an inorganic surface [163]. The challenge arises because of the differences in the driving forces that rule protein conformation behavior in solution, which is the parameter set in the protein FFs such as AMBER94 [164], and those that govern the protein adsorption behavior near the solid surface [163].

Therefore, for protein-surface interaction simulations new parameters are essential and needed to be able to consider adsorption behavior. In protein-inorganic surface interaction simulations, few parameters are developed because of the lack of experimental data [158]. Consequently, few of the existing parameters are developed, which are suitable for use in this kind of simulation [163, 165].

The default Rosetta scoring function, score12 [166], contains both statistical and physical terms [158]. This scoring function was developed to recover the folded structures of proteins.

Score12 is a linear combination of different variables that model varying interactions between atoms. It consists of an implicit solvation [167], a Lennard-Jones interaction, angle-depended hydrogen-bonds [168], short-range electrostatics (knowledge-based), backbone and a side-chain torsion potential extracted from Protein Data Bank (PDB), and unfolded state energy for 20 naturally occurring amino acids as a reference energy [169].

The score12 has been used as a starting point for a protein-inorganic surface interaction scoring function [158]. A distance dependent dielectric coulomb interaction term [170] has been added for considering protein side-chains interaction with charged surface ions [158]. The final scoring function in RosettaSurface is (3.8) [158]:

$$E_{total} = W_{vdW}E_{vdW} + W_{elec}E_{elec} + W_{Hbond}E_{Hbond} + W_{solv}E_{solv} \tag{3.8}$$

Where W_i are weights and E_i indicates energy functions. Table 3.1 shows some of the RosettaSurface scoring terms.

Energy term	**Discription**
fa_elec	Coulomb interactions with distance-dependent dielectric
fa_atr	Lennard-Jones attractive
fa_rep	lennard-jones repulsive
fa_intra_rep	lennard-jones repulsive between atoms in the same residue
fa_sol	Implicit solvation model based
hbond_bb_sc	Orientation-dependent backbone–side-chain hydrogen bonding
hbond_sc	Orientation-dependent side-chain–side-chain hydrogen bonding
hbond_sr_bb	Short-range orientation-dependent backbone–backbone hydrogen bonding
fa_dun	Internal energy of sidechain rotamers

Table 3.1: Rosetta scoring function terms.

In RosettaSurface software the energy function is based on physics and statistics-derived potential terms and it does not match the actual energy units such as kcal/mol.

The energy unit in this software is REU for (Rosetta Energy Unit) and it should only be interpreted in relative terms, i.e. by ranking the affinity of various structures with respect to one another.

3.8 Binding Energy Calculations

The most important thermodynamic quantity in organic-inorganic interactions is the change in the free energy [30]. Our assumption in the Gibbs free energy for such a system is the difference between energy at constant pressure and temperature in solution, where the peptide is far from the surface and in the adsorbed state, where the peptide and the surface are in contact with each other.

Binding energies for 100 complex structures with lowest energy for each peptide were calculated and a histogram with a 0.1 energy step was drawn. The most frequent binding energy value among the top 100 complex structures was selected as a representative binding energy for each peptide. The calculated binding energy along with a Rosetta score for the adsorbed state of each peptide-surface complex was used as a reference to identify the strong and weak binders.

Chapter 4

Results

4.1 Library

All 20 naturally occurring amino acids were grouped into seven different classes to cover all chemical properties. In each group all the possible permutations of amino acids have been considered for producing 4-mer peptides. This classification includes 10,965 different sequences. The 50 ns MD simulation was carried out to produce all possible conformations for each peptide which was 50,000 different conformations for each sequence.

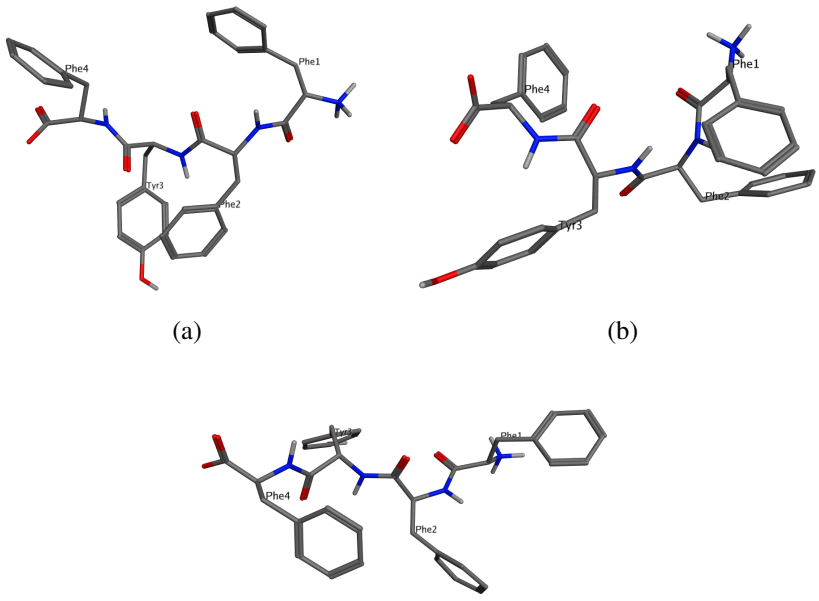
4.2 Clustering

The position of the backbone for all the conformations of a sequence has been identified and similar ones have been clustered; the conformation which was in the middle of the group geometrically was used as a representative for all the members of the group. Table 4.1 shows the number of obtained conformations for each class of amino acids as well as the their abbreviations.

	Amino acids involved	Number of sequences	Number of structures	Number of representative structure after clustering
Acidic	DE	16	800000	50
Basic	RHK	81	4050000	200
Aromatic	WFY	81	4050000	286
Nucleophilic	TSC	81	4050000	322
Special-cases	CGP	81	4050000	328
Polar	NQSTY	625	31250000	2166
Non-polar	MFPWVACGIL	10000	500000000	33636

Table 4.1: The amino acids presents in each group, number of sequences as well as number of structures before and after clustering

Figure 4.1 is an illustration of different conformations of one sequence. These conformations belong to the ar10 peptide with the sequence (Phe,Phe,Tyr,Phe) in the aromatic group.



(a) (b)

(c)

Figure 4.1: Different conformation of peptide ar10 (FFYF).

4.3 The Shape Complementarity

The shape complementarity between selected representative conformations of peptides (from clustering) and the calcite 104 surface was used as a filtration step to narrow down the number of representatives to a manageable number. The Wigner-D function algorithm was used to perform this comparison.

This algorithm compares the shape of two molecules based on their RMSD and shape similarity scores. All the different conformations obtained from clustering have been subjected to this calculation.

The ranges of similarity for different groups are shown in Table 4.2. The similarity value (SV) is between 0 and 0.5. Identical structures have a 0.5 value. In our case the structures were not expected to be very similar as we were comparing crystalline surface of calcite 104 and the 3D structure of a peptide.

	max SV	min SV
Polar	0.147236	0.011644
Aromatic	0.147236	0.018967
Basic	0.130216	0.01844
Non-polar	0.129694	0.007801
Special cases	0.081105	0.008555
Acidic	0.064191	0.016455
Nucleophilic	0.059446	0.016451

Table 4.2: The range of similarity value (SV) between calcite 104 surface and peptides in different groups.

Polar and aromatic groups showed the best similarity compared to the other groups. For each group, except the non-polar, 10% of the peptides which had the highest similarity to the calcite 104 surface were selected. The top 1% of peptides similar to calcite 104 surface in the non-polar group were selected to obtain a reasonable number for representatives.

4.4 RosettaSurface

One percent of the non-polar sequences and ten percents of the aromatic, acidic, basic, nucleophilic, special-cases, and polar groups were selected and for each sequence only one conformation was used. The selected conformations had the most shape similarity with the calcite 104 surface.

These selected sequences were docked to the calcite 104 surface using the RosettaSurface software package.

These calculations gave the structural details of the peptides binding to the surface and the scoring function for the system. The scoring function for each sequence in solution and for adsorbed states was calculated for all the 100,000 complex structures.

The sequences with more than 60% negative energy in their adsorbed state during the simulation were considered as potential binders to the surface; this factor guaranteed to lead to a high affinity to the surface. For each sequence, 100 structures with the best score were selected for analyzing the peptide adsorption behavior on the calcite 104 surface.

In Fig. 4.2, the darker column shows the percentage of sequences with at least one pose with a negative energy during the simulation and the brighter column shows the percentage of sequences with more than 60% negative energy during the simulation.

Tables (4.3,4.4,4.5,4.6) show the ranges of energy and percentage for negative poses among the 100 structures with the lowest scoring functions for potential strong binders in each group.

Non-polar, aromatic and acidic groups did not meet the conditions for strong binders. For the aromatic group, the best negative energy belonged to a sequence with -1.475 REU and only 0.3% of its selected structures showed negative energy. In the acidic group, the best energy was -2.151 REU and 26% of its selected structures had negative energy. These numbers for the non-polar group are -1.321 REU

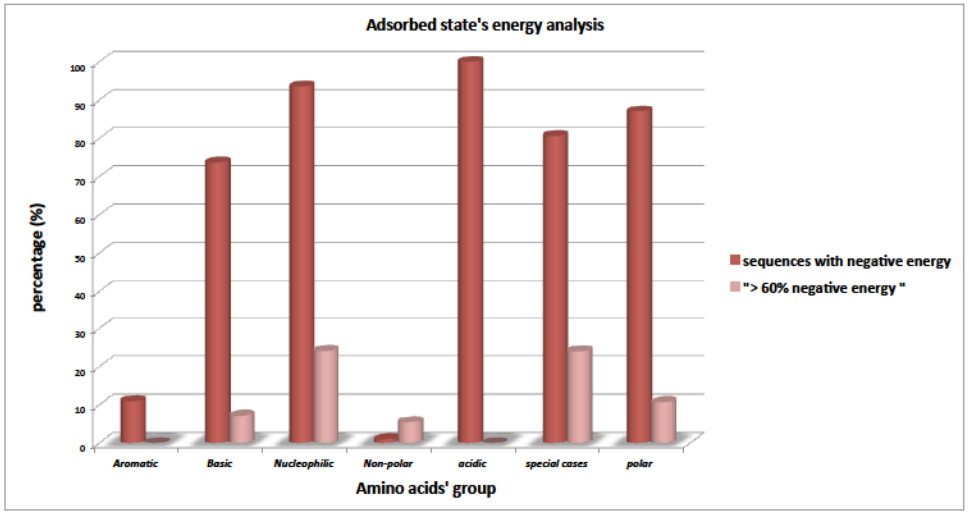


Figure 4.2: Negative energy poses analysis; darker column:percentage of sequences with at least one pose of negative energy. lighter column:percentage of sequences with more than 60% of negative energy poses.

special-cases	negative energy(%)	min adsorbed state energy(REU)	max adsorbed state energy(REU)
s41	99.83	-3.454	1.04
s51	98.86	-3.086	2.795
s33	94.55	-3.041	1.15
s15	94.48	-2.826	2.602
s24	94.71	-2.82	4.372
s39	88.99	-1.819	2.978

Table 4.3: Special-cases potential strong binders.

Nucleophilic	negative energy(%)	min adsorbed state energy(REU)	max adsorbed state energy(REU)
nu67	85.64	-4.754	1.902
nu76	59.71	-2.936	1.925
nu66	99.76	-2.928	1.978
nu1	62.07	-2.306	1.983
nu24	98.18	-2.178	3.681
nu10	76.05	-2.017	1.283
nu42	94.66	-1.017	2.294

Table 4.4: Nucleophilic potential strong binders.

Basic	negative energy(%)	min adsorbed state energy(REU)	max adsorbed state energy(REU)
b63	84.69	-3.337	4.332

Table 4.5: Basic potential strong binders.

Polar	negative energy(%)	min adsorbed state energy(REU)	max adsorbed state energy(REU)
p19	71.02	-4.599	1.689
p266	97.26	-4.567	2.44
p113	99.78	-4.445	1.302
p104	90.09	-4.130	3.149
p34	66.75	-3.656	2.435
p243	67.16	-3.561	2.846
p294	62.32	-3.448	2.233
p94	75.24	-3.365	5.485
p87	77.16	-3.364	2.438
p509	63.00	-3.065	3.106
p530	62.93	-2.634	4.34
p392	62.87	-2.351	2.621
p293	71.61	-2.264	1.778
p324	79.24	-2.007	6.359

Table 4.6: Polar potential strong binders.

and 1%.

The sequences with more than 60% negative energy during their simulation were considered as potential binders to the surface; this factor guaranteed to lead to a high affinity to the surface.

4.5 Binding Energy

As mentioned in Chapter 2, the most important thermodynamic feature in organic-inorganic interactions is the change in the free energy. The most promising approach for determining binding energy is the energy difference in the adsorbed and solution states.

For calculating the binding energy, the average of the energy for the peptide and surface in solution state for the selected poses was subtracted from the total energy for each of the selected poses in their adsorbed states. A histogram of these data was plotted to obtain the most accurate representative binding energy for each sequence. The graphs below show these results for the polar group.

The binding energy for potential binders in the polar group was calculated as -4.2 REU for p19, p266 and p113; -3.6 REU for p509; -3.5 REU for p243; -3.3 REU for p37 and p87; -3.2 REU for p324; -3.1 REU for p294; -2.9 REU for p34; -2.4 REU for p94 and p104; -2.3 REU for p293 and -2.2 REU for p530.

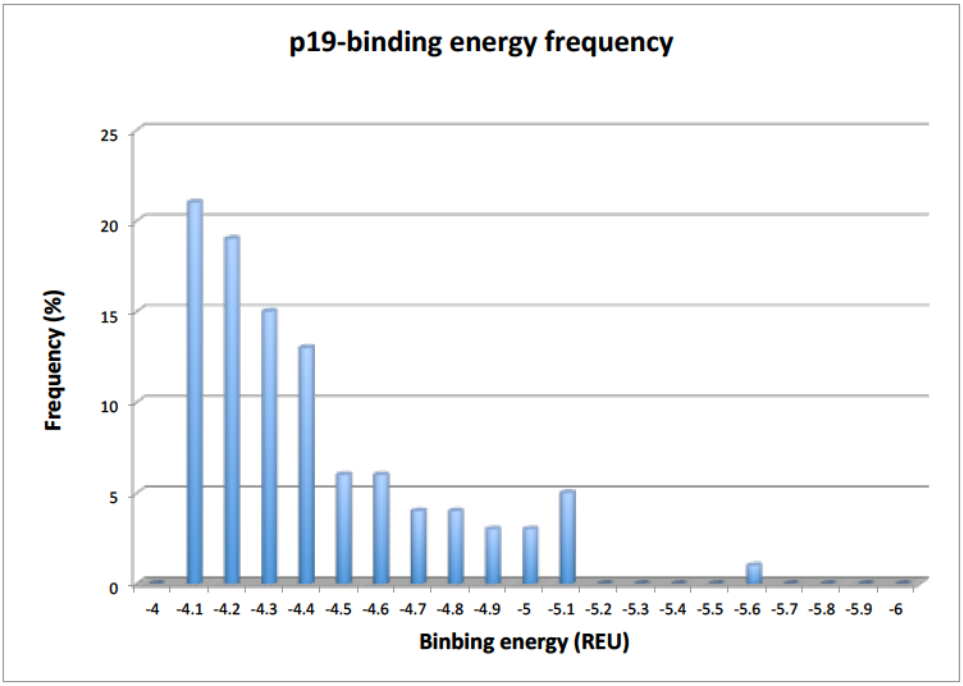


Figure 4.3: p19-Binding energy frequency in 100 lowest energy poses.

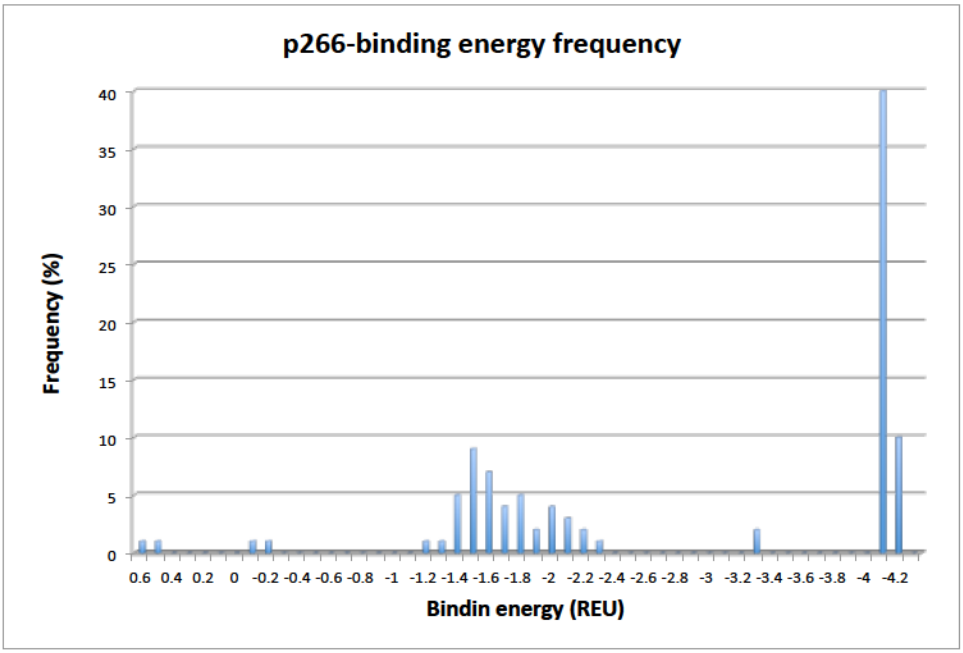


Figure 4.4: p266-Binding energy frequency in 100 lowest energy poses.

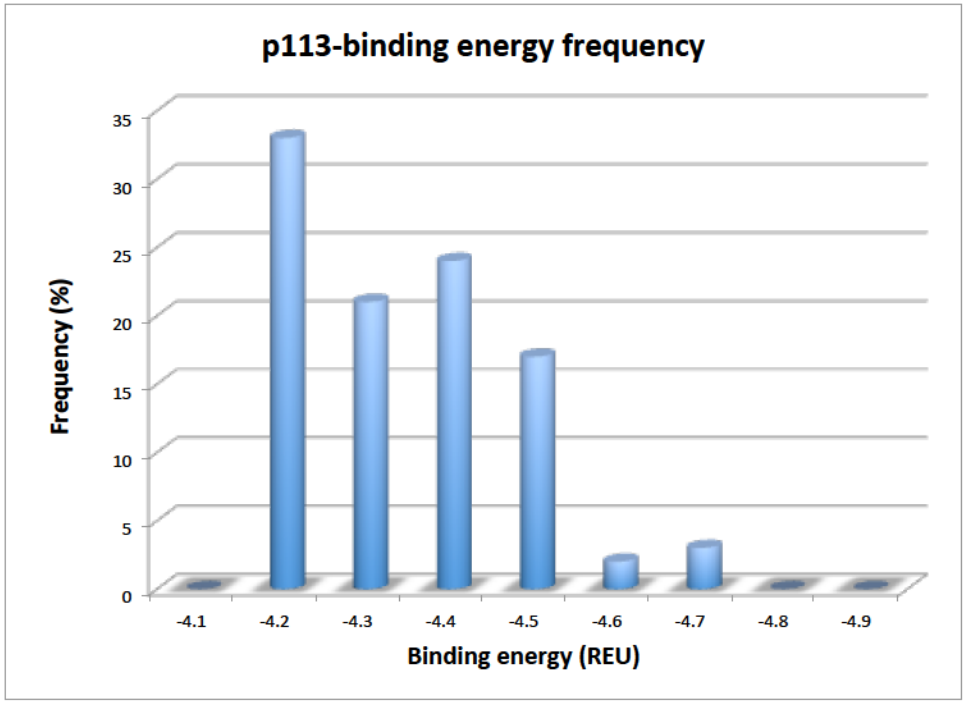


Figure 4.5: p113-Binding energy frequency in 100 lowest energy poses.

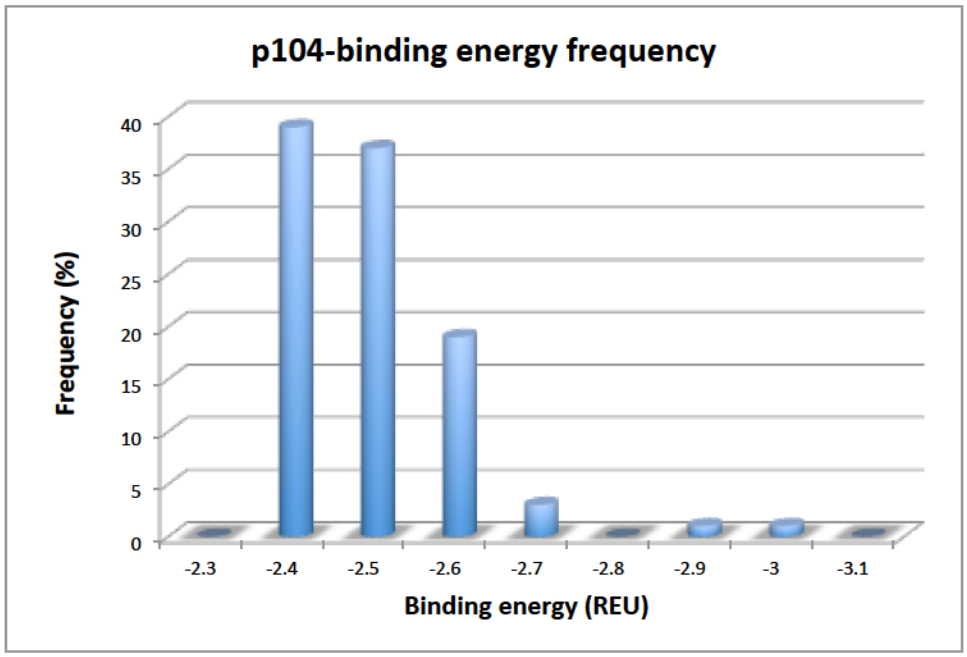


Figure 4.6: p104-Binding energy frequency in 100 lowest energy poses.

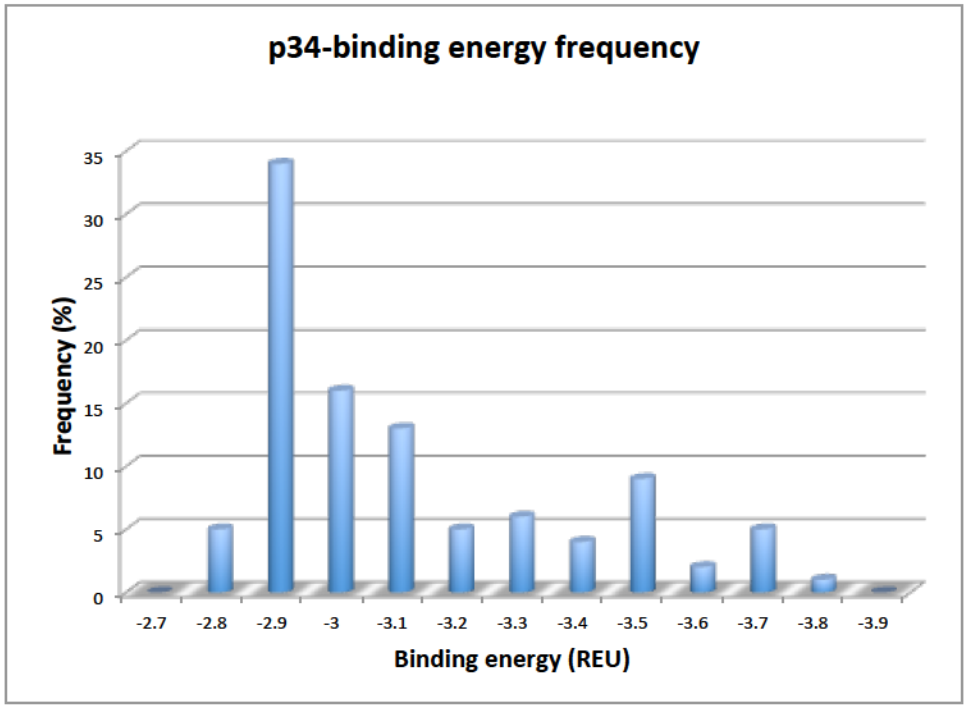


Figure 4.7: p34-Binding energy frequency in 100 lowest energy poses.

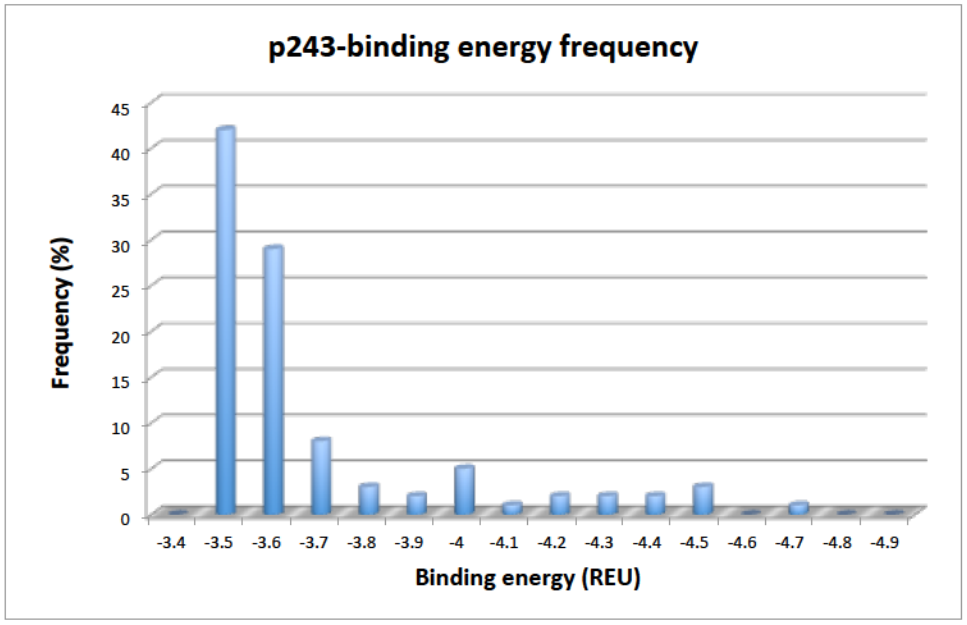


Figure 4.8: p243-Binding energy frequency in 100 lowest energy poses.

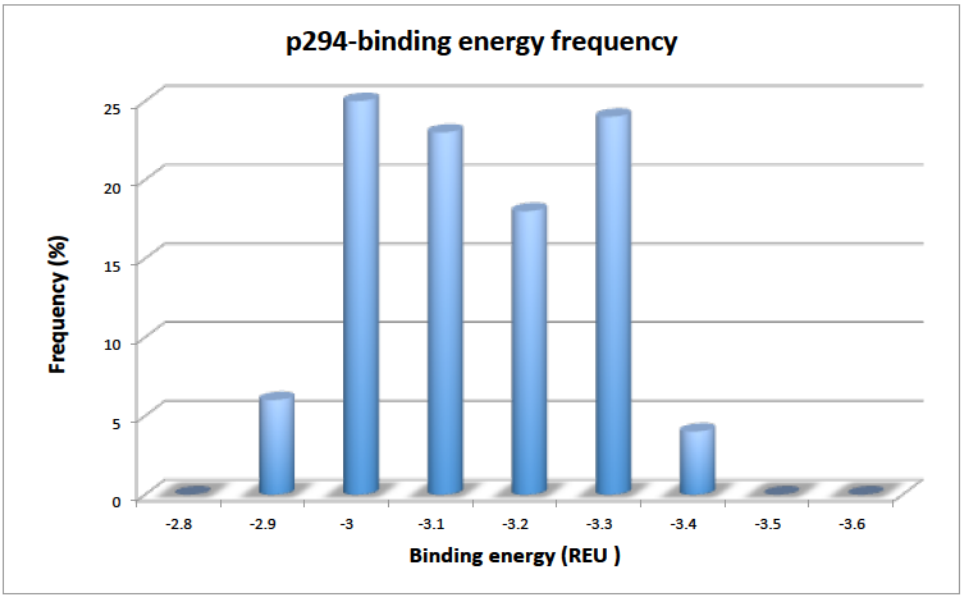


Figure 4.9: p294-Binding energy frequency in 100 lowest energy poses.

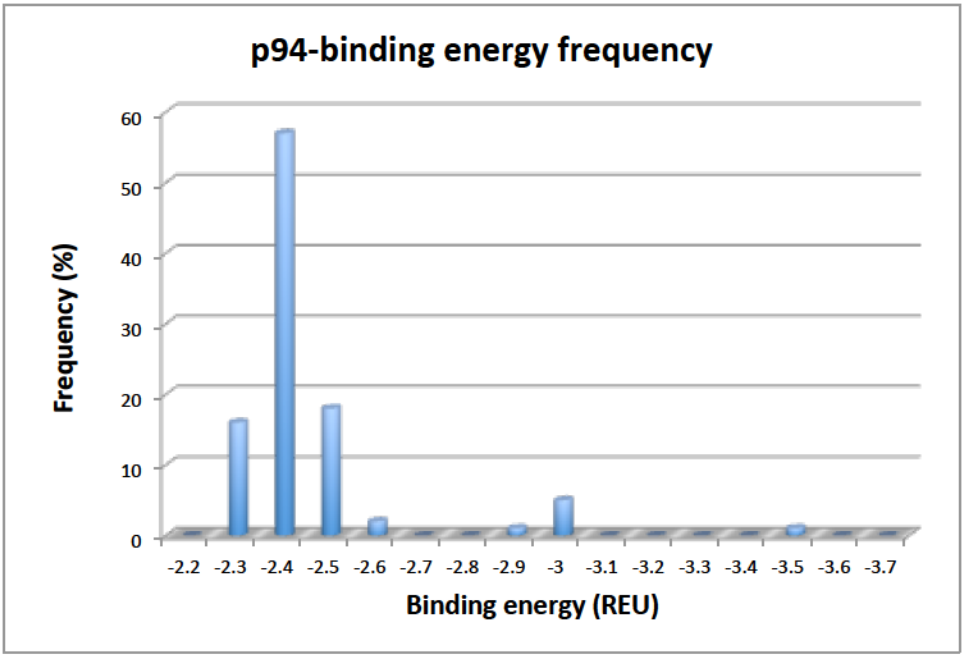


Figure 4.10: p94-Binding energy frequency in 100 lowest energy poses.

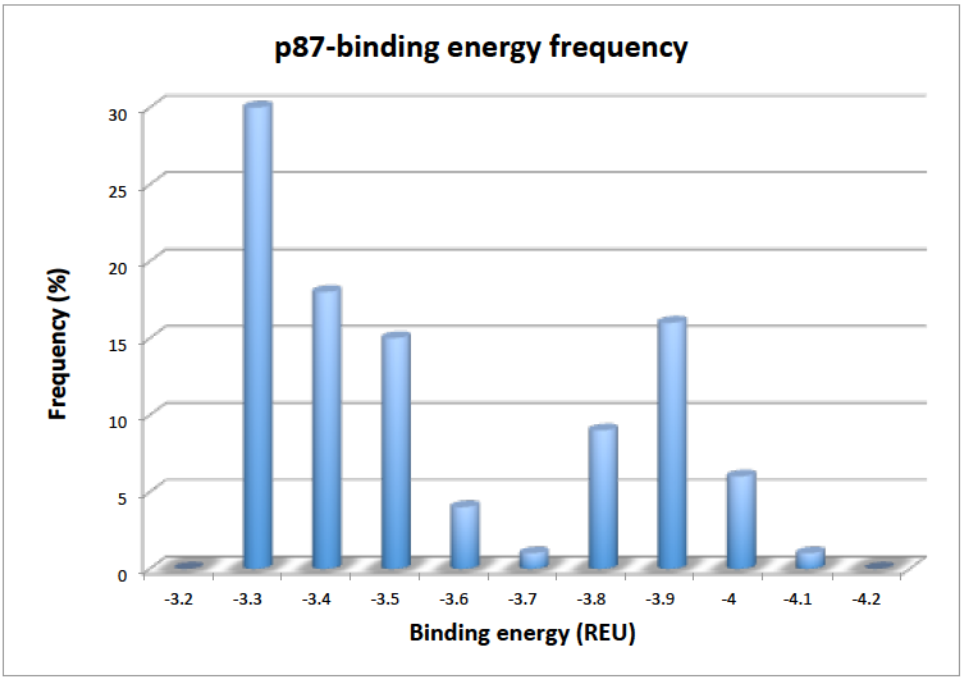


Figure 4.11: p87-Binding energy frequency in 100 lowest energy poses.

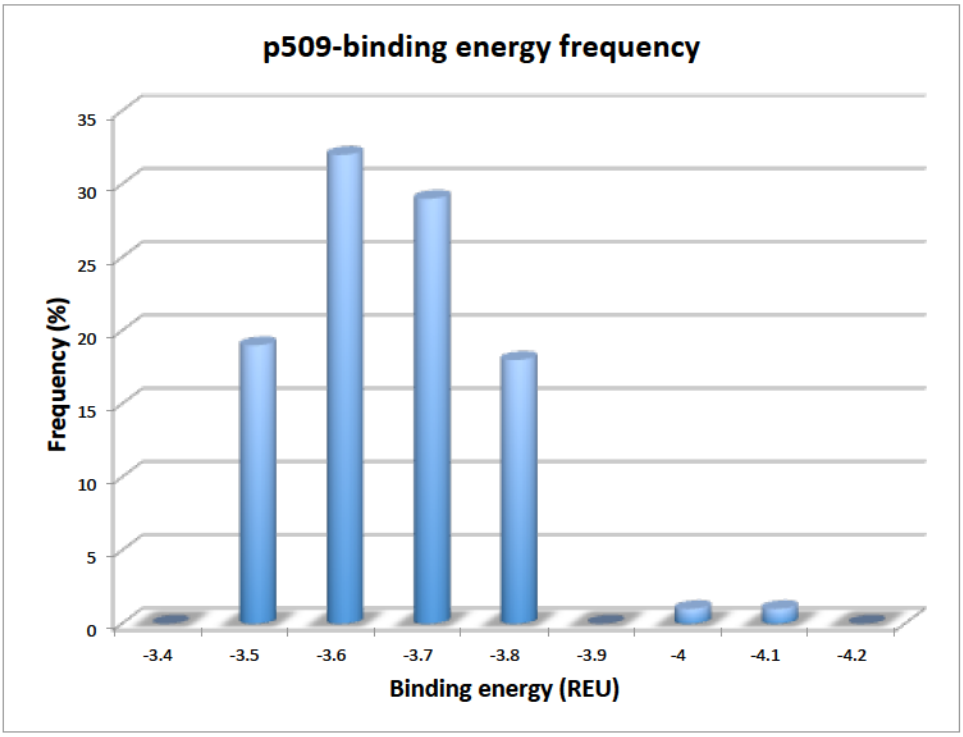


Figure 4.12: p509-Binding energy frequency in 100 lowest energy poses.

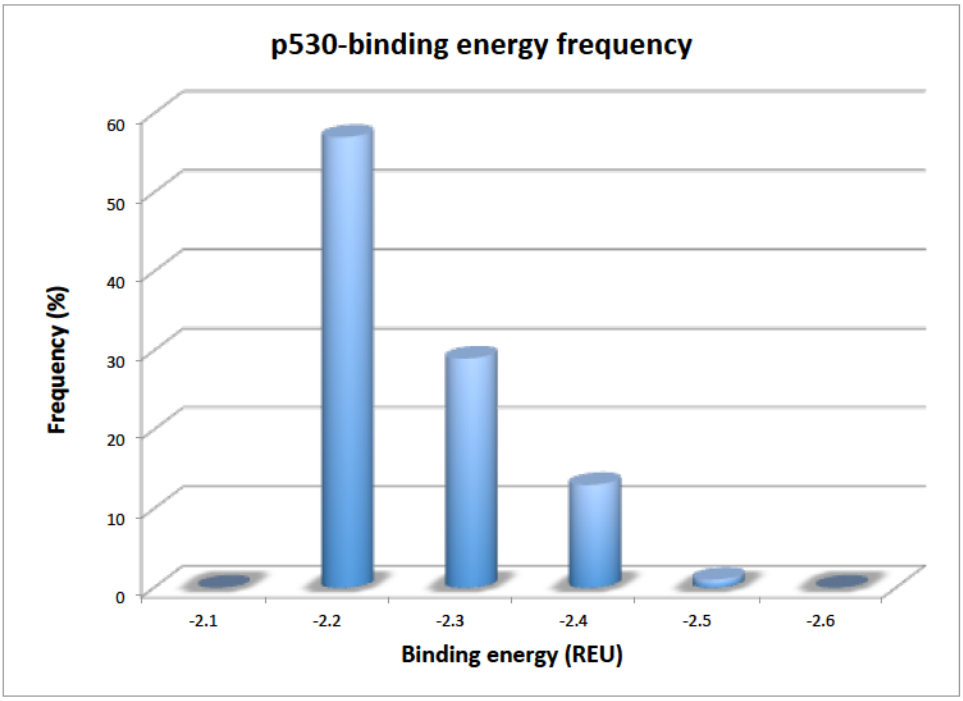


Figure 4.13: p530-Binding energy frequency in 100 lowest energy poses.

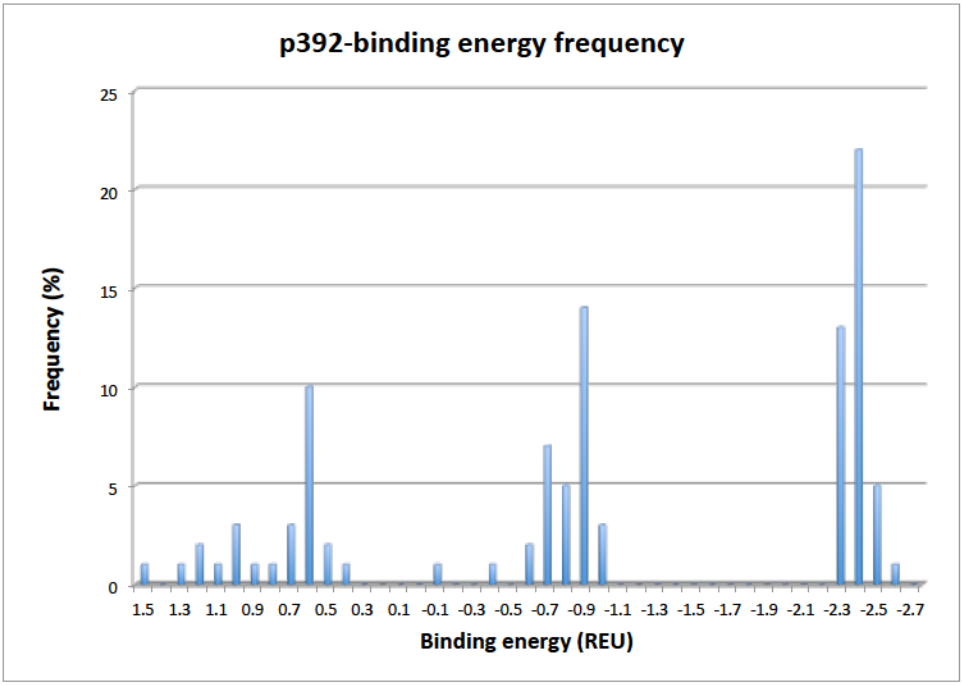


Figure 4.14: p392-Binding energy frequency in 100 lowest energy poses.

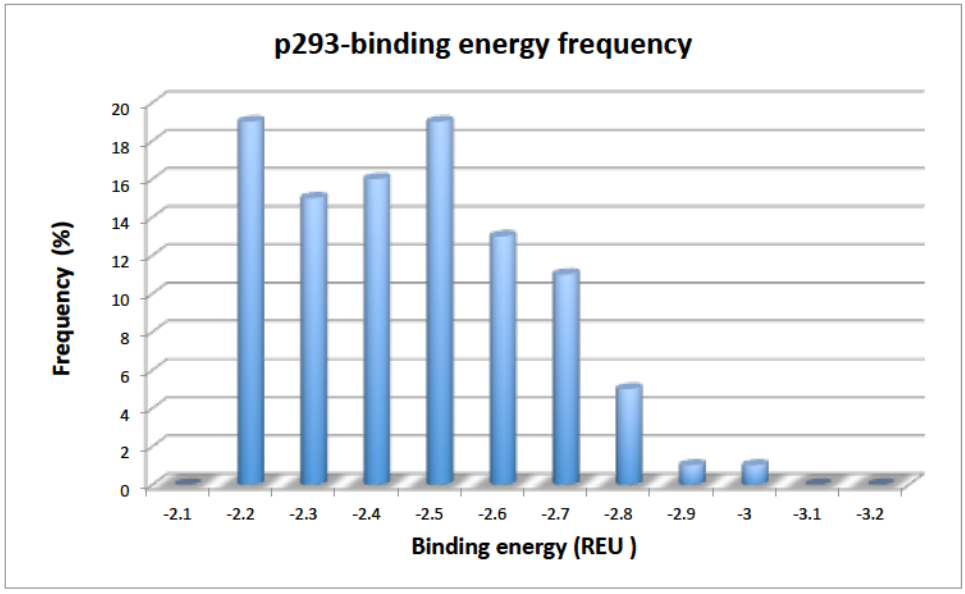


Figure 4.15: p293-Binding energy frequency in 100 lowest energy poses.

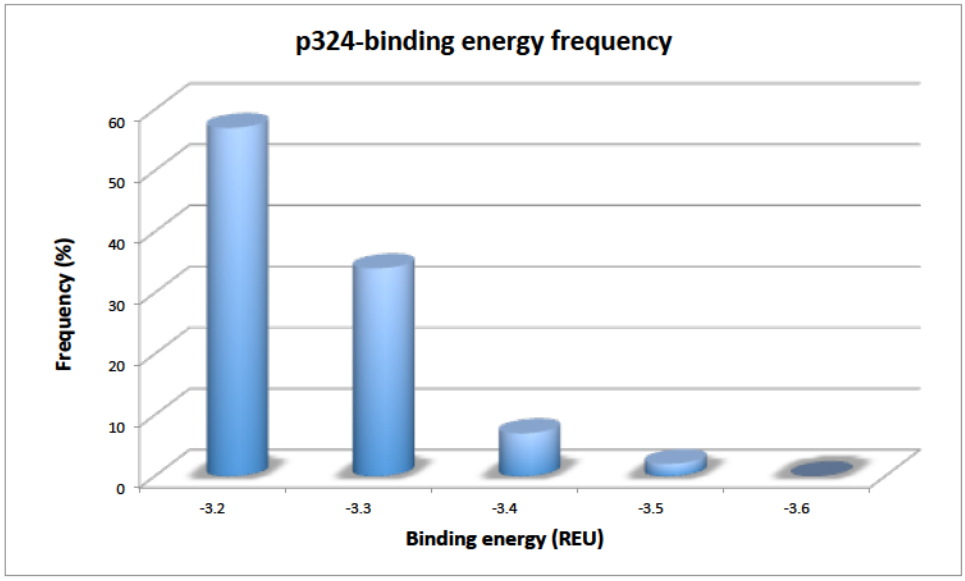


Figure 4.16: p324-Binding energy frequency in 100 lowest energy poses.

The results of this calculation for the only basic potential binder are shown in Fig 4.17. For b63 the binding energy was calculated as -2.2 REU.

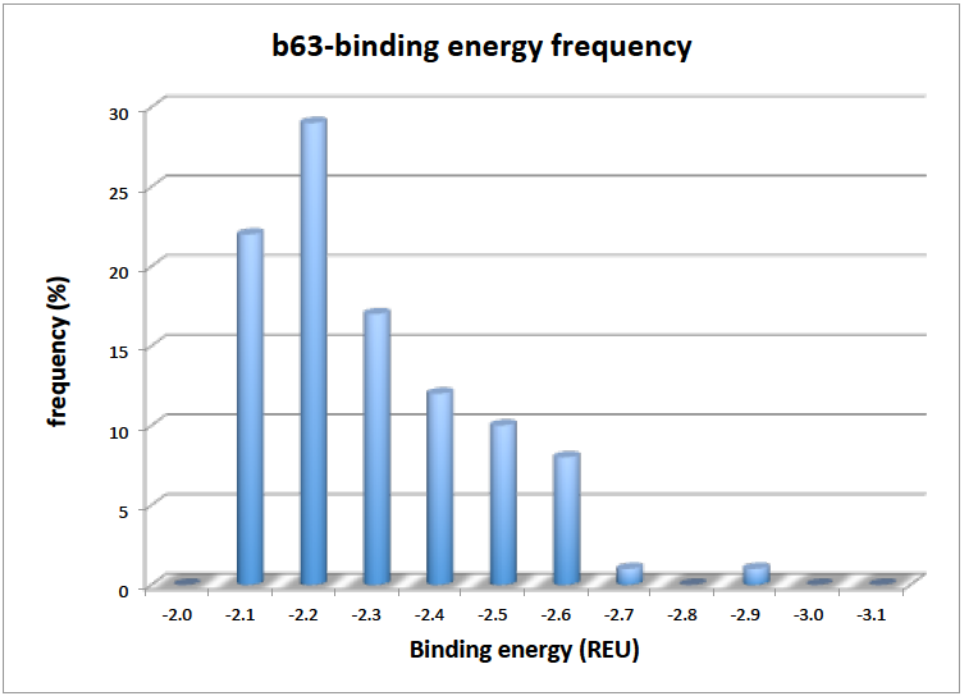


Figure 4.17: b63-Binding energy frequency in 100 lowest energy poses.

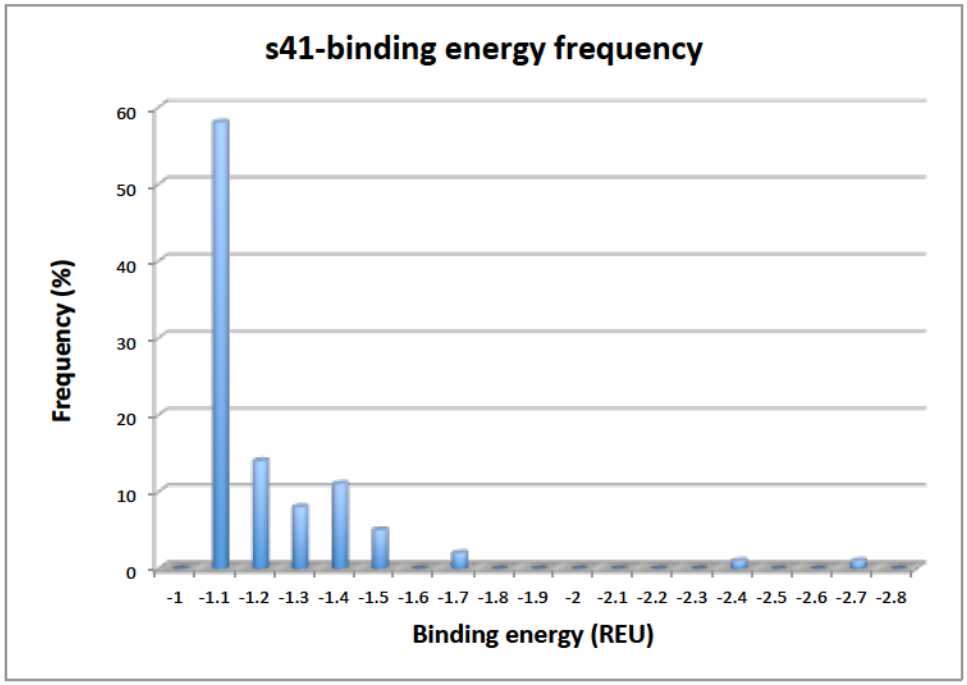


Figure 4.18: s41-Binding energy frequency in 100 lowest energy poses.

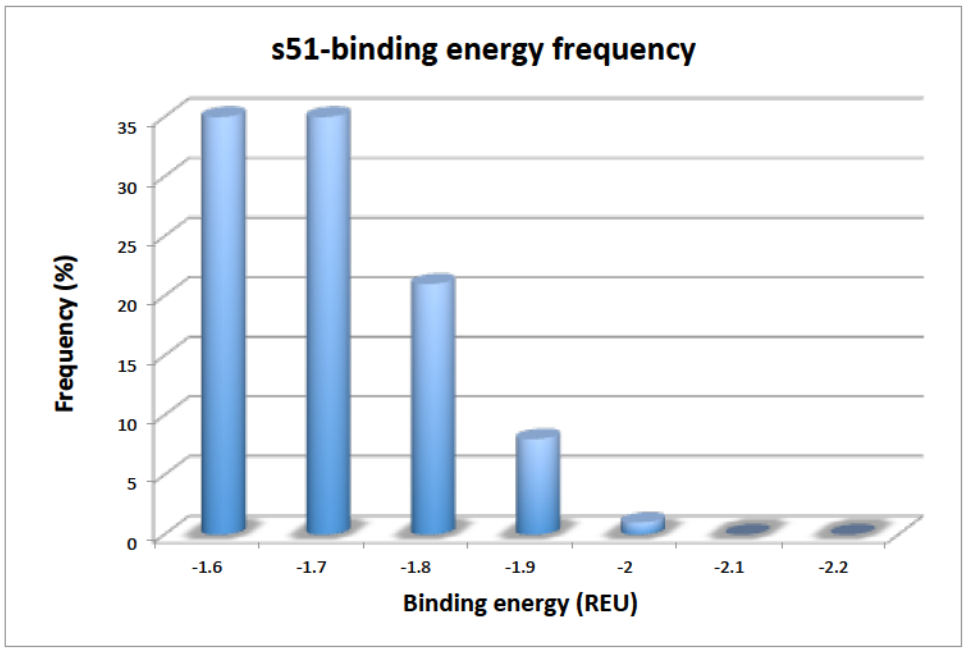


Figure 4.19: s51-Binding energy frequency in 100 lowest energy poses.

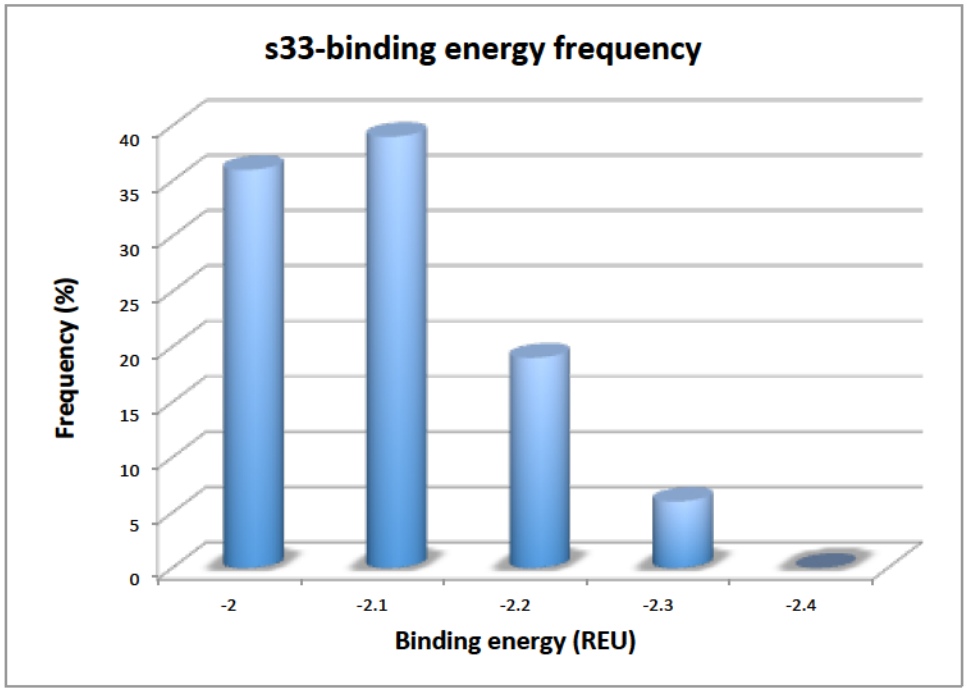


Figure 4.20: s33-Binding energy frequency in 100 lowest energy poses.

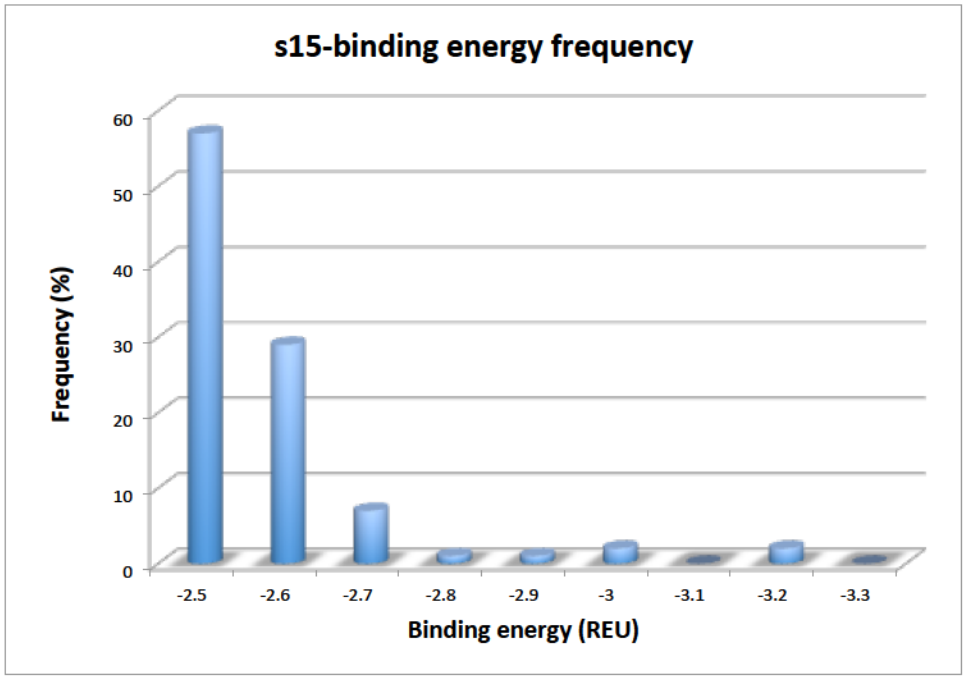


Figure 4.21: s15-Binding energy frequency in 100 lowest energy poses.

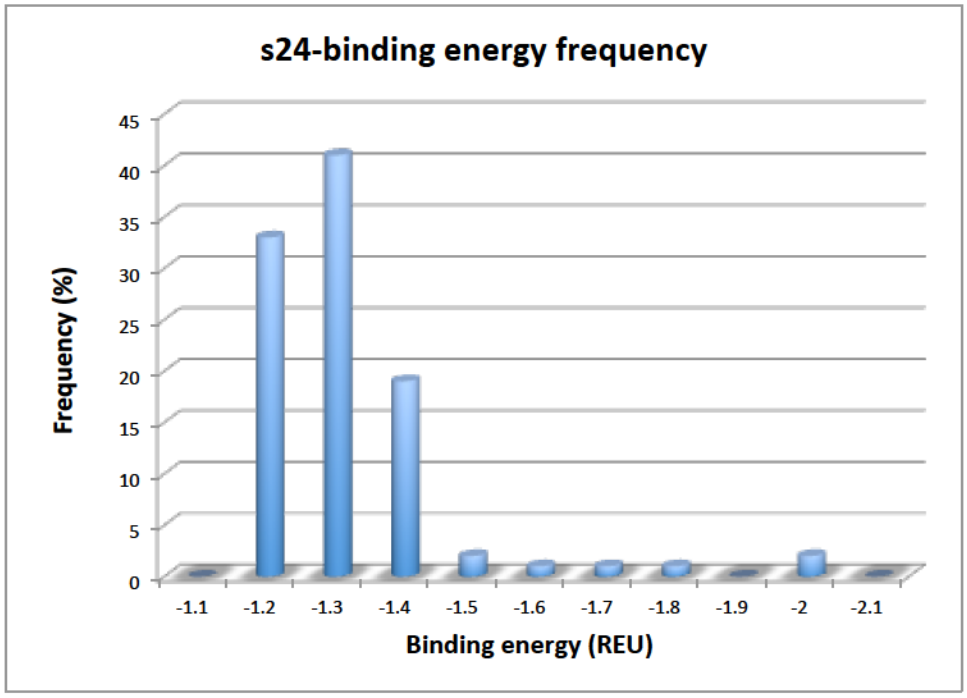


Figure 4.22: s24-Binding energy frequency in 100 lowest energy poses.

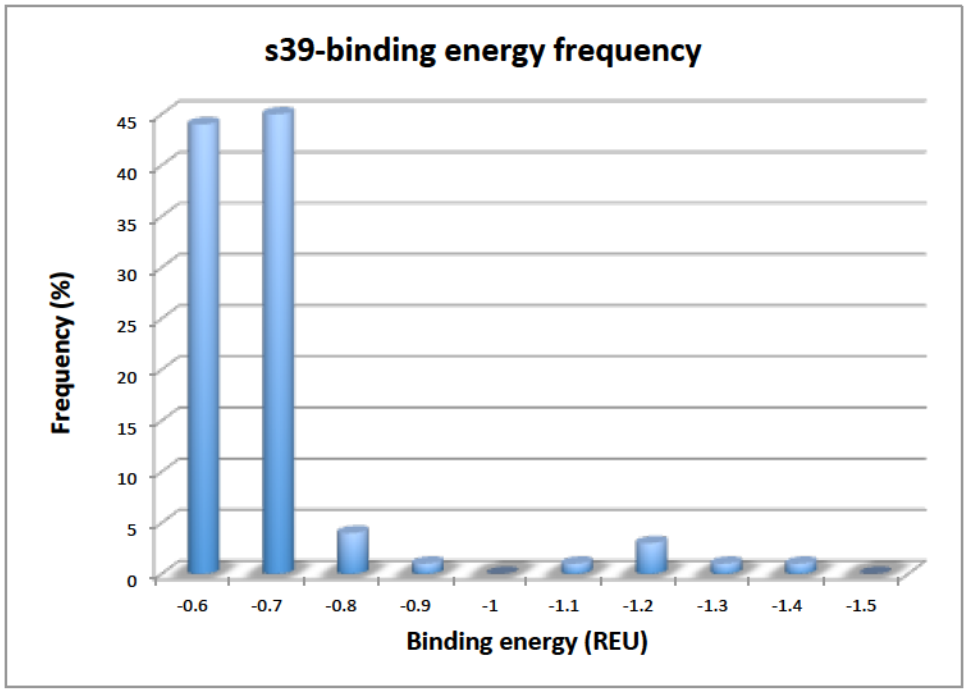


Figure 4.23: s39-Binding energy frequency in 100 lowest energy poses.

The binding energy for the special-cases group was calculated as: -2.5 REU for s15; -2.3 REU for s33; -1.7 REU for s51; -1.3 REU for s24; -1.1 REU for s41 and -0.7 REU for s39.

The strong binders were chosen based on the binding energy. The cut-off for strong binders' binding energy was assumed to be -3 REU. Table 4.7 lists the name, energy, and amino acid components of these sequences.

Strong binders	binding energy	sequence
p266	-4.2	NTNS
p113	-4.2	SSYN
p19	-4.2	TTNN
p509	-3.6	TQNY
nu67	-3.5	STTC
P87	-3.3	QSTN
P37	-3.3	QSQN
p324	-3.2	TYSS

Table 4.7: The predicted strong binding peptides.

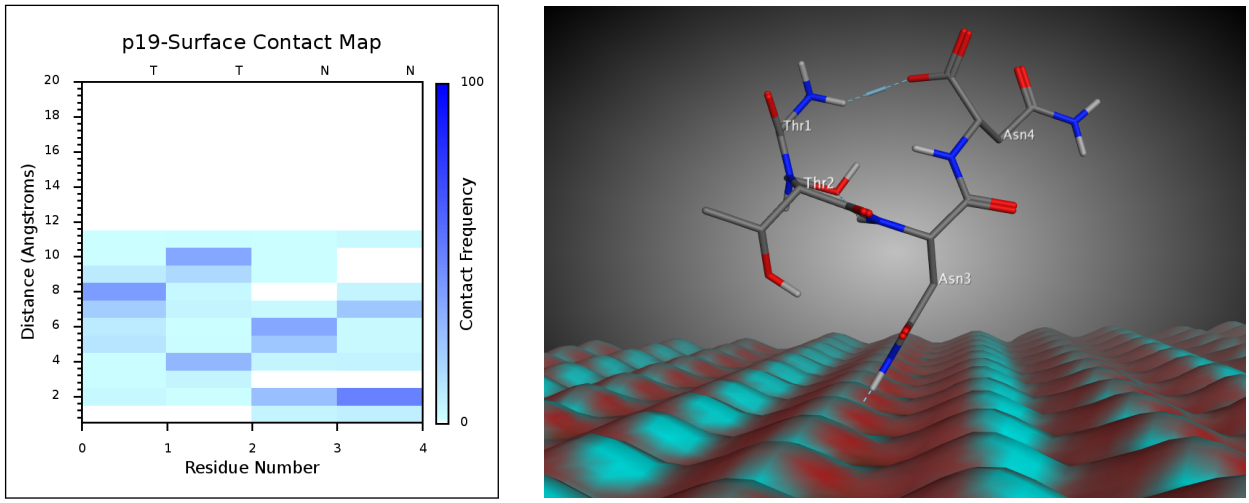
The highlighted sequences were reported to wet lab experts in the Ingenuity Lab at the University of Alberta, to validate our calculations. The summarized results show that most of the strong binders came from the polar groups. Table 4.8 shows the values for each term of the energy for strong binders.

	Total	fa_atr	fa_dun	fa_elec	fa_intra_rep	fa_rep	fa_sol	hbond_bb_sc	hbond_sc	hbond_sr_bb
p324	-2.007	***-9.194***	1.823	***-3.861***	0.041	0.947	10.490	0	-1.371	-1.635
p266	-4.567	***-10.310***	2.715	***-3.564***	0.024	0.327	11.630	-1.692	-1.213	-1.029
p113	-4.445	***-8.989***	3.597	***-4.831***	0.036	0.322	9.390	-1.292	-0.890	-0.992
p87	-3.364	***-7.050***	4.060	***-3.160***	0.023	0.165	7.757	-1.937	-0.874	0
p37	-4.034	***-7.837***	4.933	***-4.266***	0.025	0.661	8.593	0	-0.889	-2.127
p19	-4.599	***-4.198***	3.422	***-6.018***	0.019	0.309	5.148	-0.268	0	-1.129
p509	-3.065	***-10.300***	4.834	***-3.983***	0.047	0.179	10.600	-2.042	-0.533	0
nu67	-3.601	***-5.067***	0.189	***-3.532***	0.012	0.196	5.736	-0.220	-0.101	-0.988

Table 4.8: A breakdown of energy contribution for a selection of strong binders.

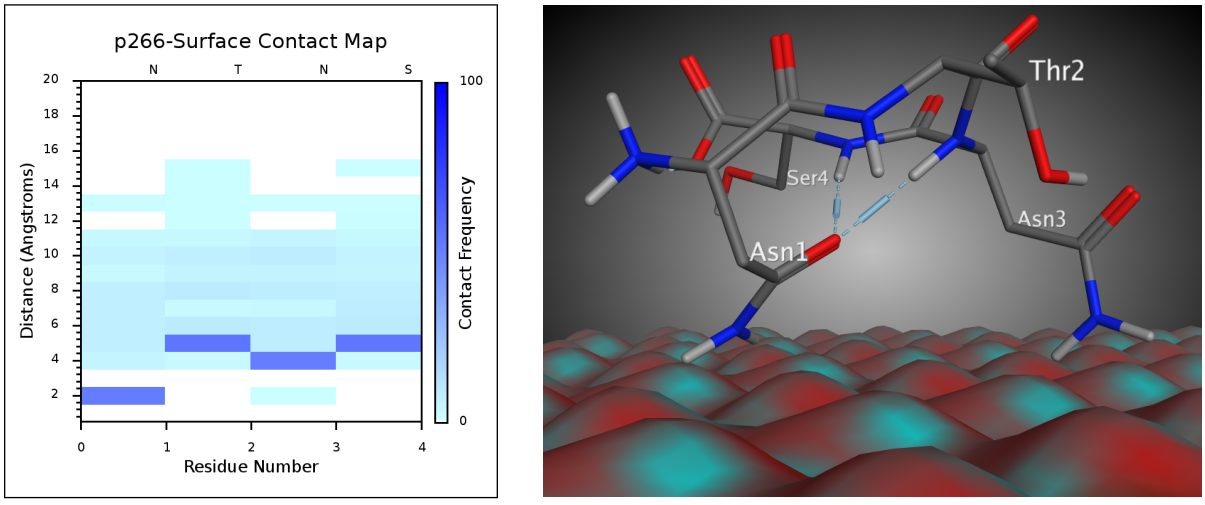
The main contributions in the calculated energy belong to the fa-atr (Lennard-Jones attractive between atoms in different residues) and the fa-elec (Coulombic electrostatic potential with a distance-dependent dielectric) groups.

A protein-surface contact map graph provides information about contacts between each residue in the sequence and the calcite 104 surface for the 100 top adsorbed state structures. The y-axis shows the perpendicular distance between each residue and the surface in angstrom. Different colors in these maps show different frequencies of the presence of each residue in a particular position.



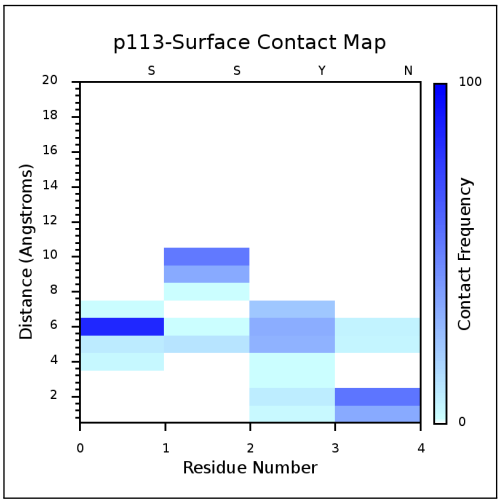
(a) (b)

Figure 4.24: a:p19 surface contact map plot. b:p19 dominant binding pose.

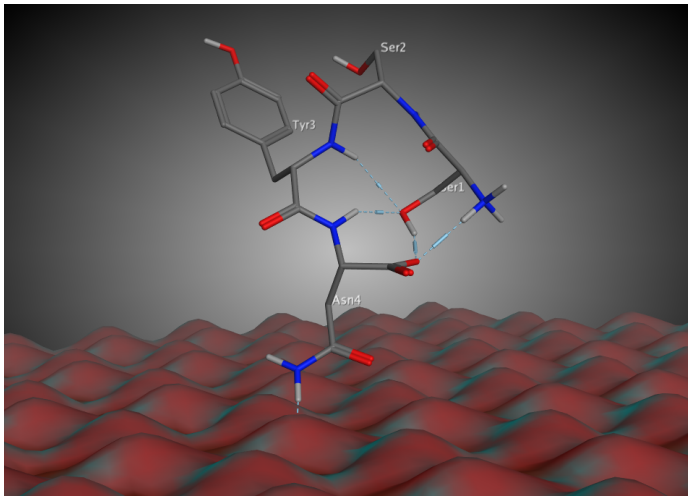


(a) (b)

Figure 4.25: a:p266 surface contact map plot. b:p266 dominant binding pose.

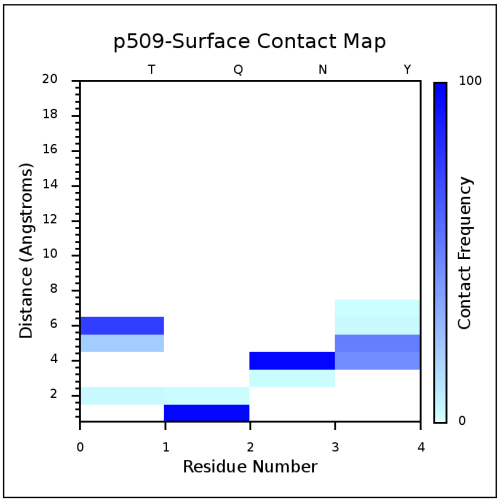


(a)

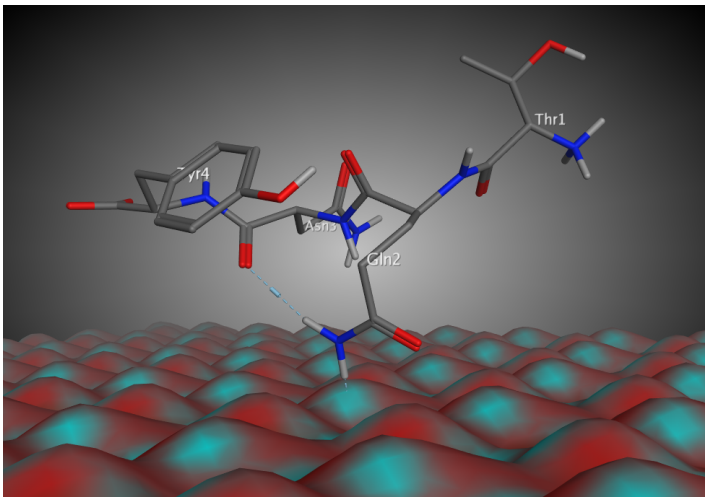


(b)

Figure 4.26: a:p113 surface contact map plot. b:p113 dominant binding pose.

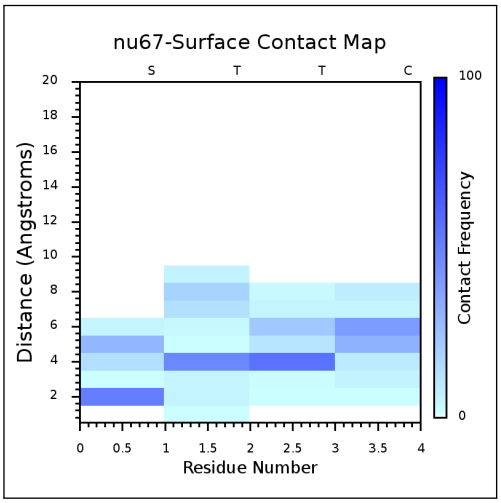


(a)

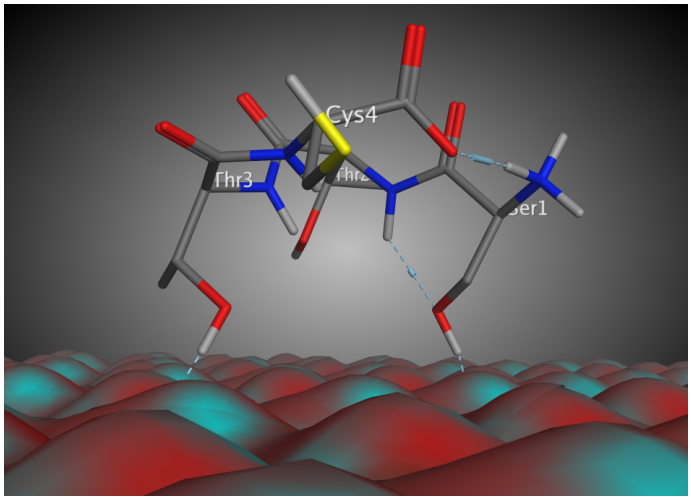


(b)

Figure 4.27: a:p509 surface contact map plot. b:p509 dominant binding pose.

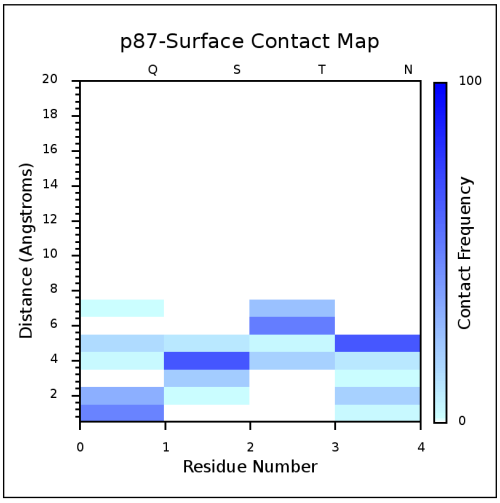


(a)

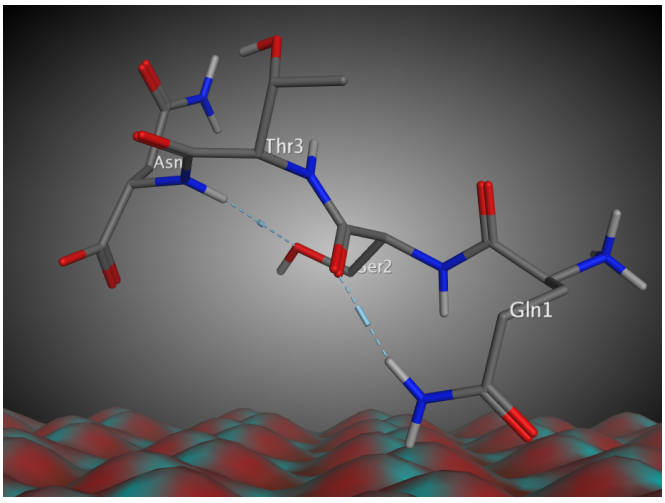


(b)

Figure 4.28: a:nu67 surface contact map plot. b:nu67 dominant binding pose.

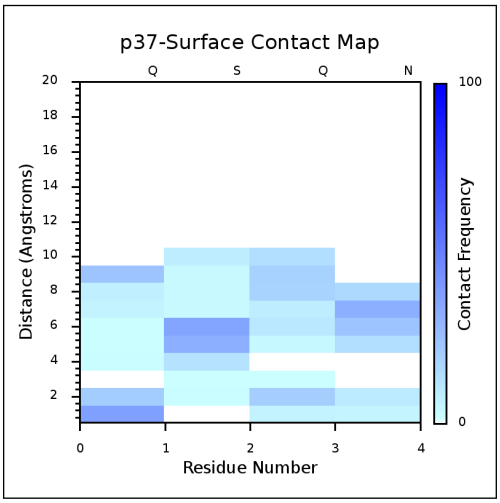


(a)

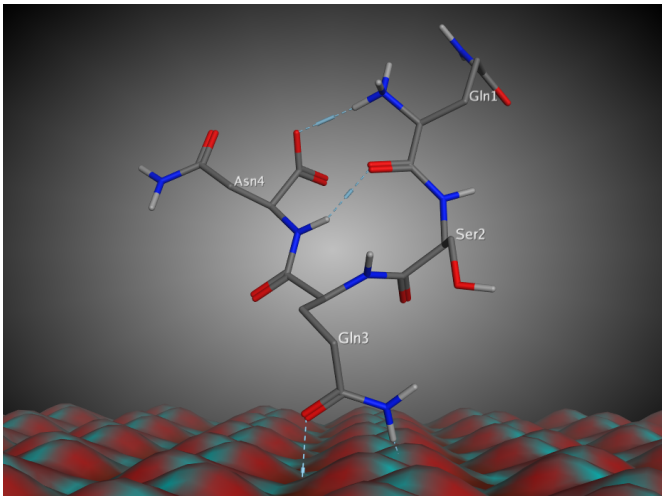


(b)

Figure 4.29: a:p87 surface contact map plot. b:p87 dominant binding pose.

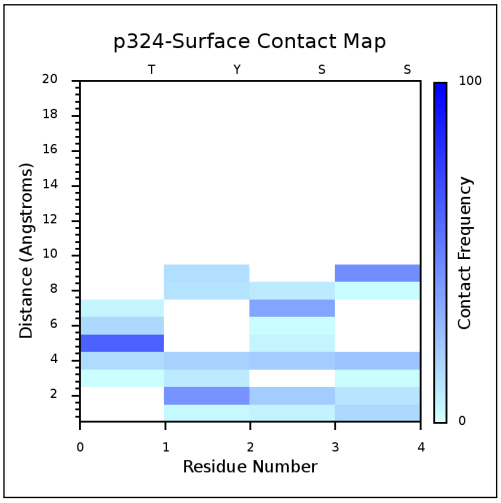


(a)

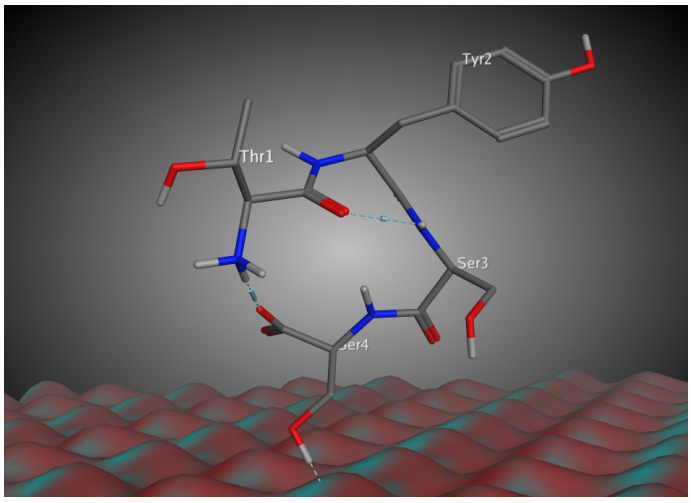


(b)

Figure 4.30: a:p37 surface contact map plot. b:p37 dominant binding pose.



(a)



(b)

Figure 4.31: a:p324 surface contact map plot. b:p324 dominant binding pose.

In p19 (TTNN), most of the poses among the top 100 poses during simulation showed no intramolecular interactions other than the N- and C-terminus ionic interaction.

In p266 (NTSN), a side-chain carbonyl of N interacts with NH of the backbone of residues at position three and four.

In p113 (SSYN), the backbone of N (NH) interacts with the O of serine at position one. Also the O of serine at position one interacts with the NH backbone of residue two.

In p509 (TQNY), no interaction other than the inter-terminal ionic interaction has been detected.

In p37 (QSQN), other than terminal, carbonyl (O) of the residue one interacts with the NH of the backbone of residue four.

In case of p87 (QSTN), the NH side-chain of glutamine interacts with the O of the backbone of serine and the O of the side-chain of serine interacts with the H of the backbone of asparagine.

In Nu67 peptide we have identified the terminal ionic interaction and the interaction between the backbone of Threonine (NH) at position two and the side-chain of serine (O).

Figure 4.32 shows a flowchart of the method and the way we selected the eight strong binders from among the $\sim$11000 different sequences.

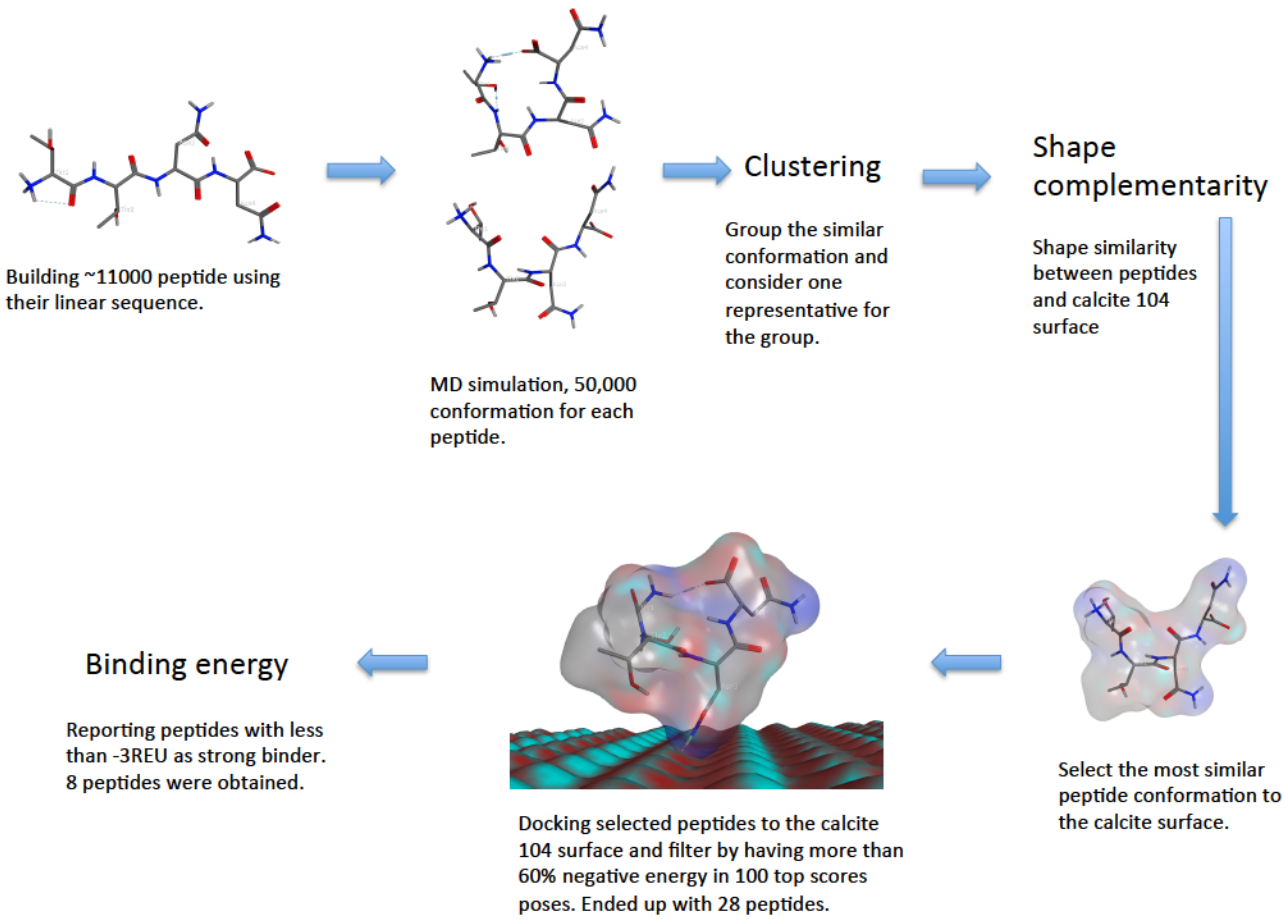


Figure 4.32: method flowchart for selection the peptides with high affinity and specificity to inorganic material.

Chapter 5

Experimental Validation

5.1 Methods[1]

5.1.1 Peptide synthesis

All amino acids [THR(tBu), TYR(tBu), ASN, SER, PRO, TRP(Boc), PHE, MET, ILE, CYS(Trt)] and preloaded resins [Fmoc-Ser(tBu), Fmoc-Asn(Trt)-, Fmoc-Cys(Trt), Fmoc-Trp(Boc), Fmoc-Phe, Fmoc-Leu Wang resins] were purchased from AAPPTec LLC. (Louisville, KY, USA) and used as received. An automated solid-phase peptide synthesizer (AAPPTec Focus XC) was used to prepare the required sequences following standard batch-wise Fmoc chemistry. Peptide sequences (see Table 5.1) were synthesized executed on their respective preloaded Wang resins employing HBTU activation chemistry. 20% pipiridine in DMF was used for deprotection. Synthesized sequences were removed from resin with 95:2.5:2.5 TFA:TIS:DI-H2O cleavage cocktail while shaking for an hour. Leaving the resins out via filtration, peptides were precipitated in 50 ml of cold ether, centrifuged at 9000 rpms for 8 minutes decanting the remaining solution, continued by re-dissolving them in a 5ml solution (1:1 ACN:H2O) and proceeded finally by lyophilization over 2 days (Labconco FreeZone 4.5 plus). Purification of all peptides was achieved through reverse-phase HPLC (AAPPTec UV3000 and two P3000) working with a Spirit Peptide 120 C18 10m 25×2.12 column (AAPPTec). Finally, peptide purities were analyzed with a LC/MSD (Agilent 1100 Series) using a Zorbax SB-C18 5m4.6×250mm (Agilent) analytical column and their mass was determined with mass spectrometer

[1]The experiments have been done at Ingenuity Lab at University of Alberta. The method detail is given by Dr. Sibel Cetinel.

Table 5.1: Peptides ID and sequences

Peptide ID	Peptide Sequence
p266	NH_2-NTNS-COOH
p113	NH_2-SSYN-COOH
p19	NH_2-TTNN-COOH
np6688	NH_2-PWFF-COOH
np4138	NH_2-PWFW-COOH
np4346	NH_2-MLIL-COOH
np4138	NH_2-PICL-COOH

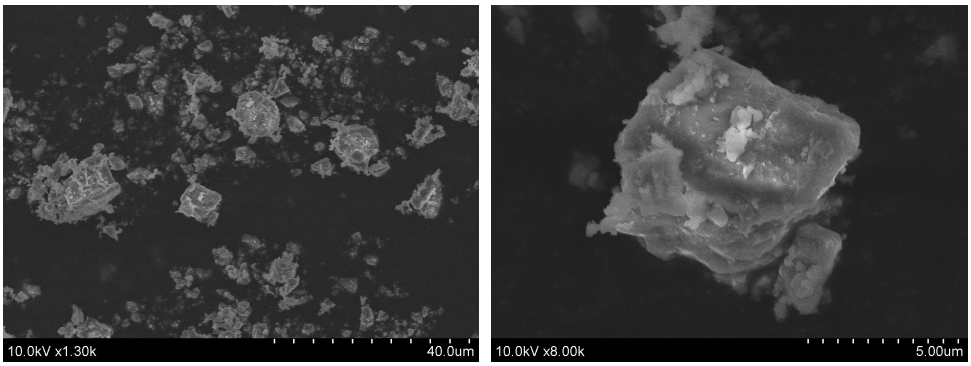
set to positive-ion mode.

5.1.2 Calcite Powder Preparation

Calcite mineral specimen ($6 \times 2 \times 1.5$) cm was received from the mine Durango, Mexico, through John Betts, Fine Minerals, NY (figure 5.1). Mineral was grinded in Mortar Grinder RM200 (Retsch Ltd.) at Thin Section Lab., Earth and Atmospheric Sciences, University of Alberta. Grinded mineral was filtered through a sieve shaker using a 106m sieve. Powder containing smaller then 106m (in diameter) size particles was freshly used for the binding experiments.

Figure 5.1: Calcite mineral from Durango, Mexico

The size of the powder was analyzed under SEM (Hitachi FE-SEM S4800). Samples were placed on a SEM stub with carbon tape. The images were taken under 10kv and 20uA and a working distance of 8 (5.2).



(a) (b)

Figure 5.2: SEM images of calcite powder filtered through 106m sieve.

5.1.3 Peptide Binding

Calcite powder was washed twice with water before the experiment. For each binding assay 45 mg powder was incubated with 0.1 mg peptide in 500ul diH2O. Incubation was performed on a rotary shaker for an hour, in room temperature. Following the incubation period, powder was settled with two minuets, centrifugation at 3000 rpm and unbound peptide was recovered. Unbound peptide solution was measured with NanoDrop (in UV-Vis mode). The absorbance at 227 nm was recorded and bound peptide concentration was calculated. For each peptide 0.1 mg as the starting concentration was used. A standard curve for each of them (0.1, 0.2, 0.4, 0.8, 1 mg/ml standards) was prepared and used to calculate the unbound amount. This number was subtracted from initial amount of the peptide to find the concentration of bonded peptide. Each experiment repeated at least twice with duplicate measurements.

5.2 Results

Figure 5.3 shows the percentage of binding to calcite for each peptide. In this study, linear regression between the experimental percentage of binding and computationally calculated minimum binding energy was performed and R-squared was calculated (Fig. 5.4) Three positive and four negative binders were selected to experimentally validate computational modeling predictions. Obviously, if we could do the experiment on more peptides we could have got more precise results and

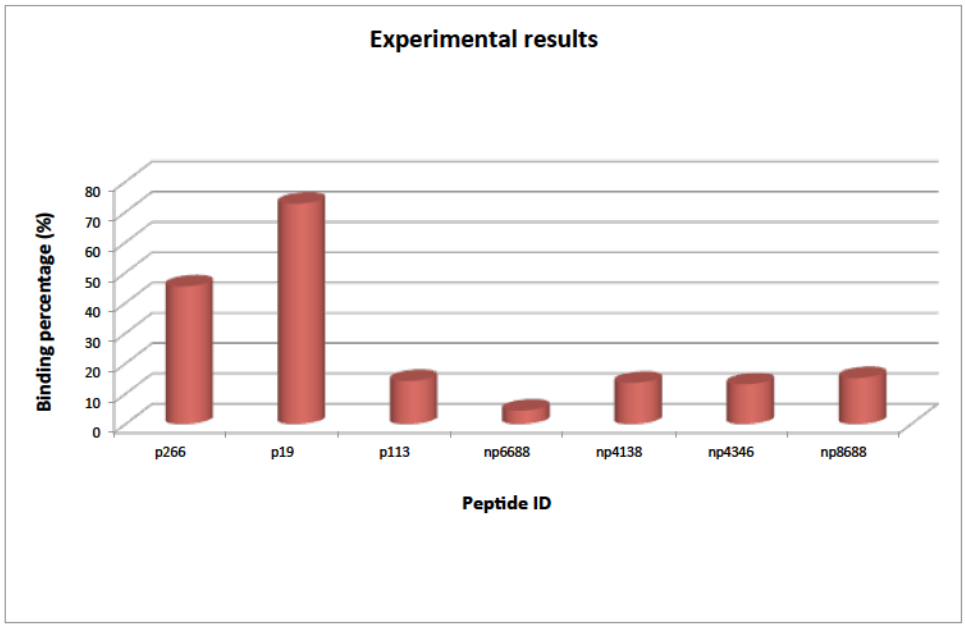


Figure 5.3: Percentage of bound peptide towards calcite

therefore, a better explanation for how far our computational model could predict the behavior of peptides in the experimental study.

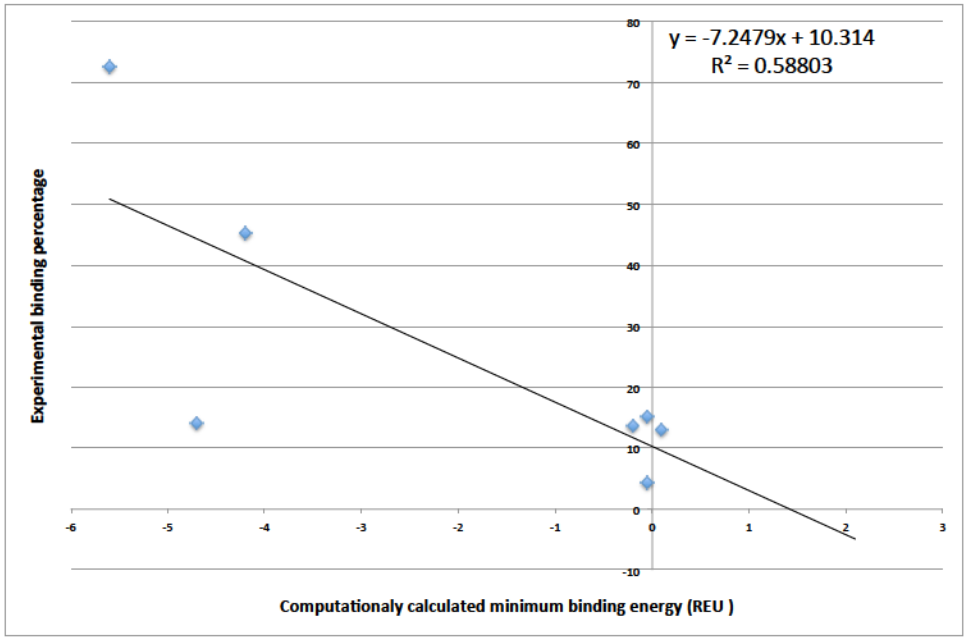


Figure 5.4: Linear regression between experimental bonding percentage and computational binding energy

In accordance to Pearson's correlation coefficient (R), our result which is -0.767 shows that there is a strong negative correlation between the peptide-calcite 104 surface binding energy, calculated by our computational model, and peptide-calcite 104 surface binding percentage, which was obtained in experimental study. By assuming the p value to be 0.05 (which is a common level for educational research) and considering the degree of freedom to be 5 (as we have 7 pair of data) the critical value for r, based on the correlation coefficient table, is -0.754. Therefore, based on our result for r, we can be 95% confident to reject the null hypothesis and announce a statistically significant relationship between our variables.

The determination coefficient (R-squared) in our study, which is 0.588, shows that our first variable, peptide-calcite 104 surface binding energy, can impart 58.8% share in peptide-calcite 104 surface binding percentage occurrence. In other words, our computational model can predict up to 58.8% of the peptide-calcite 104 surface binding percentage for each experiments. Furthermore, the adjusted R-squared for our results, which is obtained by removing biased data, is 0.506, and still is a big number for prediction the experimental results. Although the Pearson's correlation coefficient does not resemble the cause relation between two variables, in our study it shows a good linear relationship.

Regression Statistics	
R	- 0.766834399
R Square	0.588034995
Adjusted R Square	0.505641994
Standard Error	17.21768559
Observations	7

Figure 5.5: regression statics for estimating the relationship between experimental and computational study

By removing the outlier (data for p113) and performing another linear regression R-squared went up to 0.95 (see Fig. 5.6). The outlier can be predicted before the experiments and by just considering the computational study. This will be explained in more detail in next chapter.

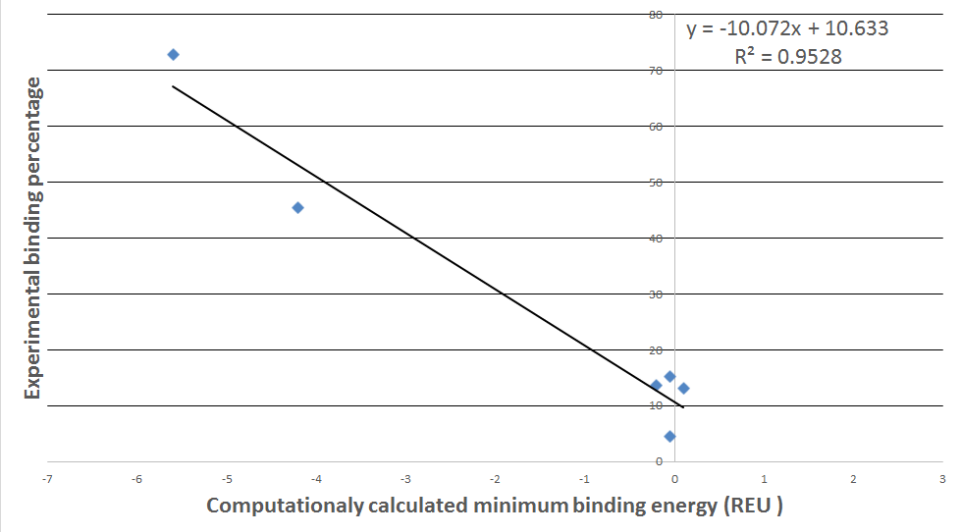


Figure 5.6: Linear regression between experimental bonding percentage and computational binding energy without outlier

Chapter 6

Conclusions and Future Work

6.1 Conclusion

Peptide selection for binding to an inorganic material is an important tool in biotechnology, material science, medicine and so on. Up until now, there has not been much progress in improving this technique and during the past 10 years all the studies and associated works did not have a sufficiently substantial contribution to make a major breakthrough in this approach [74]. Most of the experimental studies in this field have faced many challenges such as having biased libraries and the absence of dominant sequences as positive controls, which mainly made the wet lab studies in this area infeasible and without significant progress [74].

This study has introduced a new method for the selection of peptides for inorganic material binding by *in silico* modeling. We have created a library of 4-mer peptides with different chemical properties. We have been able to screen the library against an arbitary target material to identify and rank the strongest potential binders. In this study we have utilized findings of previous studies in this field and built on them.

First of all each peptide-inorganic binding pair is unique and has different properties and binding characteristics [60, 70, 71, 72, 73, 88, 171]. Based on this assumption we could make a pool of different peptides. The amino acids were classified based on their chemical properties into seven groups (polar, non-polar, acidic, basic, aromatic, nucleophilic and special-cases). 4-mer peptides could be built in each group. They were selected to reduce the computational cost and represent a

starting point for peptide selection for a specific surface. The sequences of amino acids in a given peptide are instrumental in determining the binding affinity between peptide and the material surface [68]. Therefore, all the possible permutations of amino acids in each group were considered.

Hnilova et al. found out that peptide molecular conformation has a huge effect on binding affinity [58]. In this study, possible conformations of peptides were obtained by MD simulation. The fully unrestrained system was simulated for 50 ns at 300 K, generating 50,000 conformations of each peptide.

Clustering grouped the similar structures and reduced the number of conformations for each peptide. Conformations at an interval of 20 ps of the entire MD simulation were clustered based on the RMSD of the peptide backbone from 1 to 100 clusters. The highest populated clusters, which gave more than 70% of the total conformations in a trajectory, were chosen as peptide representative structures.

On the other hand, molecular recognition of a peptide by an inorganic material is mostly based on molecular complementarity between a peptide (tertiary structure) and material crystallography [39, 172, 173]. Shape complementarity between peptides with different conformations and the calcite 104 surface was evaluated using the Wigner-D function algorithm. For polar, aromatic, acidic, basic, nucleophilic, and special-cases groups, 10% of structures which have the most similarity to the calcite 104 surface were selected for further analysis. This number for the non-polar group was reduced to 1% as the number of peptides and their conformations in this group were much greater than those in the other groups. In the non-polar group, the top 1% were selected to obtain a manageable number of conformations. For each peptide, conformations with the most similarity to the surface were selected for the next step. Polar and aromatic groups had the highest affinity among their members for the calcite 104 surface. However, the aromatic group similarity was not as trustworthy as the other groups, because its members had an aromatic ring and it could be in good agreement with the flat crystalline surface of calcite, while there could be no strong interaction between the two.

The next step in our protocol was using RosettaSurface to dock the selected conformation of each peptide to the calcite 104 surface. This simulation is based

on the optimization of the protein rigid-body orientation, side-chain and backbone conformations on a solid surface. This results in the conformations of a peptide adsorbed onto a surface as well as the energy for the whole system (peptide and calcite 104 surface) in solution and for adsorbed states. Post processing calculation on these simulations showed the perpendicular distance of each amino acid in the peptide from the given surface. Based on this information, a strongly interacting amino acid with the surface was identified.

The most important thermodynamic feature in organic-inorganic interactions is the change in the free energy [30]. Therefore, comparing the binding energy of each peptide was the last parameter for identifying the strong binders.

As peptides were very small (4-mer) the carboxyl-terminus and amino-terminus could feel each other's charges and form a strong ionic bond; this resulted in the peptide forming a loop configuration. This arrangement allowed only one amino acid at a time to interact with the surface; this applied to all strong binders except p87. In case of p87 (QSTN), the NH side-chain of glutamine interacted with the O of the backbone of serine and the O of the side-chain of serine interacted with the H of the backbone of asparagine. These intramolecular interactions might have been the reason to prevent the ionic bond formation between N- and C-terminus.

P266 (NTSN), p113 (SSYN) and p19 (TTNN) showed the best binding energy (-4.2 REU). In all of them, asparagine (N) had a strong interaction with the surface. The interactions were through hydrogen bonds between NH_2 in the asparagine's side-chain and CO_3 in the surface. These three peptides were used for experimental validation. This specific interaction for p19 is shown in Figure 6.1.

In p266 (NTSN), a side-chain carbonyl of N interacted with NH of the backbone of residues at position three and four; this was likely positioning of the N in right position and stabilizing the interaction of NH_2 with the surface and hence this structure had the best binding energy.

Experimental results showed that the p266 and p19 are strong binders for calcite but p113 did not show strong propensity to bind to the surface. On the other hand, our computational study showed that the predicted strong binders to interact with that surface with the same residue and in the same way. The reason for this false

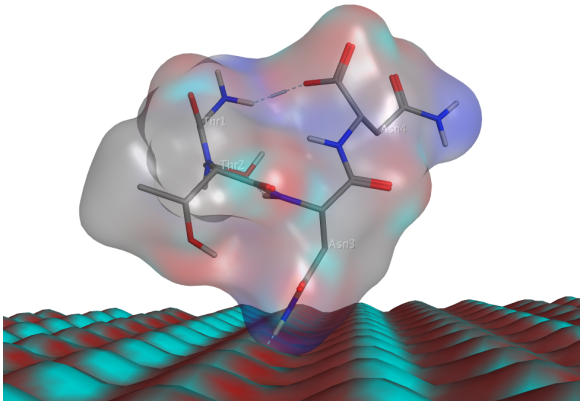


Figure 6.1: p19 interaction with calcite 104 surface.

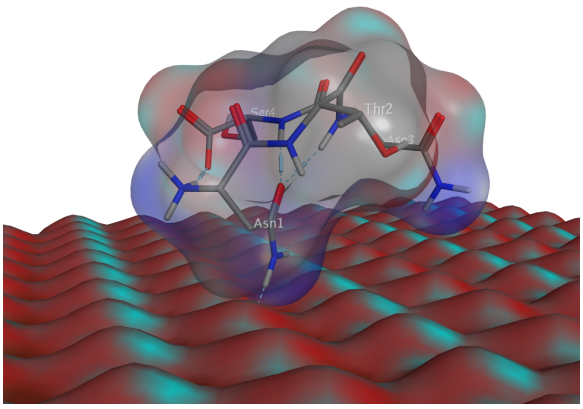


Figure 6.2: p266 interactions.

positive result in the computational study can be explained by intramolecular interactions; the p113 peptide contained various intramolecular interactions and having low binding energy could be partly because of these bindings within the peptide. Whereas, in p19 only the N- and the C-terminus interacted with each other and the negative energy could be mostly because of the peptide and the given surface interaction.

Next strong binders were p509 (TQNY), p87 (QSTN), and p37 (QSQN) with the binding energy of -3.6, -3.3, and -3.3 REU, respectively. The interacting amino acid in these peptides was glutamine (Q). The interaction happened through the side-chain (NH_2) of glutamine with CO_3 of the calcite 104 surface.

In the top three strong binders N was the interacting amino acid and when it did not interact, there was an increase in the binding energy. N and Q are almost similar and their interactions with the surface were similar. The difference between them is because of the additional CH_2 in the glutamine (Q) side-chain which makes the side-chain amide more electron-rich. This likely reflects on the hydrogen involved in H-bonding with the surface and makes it less electropositive and weakens the H-bond.

The other sequence that showed a strong binding energy was nu67 (STTC) with -3.5 REU. In this case the binding to the surface happened through a hydrogen bond between the OH in the serine side-chain and the CO_3 on the calcite 104 surface.

In all strong binders the largest contribution to the binding energy belonged to electrostatic interactions and the Lennard-Jones attraction potential. We concluded that in peptide-inorganic surface interactions the non-bonded interaction parameters (e.g., electrostatic and van der Waals) played a major role in peptide adsorption on the inorganic surface, whereas, the peptide conformation was influenced by bonded parameters.

At first glance, charged amino acids can come close to a surface with an opposite charge by simple electrostatic interactions. Surfaces and amino acids charges are pH dependent; the pH condition in solution will define the sign of the electrostatic interaction between a peptide and a surface. Thus, considering the simple electrostatic interaction as the only factor in an organic-inorganic interaction, the

prediction of amino acids adsorption on a given surface should be possible.

Some studies showed this electrostatic selectivity of amino acids on different surfaces such as quartz and calcite [174]. On the other hand, an experimental study showed the adsorption of negatively charged species on negatively charged surfaces (pyrite and TiO_2) [175, 176]. It can be concluded that the simple electrostatic model is not sufficient to explain the selectivity behavior of a peptide over a specific surface. It can be considered as a non-specific interaction [177].

Hydrogen bonding is one of the specific adsorption mechanisms in organic-inorganic interactions. Although H-bonds are weak, they bring strength and specificity to two molecules interaction by forming adducts [177].

Spectroscopic and molecular modeling of peptides adsorption on silica in the absence of water showed the hydrogen bonding between glycine carboxylic acid group and silanol from the silica surface [178]. In another study, H-bonding patterns have been seen in presence of the water molecules for glycine adsorption over silica [179, 180]. One can say this interaction is specific and well-defined in glycine adsorption on silica [177].

Another example can be the presence of electrostatic and H-bonding in lysine adsorption on quartz [181, 182] .

In this study, specific hydrogen bonding was seen between all strong binders and calcite 104 surface. Moreover, the calculations showed that one of the most important components of energy belonged to electrostatic interactions. Hence, the hydrogen bonding between the different strong peptides and the surface can be assumed as a recognition mechanism of peptides over calcite 104 surface.

6.2 Future Work

During this study we have experienced limitations, which adversely affected our achievements. We could have gained more accurate results and had a better understanding of organic-inorganic interactions if more progress was made on the experimental front concurrently with our computational work. Thus, I believe there are some indeterminate areas which with some alterations or using other tools we

could shed some light in future work.

First of all, considering explicit solvation for peptide docking to inorganic material docking using RossetaSurface could be another approach, which could help us better understand the nature of peptide-inorganic H-bond interactions, which in turn could help us improve the peptide selection.

Another important future work could be improving the force field for other inorganic materials. We can select the peptide computationally and have feedback from experiments and make a validation loop to obtain a more accurate computational model.

In this study, MD simulations of peptides have been performed in an implicit solvation because of the limited computational resources. Blue Gene is a massively parallel supercomputer that can make these simulations much faster and simplify the explicit solvation simulations. Furthermore, we can run longer simulations which give us more precise results. In this case, more accurate conformations can be obtained in a reasonable time.

Extending the 4-mer selected peptides to 12-mer by combining selected strong peptides could be an immediate future work following this project.

Bibliography

[1] J. A. Brierley, "A perspective on developments in biohydrometallurgy," *Hydrometallurgy*, vol. 94, no. 1, pp. 2–7, 2008.

[2] C. Brierley, "How will biomining be applied in future?," *Transactions of Nonferrous Metals Society of China*, vol. 18, no. 6, pp. 1302–1310, 2008.

[3] D. E. Rawlings, D. Dew, and C. du Plessis, "Biomineralization of metal-containing ores and concentrates," *TRENDS in Biotechnology*, vol. 21, no. 1, pp. 38–44, 2003.

[4] M. Boon, "The mechanism of directand indirect bacterial oxidation of sulphide minerals," *Hydrometallurgy*, vol. 62, no. 1, pp. 67–70, 2001.

[5] D. E. Rawlings, "Heavy metal mining using microbes 1," *Annual Reviews in Microbiology*, vol. 56, no. 1, pp. 65–91, 2002.

[6] D. E. Rawlings, "Biomining (mineral bioleaching, mineral biooxidation)," in *Encyclopedia of Geobiology*, pp. 182–185, Springer, 2011.

[7] D. B. Johnson, "Biomining-biotechnologies for extracting and recovering metals from ores and waste materials," *Current Opinion in Biotechnology*, vol. 30, pp. 24–31, 2014.

[8] D. E. Rawlings and D. B. Johnson, "The microbiology of biomining: development and optimization of mineral-oxidizing microbial consortia," *Microbiology*, vol. 153, no. 2, pp. 315–324, 2007.

[9] T. Rohwerder, T. Gehrke, K. Kinzler, and W. Sand, "Bioleaching review part A," *Applied Microbiology and Biotechnology*, vol. 63, no. 3, pp. 239–248, 2003.

[10] G. Olson, J. Brierley, and C. Brierley, "Bioleaching review part B," *Applied Microbiology and Biotechnology*, vol. 63, no. 3, pp. 249–257, 2003.

[11] D. B. Johnson, "Development and application of biotechnologies in the metal mining industry," *Environmental Science and Pollution Research*, vol. 20, no. 11, pp. 7768–7776, 2013.

[12] D. B. Johnson, T. Kanao, and S. Hedrich, "Redox transformations of iron at extremely low pH: Fundamental and applied aspects," *Frontiers in Microbiology*, vol. 3, 2012.

[13] C. L. Brierley and J. A. Brierley, "Progress in bioleaching: part B: Applications of microbial processes by the minerals industries," *Applied Microbiology and Biotechnology*, vol. 97, no. 17, pp. 7543–7552, 2013.

[14] K. L. Temple and A. R. Colmer, "The autotrophic oxidation of iron by a new bacterium: Thiobacillus ferrooxidans," *Journal of Bacteriology*, vol. 62, no. 5, p. 605, 1951.

[15] L. C. Bryner and A. Jameson, "Microorganisms in leaching sulfide minerals," *Applied Microbiology*, vol. 6, no. 4, p. 281, 1958.

[16] S. R. Zimmerley, D. G. Wilson, and J. D. Prater, "Cyclic leaching process employing iron oxidizing bacteria," Apr. 8 1958. US Patent 2,829,964.

[17] M. Edraki, T. Baumgartl, E. Manlapig, D. Bradshaw, D. M. Franks, and C. J. Moran, "Designing mine tailings for better environmental, social and economic outcomes: A review of alternative approaches," *Journal of Cleaner Production*, vol. 84, pp. 411–420, 2014.

[18] C. Wang, D. Harbottle, Q. Liu, and Z. Xu, "Current state of fine mineral tailings treatment: A critical review on theory and practice," *Minerals Engineering*, vol. 58, pp. 113–131, 2014.

[19] B. G. Lottermoser, "Recycling, reuse and rehabilitation of mine wastes," *Elements*, vol. 7, no. 6, pp. 405–410, 2011.

[20] A. Watson, P. Corser, E. Garces Pardo, T. Lopez Christian, and J. Vandekeybus, "A comparison of alternative tailings disposal methods–the promises and realities," *Proceedings from Mine Waste, Perth, Australia*.

[21] X. Lu and H. Wang, "Microbial oxidation of sulfide tailings and the environmental consequences," *Elements*, vol. 8, no. 2, pp. 119–124, 2012.

[22] H. E. Jamieson, "Geochemistry and mineralogy of solid mine waste: Essential knowledge for predicting environmental impact," *Elements*, vol. 7, no. 6, pp. 381–386, 2011.

[23] M. Leblanc, J. Morales, J. Borrego, and F. Elbaz-Poulichet, "4,500-year-old mining pollution in southwestern Spain: Long-term implications for modern mining pollution," *Economic Geology*, vol. 95, no. 3, pp. 655–662, 2000.

[24] X. Tan, W. G. Wu, Z. M. Wang, and H. LIU, "Experimental study on comprehensive recovery of valuable minerals from flotation tailings in copper-molybdenum mine," *Advanced Materials Research*, vol. 878, pp. 234–243, 2014.

[25] D. B. Johnson and C. A. du Plessis, "Biomining in reverse gear: Using bacteria to extract metals from oxidised ores," *Minerals Engineering*, vol. 75, pp. 2–5, 2015.

[26] M. Vera, A. Schippers, and W. Sand, "Progress in bioleaching: Fundamentals and mechanisms of bacterial metal sulfide oxidationpart A," *Applied Microbiology and Biotechnology*, vol. 97, no. 17, pp. 7529–7541, 2013.

[27] D. B. Johnson and K. B. Hallberg, "Pitfalls of passive mine water treatment," *Reviews in Environmental Science and Biotechnology*, vol. 1, no. 4, pp. 335–343, 2002.

[28] D. B. Johnson, B. M. Grail, and K. B. Hallberg, "A new direction for biomining: Extraction of metals by reductive dissolution of oxidized ores," *Minerals*, vol. 3, no. 1, pp. 49–58, 2013.

[29] P. Calvert, "Biomimetic ceramics and hard composites," *Biomimetics*, p. 14, 1995.

[30] T. R. Walsh, "Fundamentals of peptide-materials interfaces," in *Bio-Inspired Nanotechnology*, pp. 17–36, Springer, 2014.

[31] M. A. Meyers, P.-Y. Chen, A. Y.-M. Lin, and Y. Seki, "Biological materials: Structure and mechanical properties," *Progress in Materials Science*, vol. 53, no. 1, pp. 1–206, 2008.

[32] M. B. Dickerson, K. H. Sandhage, and R. R. Naik, "Protein-and peptide-directed syntheses of inorganic materials," *Chemical Reviews*, vol. 108, no. 11, pp. 4935–4978, 2008.

[33] C.-L. Chen and N. L. Rosi, "Peptide-based methods for the preparation of nanostructured inorganic materials," *Angewandte Chemie International Edition*, vol. 49, no. 11, pp. 1924–1942, 2010.

[34] L. Addadi and S. Weiner, "Control and design principles in biological mineralization," *Angewandte Chemie International Edition in English*, vol. 31, no. 2, pp. 153–169, 1992.

[35] F. Nudelman, K. Pieterse, A. George, P. H. Bomans, H. Friedrich, L. J. Brylka, P. A. Hilbers, G. de With, N. A. Sommerdijk, *et al.*, "The role of collagen in bone apatite formation in the presence of hydroxyapatite nucleation inhibitors," *Nature Materials*, vol. 9, no. 12, pp. 1004–1009, 2010.

[36] P. Das and M. Reches, "Insights into the interactions of amino acids and peptides with inorganic materials using single molecule force spectroscopy," *Peptide Science*, 2015.

[37] B. D. Ratner, A. S. Hoffman, F. J. Schoen, and J. E. Lemons, *Biomaterials science: An introduction to materials in medicine*. USA: Elsevier Academic Press, 2004.

[38] T. Horbett and J. Brash, "Proteins at interfaces II: Fundamentals and applications american chemical society," 1995.

[39] C. Tamerler and M. Sarikaya, "Genetically designed peptide-based molecular materials," *ACS Nano*, vol. 3, no. 7, pp. 1606–1615, 2009.

[40] Y. Wei and R. A. Latour, "Determination of the adsorption free energy for peptide- surface interactions by SPR spectroscopy," *Langmuir*, vol. 24, no. 13, pp. 6721–6729, 2008.

[41] M. J. Pilling, P. W. Seakins, *et al.*, *Reaction kinetics*. Oxford: Oxford University Press, 2001.

[42] P. Atkins and J. de Paula, *Physical chemistry volume 1: Thermodynamics and kinetics*. New York: Macmillan Higher Education, 2009.

[43] V. Humblot, J. Landoulsi, and C.-M. Pradier, "Mechanisms of adsorption of short peptides on metal and oxide surfaces," *Peptide Materials: From Nanostructures to Applications*, pp. 289–311, 2013.

[44] I. Lynch and K. A. Dawson, "Protein-nanoparticle interactions," *Nano Today*, vol. 3, no. 1, pp. 40–47, 2008.

[45] K. Chen, Y. Xu, S. Rana, O. R. Miranda, P. L. Dubin, V. M. Rotello, L. Sun, and X. Guo, "Electrostatic selectivity in protein–nanoparticle interactions," *Biomacromolecules*, vol. 12, no. 7, pp. 2552–2561, 2011.

[46] K. Rezwan, A. Studart, J. Vörös, and L. Gauckler, "Change of ζ potential of biocompatible colloidal oxide particles upon adsorption of bovine serum albumin and lysozyme," *The Journal of Physical Chemistry B*, vol. 109, no. 30, pp. 14469–14474, 2005.

[47] M. J. Limo, C. C. Perry, A. Thyparambil, Y. Wei, and R. A. Latour, "Experimental characterization of peptide–surface interactions," in *Bio-Inspired Nanotechnology*, pp. 37–94, Springer, 2014.

[48] J. O. Bockris and A. K. Reddy, *Modern electrochemistry: A Molecular Level Approach*. New York: Plenum Press, 1970.

[49] J. Léonil, D. Mollé, S. Bouhallab, and G. Henry, "Precipitation of hydrophobic peptides from tryptic casein hydrolysate by salt and pH," *Enzyme and Microbial Technology*, vol. 16, no. 7, pp. 591–595, 1994.

[50] P. Sengupta, K. Garai, B. Sahoo, Y. Shi, D. J. Callaway, and S. Maiti, "The amyloid β peptide (Aβ1-40) is thermodynamically soluble at physiological concentrations," *Biochemistry*, vol. 42, no. 35, pp. 10506–10513, 2003.

[51] C.-M. Pradier, V. Humblot, L. Stievano, C. Méthivier, and J.-F. Lambert, "Salt concentration and pH-dependent adsorption of two polypeptides on planar and divided alumina surfaces. In situ IR investigations," *Langmuir*, vol. 23, no. 5, pp. 2463–2471, 2007.

[52] S. V. Patwardhan, F. S. Emami, R. J. Berry, S. E. Jones, R. R. Naik, O. Deschaume, H. Heinz, and C. C. Perry, "Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption," *Journal of the American Chemical Society*, vol. 134, no. 14, pp. 6244–6256, 2012.

[53] A. A. Skelton, T. Liang, and T. R. Walsh, "Interplay of sequence, conformation, and binding at the peptide- titania interface as mediated by water," *ACS Applied Materials & Interfaces*, vol. 1, no. 7, pp. 1482–1491, 2009.

[54] J. Schneider and L. Colombi Ciacchi, "Specific material recognition by small peptides mediated by the interfacial solvent structure," *Journal of the American Chemical Society*, vol. 134, no. 4, pp. 2407–2413, 2012.

[55] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *Journal of Molecular Biology*, vol. 7, no. 1, pp. 95–99, 1963.

[56] H. M. Albert, L. Nicholas, and V. P. Rohit, "Describing sequence-ensemble relationships for intrinsically disordered proteins," *Biochemical Journal*, vol. 449, no. 2, pp. 307–318, 2013.

[57] S. Collino and J. S. Evans, "Molecular specifications of a mineral modulation sequence derived from the aragonite-promoting protein n16," *Biomacromolecules*, vol. 9, no. 7, pp. 1909–1918, 2008.

[58] M. Hnilova, E. E. Oren, U. O. Seker, B. R. Wilson, S. Collino, J. S. Evans, C. Tamerler, and M. Sarikaya, "Effect of molecular conformations on the adsorption behavior of gold-binding peptides," *Langmuir*, vol. 24, no. 21, pp. 12440–12445, 2008.

[59] J. M. Slocik, A. O. Govorov, and R. R. Naik, "Plasmonic circular dichroism of peptide-functionalized gold nanoparticles," *Nano letters*, vol. 11, no. 2, pp. 701–705, 2011.

[60] E. E. Oren, R. Notman, I. W. Kim, J. S. Evans, T. R. Walsh, R. Samudrala, C. Tamerler, and M. Sarikaya, "Probing the molecular mechanisms of quartz-binding peptides," *Langmuir*, vol. 26, no. 13, pp. 11003–11009, 2010.

[61] M. R. Jensen, L. Salmon, G. Nodet, and M. Blackledge, "Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts," *Journal of the American Chemical Society*, vol. 132, no. 4, pp. 1270–1272, 2010.

[62] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, "Function and structure of inherently disordered proteins," *Current Opinion in Structural Biology*, vol. 18, no. 6, pp. 756–764, 2008.

[63] P. Tompa, "Intrinsically disordered proteins: A 10-year recap," *Trends in Biochemical Sciences*, vol. 37, no. 12, pp. 509–516, 2012.

[64] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky, and Z. Obradovic, "Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1882–1898, 2007.

[65] E. C. Keene, J. S. Evans, and L. A. Estroff, "Matrix interactions in biomineralization: aragonite nucleation by an intrinsically disordered nacre polypeptide, n16n, associated with a β-chitin substrate," *Crystal Growth & Design*, vol. 10, no. 3, pp. 1383–1389, 2010.

[66] C. R. So, J. L. Kulp III, E. E. Oren, H. Zareie, C. Tamerler, J. S. Evans, and M. Sarikaya, "Molecular recognition and supramolecular self-assembly of a genetically engineered gold binding peptide on Au $\{111\}$," *Acs Nano*, vol. 3, no. 6, pp. 1525–1531, 2009.

[67] S. Z. Nergiz, J. M. Slocik, R. R. Naik, and S. Singamaneni, "Surface defect sites facilitate fibrillation: An insight into adsorption of gold-binding peptides on Au (111)," *Phys. Chem. Chem. Phys.*, vol. 15, no. 28, pp. 11629–11633, 2013.

[68] E. E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U. O. S. Seker, M. Sarikaya, and R. Samudrala, "A novel knowledge-based approach to design inorganic-binding peptides," *Bioinformatics*, vol. 23, no. 21, pp. 2816–2822, 2007.

[69] K. Goede, P. Busch, and M. Grundmann, "Binding specificity of a peptide on semiconductor surfaces," *Nano Letters*, vol. 4, no. 11, pp. 2115–2120, 2004.

[70] S. R. Whaley, D. English, E. L. Hu, P. F. Barbara, and A. M. Belcher, "Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly," *Nature*, vol. 405, no. 6787, pp. 665–668, 2000.

[71] M. Sarikaya, C. Tamerler, A. K.-Y. Jen, K. Schulten, and F. Baneyx, "Molecular biomimetics: nanotechnology through biology," *Nature Materials*, vol. 2, no. 9, pp. 577–585, 2003.

[72] R. R. Naik, L. L. Brott, S. J. Clarson, and M. O. Stone, "Silica-precipitating peptides isolated from a combinatorial phage display peptide library," *Journal of Nanoscience and Nanotechnology*, vol. 2, no. 1, pp. 95–100, 2002.

[73] M. Sarikaya, C. Tamerler, D. T. Schwartz, and F. Baneyx, "Materials assembly and formation using engineered polypeptides," *Annu. Rev. Mater. Res.*, vol. 34, pp. 373–408, 2004.

[74] J. M. Slocik and R. R. Naik, "Peptide-nanoparticle strategies, interactions, and challenges," in *Bio-Inspired Nanotechnology*, pp. 1–16, Springer, 2014.

[75] G. P. Smith, "Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.

[76] G. P. Smith and V. A. Petrenko, "Phage display," *Chemical Reviews*, vol. 97, no. 2, pp. 391–410, 1997.

[77] H.-E. Jin, W.-J. Chung, and S.-W. Lee, "Phage display for the discovery of hydroxyapatite-associated peptides.," *Methods in Enzymology*, vol. 532, pp. 305–323, 2012.

[78] S. Y. Lee, J. H. Choi, and Z. Xu, "Microbial cell-surface display," *Trends in Biotechnology*, vol. 21, no. 1, pp. 45–52, 2003.

[79] S. S. Sidhu, "Engineering M13 for phage display," *Biomolecular Engineering*, vol. 18, no. 2, pp. 57–63, 2001.

[80] V. Petrenko, G. Smith, X. Gong, and T. Quinn, "A library of organic landscapes on filamentous phage," *Protein Engineering*, vol. 9, no. 9, pp. 797–801, 1996.

[81] S. Brown, "Metal-recognition by repeating polypeptides," *Nature biotechnology*, vol. 15, no. 3, pp. 269–272, 1997.

[82] U. O. S. Seker, B. Wilson, S. Dincer, I. W. Kim, E. E. Oren, J. S. Evans, C. Tamerler, and M. Sarikaya, "Adsorption behavior of linear and cyclic genetically engineered platinum binding peptides," *Langmuir*, vol. 23, no. 15, pp. 7895–7900, 2007.

[83] K. Tao, J. Wang, Y. Li, D. Xia, H. Shan, H. Xu, and J. R. Lu, "Short peptide-directed synthesis of one-dimensional platinum nanostructures with controllable morphologies," *Scientific Reports*, vol. 3, 2013.

[84] L. L. Brott, R. R. Naik, D. J. Pikas, S. M. Kirkpatrick, D. W. Tomlin, P. W. Whitlock, S. J. Clarson, and M. O. Stone, "Ultrafast holographic nanopatterning of biocatalytically formed silica," *Nature*, vol. 413, no. 6853, pp. 291–293, 2001.

[85] C. K. Thai, H. Dai, M. Sastry, M. Sarikaya, D. T. Schwartz, and F. Baneyx, "Identification and characterization of Cu2O- and ZnO-binding polypeptides by escherichia coli cell surface display: Toward an understanding of metal oxide binding," *Biotechnology and Bioengineering*, vol. 87, no. 2, pp. 129–137, 2004.

[86] W. Zhou, D. T. Schwartz, and F. Baneyx, "Single-pot biofabrication of zinc sulfide immuno-quantum dots," *Journal of the American Chemical Society*, vol. 132, no. 13, pp. 4731–4738, 2010.

[87] D. Rothenstein, B. Claasen, B. Omiecienski, P. Lammel, and J. Bill, "Isolation of ZnO-binding 12-mer peptides and determination of their binding epitopes by NMR spectroscopy," *Journal of the American Chemical Society*, vol. 134, no. 30, pp. 12547–12556, 2012.

[88] K.-I. Sano, H. Sasaki, and K. Shiba, "Specificity and biomineralization activities of Ti-binding peptide-1 (TBP-1)," *Langmuir*, vol. 21, no. 7, pp. 3090–3095, 2005.

[89] V. Puddu, J. M. Slocik, R. R. Naik, and C. C. Perry, "Titania binding peptides as templates in the biomimetic synthesis of stable titania nanosols: Insight into the role of buffers in peptide-mediated mineralization," *Langmuir*, vol. 29, no. 30, pp. 9464–9472, 2013.

[90] H. Yazici, H. Fong, B. Wilson, E. Oren, F. Amos, H. Zhang, J. Evans, M. Snead, M. Sarikaya, and C. Tamerler, "Biological response on a titanium implant-grade surface functionalized with modular peptides," *Acta Biomaterialia*, vol. 9, no. 2, pp. 5341–5352, 2013.

[91] K. Hashimoto, M. Yoshinari, K. Matsuzaka, K. Shiba, and T. Inoue, "Identification of peptide motif that binds to the surface of zirconia," *Dental Materials Journal*, vol. 30, no. 6, pp. 935–940, 2011.

[92] D. J. Gaskin, K. Starck, and E. N. Vulfson, "Identification of inorganic crystal-specific sequences using phage display combinatorial library of short peptides: A feasibility study," *Biotechnology Letters*, vol. 22, no. 15, pp. 1211–1216, 2000.

[93] E. M. Krauland, B. R. Peelle, K. D. Wittrup, and A. M. Belcher, "Peptide tags for enhanced cellular and protein adhesion to single-crystalline sapphire," *Biotechnology and Bioengineering*, vol. 97, no. 5, pp. 1009–1020, 2007.

[94] M. Gungormus, H. Fong, I. W. Kim, J. S. Evans, C. Tamerler, and M. Sarikaya, "Regulation of in vitro calcium phosphate mineralization by combinatorially selected hydroxyapatite-binding peptides," *Biomacromolecules*, vol. 9, no. 3, pp. 966–973, 2008.

[95] S.-W. Lee, C. Mao, C. E. Flynn, and A. M. Belcher, "Ordering of quantum dots using genetically engineered viruses," *Science*, vol. 296, no. 5569, pp. 892–895, 2002.

[96] M. Matmor and N. Ashkenasy, "Modulating semiconductor surface electronic properties by inorganic peptide–binders sequence design," *Journal of the American Chemical Society*, vol. 134, no. 50, pp. 20403–20411, 2012.

[97] C. R. So, Y. Hayamizu, H. Yazici, C. Gresswell, D. Khatayevich, C. Tamerler, and M. Sarikaya, "Controlling self-assembly of engineered peptides on graphite by rational mutation," *Acs Nano*, vol. 6, no. 2, pp. 1648–1656, 2012.

[98] B. Akdim, R. Pachter, S. S. Kim, R. R. Naik, T. R. Walsh, S. Trohalaki, G. Hong, Z. Kuang, and B. L. Farmer, "Electronic properties of a graphene device with peptide adsorption: insight from simulation," *ACS Applied Materials & Interfaces*, vol. 5, no. 15, pp. 7470–7477, 2013.

[99] B. Gabryelczyk, G. R. Szilvay, M. Salomäki, P. Laaksonen, and M. B. Linder, "Selection and characterization of peptides binding to diamond-like carbon," *Colloids and Surfaces B: Biointerfaces*, vol. 110, pp. 66–73, 2013.

[100] R. Derda, S. K. Tang, S. C. Li, S. Ng, W. Matochko, and M. R. Jafari, "Diversity of phage-displayed libraries of peptides during panning and amplification," *Molecules*, vol. 16, no. 2, pp. 1776–1803, 2011.

[101] W. L. Matochko, K. Chu, B. Jin, S. W. Lee, G. M. Whitesides, and R. Derda, "Deep sequencing analysis of phage libraries using Illumina platform," *Methods*, vol. 58, no. 1, pp. 47–55, 2012.

[102] A. Lyubartsev, Y. Tu, and A. Laaksonen, "Hierarchical multiscale modelling scheme from first principles to mesoscale," *Journal of Computational and Theoretical Nanoscience*, vol. 6, no. 5, pp. 951–959, 2009.

[103] F. Iori, R. Di Felice, E. Molinari, and S. Corni, "GolP: An atomistic force-field to describe the interaction of proteins with Au (111) surfaces in water," *Journal of Computational Chemistry*, vol. 30, no. 9, pp. 1465–1476, 2009.

[104] R. A. Latour, "Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions (Review)," *Biointerphases*, vol. 3, no. 3, pp. FC2–FC12, 2008.

[105] M. Corno, A. Rimola, V. Bolis, and P. Ugliengo, "Hydroxyapatite as a key biomaterial: Quantum-mechanical simulation of its surfaces in interaction with biomolecules," *Physical Chemistry Chemical Physics*, vol. 12, no. 24, pp. 6309–6329, 2010.

[106] R. Di Felice and A. Selloni, "Adsorption modes of cysteine on Au (111): Thiolate, amino-thiolate, disulfide," *The Journal of Chemical Physics*, vol. 120, no. 10, pp. 4906–4914, 2004.

[107] R. Di Felice, A. Selloni, and E. Molinari, "DFT study of cysteine adsorption on Au (111)," *The Journal of Physical Chemistry B*, vol. 107, no. 5, pp. 1151–1156, 2003.

[108] D. I. Sayago, M. Polcik, G. Nisbet, C. L. Lamont, and D. P. Woodruff, "Local structure determination of a chiral adsorbate: Alanine on Cu (110)," *Surface Science*, vol. 590, no. 1, pp. 76–87, 2005.

[109] M. Nyberg, J. Hasselström, O. Karis, N. Wassdahl, M. Weinelt, A. Nilsson, and L. G. Pettersson, "The electronic structure and surface chemistry of glycine adsorbed on Cu (110)," *The Journal of Chemical Physics*, vol. 112, no. 12, pp. 5420–5427, 2000.

[110] L. M. Ghiringhelli, P. Schravendijk, and L. Delle Site, "Adsorption of alanine on a Ni (111) surface: A multiscale modeling oriented density functional study," *Physical Review B*, vol. 74, no. 3, p. 035437, 2006.

[111] R. Tonner, "Adsorption of proline and glycine on the Tio2 (110) surface: A density functional theory study," *ChemPhysChem*, vol. 11, no. 5, pp. 1053–1061, 2010.

[112] C. L. Freeman, J. H. Harding, D. Quigley, and P. M. Rodger, "Simulations of ovocleidin-17 binding to calcite surfaces and its implications for eggshell formation," *The Journal of Physical Chemistry C*, vol. 115, no. 16, pp. 8175–8183, 2011.

[113] F. Wang, S. J. Stuart, and R. A. Latour, "Calculation of adsorption free energy for solute-surface interactions using biased replica-exchange molecular dynamics," *Biointerphases*, vol. 3, no. 1, pp. 9–18, 2008.

[114] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, "Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system," *Proteins: Structure, Function, and Bioinformatics*, vol. 4, no. 1, pp. 31–47, 1988.

[115] J. Tirado-rives, "The OPlS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin," *J am Chem Soc*, vol. 110, pp. 1657–1666, 1988.

[116] M. Karplus *et al.*, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J Comput Chem*, vol. 4, p. 187217, 1983.

[117] R. T. Cygan, J.-J. Liang, and A. G. Kalinichev, "Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field," *The Journal of Physical Chemistry B*, vol. 108, no. 4, pp. 1255–1266, 2004.

[118] B. J. Teppen, K. Rasmussen, P. M. Bertsch, D. M. Miller, and L. Schaefer, "Molecular dynamics modeling of clay minerals. 1. Gibbsite, kaolinite, pyrophyllite, and beidellite," *The Journal of Physical Chemistry B*, vol. 101, no. 9, pp. 1579–1587, 1997.

[119] H. Heinz, T.-J. Lin, R. Kishore Mishra, and F. S. Emami, "Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: The INTERFACE force field," *Langmuir*, vol. 29, no. 6, pp. 1754–1765, 2013.

[120] M. C. Pitman and A. C. Van Duin, "Dynamics of confined reactive water in smectite clay–zeolite composites," *Journal of the American Chemical Society*, vol. 134, no. 6, pp. 3042–3053, 2012.

[121] J. A. Greathouse, D. B. Hart, and M. E. Ochs, "Alcohol and thiol adsorption on (oxy) hydroxide and carbon surfaces: Molecular dynamics simulation and desorption experiments," *The Journal of Physical Chemistry C*, vol. 116, no. 51, pp. 26756–26764, 2012.

[122] M. Veteška, M. Pospíšil, K. Melánová, L. Beneš, and V. Zima, "Structure analysis of hydrotalcite intercalated with pyrenetetrasulfonate; experiments and molecular modelling," *Journal of Molecular Modeling*, vol. 14, no. 12, pp. 1119–1129, 2008.

[123] P. Kovář, M. Pospíšil, E. Káfuňková, K. Lang, and F. Kovanda, "Mg-Al layered double hydroxide intercalated with porphyrin anions: Molecular simulations and experiments," *Journal of Molecular Modeling*, vol. 16, no. 2, pp. 223–233, 2010.

[124] P. Praus, M. Veteška, and M. Pospíšil, "Adsorption of phenol and aniline on natural and organically modified montmorillonite: Experiment and molecular modelling," *Molecular Simulation*, vol. 37, no. 11, pp. 964–974, 2011.

[125] C. Liu, H. Li, C. T. Johnston, S. A. Boyd, and B. J. Teppen, "Relating clay structural factors to dioxin adsorption by smectites: Molecular dynamics simulations," *Soil Science Society of America Journal*, vol. 76, no. 1, pp. 110–120, 2012.

[126] P. Padma Kumar, A. G. Kalinichev, and R. J. Kirkpatrick, "Hydration, swelling, interlayer structure, and hydrogen bonding in organolayered double hydroxides: Insights from molecular dynamics simulation of citrate-intercalated hydrotalcite," *The Journal of Physical Chemistry B*, vol. 110, no. 9, pp. 3841–3844, 2006.

[127] P. P. Kumar, A. G. Kalinichev, and R. J. Kirkpatrick, "Molecular dynamics simulation of the energetics and structure of layered double hydroxides intercalated with carboxylic acids," *The Journal of Physical Chemistry C*, vol. 111, no. 36, pp. 13517–13523, 2007.

[128] A. G. Kalinichev, P. Padma Kumar, and R. James Kirkpatrick, "Molecular dynamics computer simulations of the effects of hydrogen bonding on the properties of layered double hydroxides intercalated with organic acids," *Philosophical Magazine*, vol. 90, no. 17-18, pp. 2475–2488, 2010.

[129] H. C. Greenwell, A. A. Bowden, B. Chen, P. Boulet, J. R. Evans, P. V. Coveney, and A. Whiting, "Intercalation and in situ polymerization of poly (alkylene oxide) derivatives within M+-montmorillonite (M= Li, Na, K)," *Journal of Materials Chemistry*, vol. 16, no. 11, pp. 1082–1094, 2006.

[130] R. L. Anderson, H. C. Greenwell, J. L. Suter, P. V. Coveney, and M.-A. Thyveetil, "Determining materials properties of natural composites using molecular simulation," *Journal of Materials Chemistry*, vol. 19, no. 39, pp. 7251–7262, 2009.

[131] L. Kulhánková, J. Tokarskỳ, P. Peikertová, K. M. Kutláková, L. Ivánek, and P. Čapková, "Montmorillonite intercalated by conducting polyanilines," *Journal of Physics and Chemistry of Solids*, vol. 73, no. 12, pp. 1530–1533, 2012.

[132] M.-A. Thyveetil, P. V. Coveney, H. C. Greenwell, and J. L. Suter, "Computer simulation study of the structural stability and materials properties of DNA-intercalated layered double hydroxides," *Journal of the American Chemical Society*, vol. 130, no. 14, pp. 4742–4756, 2008.

[133] M.-A. Thyveetil, P. V. Coveney, H. C. Greenwell, and J. L. Suter, "Role of host layer flexibility in DNA guest intercalation revealed by computer simulation of layered nanomaterials," *Journal of the American Chemical Society*, vol. 130, no. 37, pp. 12485–12495, 2008.

[134] J. B. Swadling, P. V. Coveney, and H. C. Greenwell, "Clay minerals mediate folding and regioselective interactions of RNA: a large-scale atomistic simulation study," *Journal of the American Chemical Society*, vol. 132, no. 39, pp. 13750–13764, 2010.

[135] J. B. Swadling, P. V. Coveney, and H. C. Greenwell, "Stability of free and mineral-protected nucleic acids: Implications for the RNA world," *Geochimica et Cosmochimica Acta*, vol. 83, pp. 360–378, 2012.

[136] J. B. Swadling, J. L. Suter, H. C. Greenwell, and P. V. Coveney, "Influence of surface chemistry and charge on mineral–RNA interactions," *Langmuir*, vol. 29, no. 5, pp. 1573–1583, 2013.

[137] S. R. Euston and M. A. Naser, "Simulating the equation of state of model globular proteins adsorbed at a surface," *Langmuir*, vol. 21, no. 9, pp. 4227–4235, 2005.

[138] M. Bachmann and W. Janke, "Conformational transitions of nongrafted polymers near an absorbing substrate," *Physical Review Letters*, vol. 95, no. 5, p. 058102, 2005.

[139] M. Lund, T. Åkesson, and B. Jönsson, "Enhanced protein adsorption due to charge regulation," *Langmuir*, vol. 21, no. 18, pp. 8385–8388, 2005.

[140] J. Zhou, H.-K. Tsao, Y.-J. Sheng, and S. Jiang, "Monte carlo simulations of antibody adsorption and orientation on charged surfaces," *The Journal of Chemical Physics*, vol. 121, no. 2, pp. 1050–1057, 2004.

[141] J. Zhou, S. Chen, and S. Jiang, "Orientation of adsorbed antibodies on charged surfaces by computer simulation based on a united-residue model," *Langmuir*, vol. 19, no. 8, pp. 3472–3478, 2003.

[142] Y. Dai and J. S. Evans, "An energy-based mapping method for identifying the in-plane orientations of polypeptides and other macromolecules at crystalline interfaces," *The Journal of Chemical Physics*, vol. 112, no. 11, pp. 5144–5157, 2000.

[143] D. Song and D. Forciniti, "Monte carlo simulations of peptide adsorption on solid surfaces (Monte Carlo simulations of peptide adsorption)," *The Journal of Chemical Physics*, vol. 115, no. 17, pp. 8089–8100, 2001.

[144] A. Mungikar and D. Forciniti, "Effect of cosolvents on the adsorption of peptides at the solid-liquid interface," *Biomacromolecules*, vol. 7, no. 1, pp. 239–251, 2006.

[145] D. L. Masica, S. B. Schrier, E. A. Specht, and J. J. Gray, "De novo design of peptide- calcite biomineralization systems," *Journal of the American Chemical Society*, vol. 132, no. 35, pp. 12252–12262, 2010.

[146] D. Case, T. Darden, T. E. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, *et al.*, "AMBER 12," *University of California, San Francisco*, vol. 1, no. 3, 2012.

[147] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 3, pp. 712–725, 2006.

[148] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *Journal of the American Chemical Society*, vol. 112, no. 16, pp. 6127–6129, 1990.

[149] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham, "Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms," *Journal of Chemical Theory and Computation*, vol. 3, no. 6, pp. 2312–2334, 2007.

[150] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 224–227, 1979.

[151] S. M. Saberi Fathi, D. T. White, and J. A. Tuszynski, "Geometrical comparison of two protein structures using Wigner-D functions," *Proteins: Structure, Function, and Bioinformatics*, vol. 82, no. 10, pp. 2756–2769, 2014.

[152] D. Potts, J. Prestin, and A. Vollrath, "A fast algorithm for nonequispaced fourier transforms on the rotation group," *Numerical Algorithms*, vol. 52, no. 3, pp. 355–384, 2009.

[153] R. Hielscher, D. Potts, J. Prestin, H. Schaeben, and M. Schmalz, "The radon transform on SO (3): A fourier slice theorem and numerical inversion," *Inverse Problems*, vol. 24, no. 2, p. 025011, 2008.

[154] H. Lipson, C. A. Taylor, and I. Fankuchen, "Fourier transforms and X-ray diffraction," *Physics Today*, vol. 12, no. 10, pp. 56–56, 2009.

[155] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[156] K. Makrodimitris, D. L. Masica, E. T. Kim, and J. J. Gray, "Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite," *Journal of the American Chemical Society*, vol. 129, no. 44, pp. 13713–13722, 2007.

[157] D. L. Masica and J. J. Gray, "Solution-and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system," *Biophysical Journal*, vol. 96, no. 8, pp. 3082–3091, 2009.

[158] M. S. Pacella, R. A. Da Chen Emily Koo, and J. J. G. Thottungal, "Using the Rosettasurface algorithm to predict protein structure at mineral surfaces," *Methods in Enzymology*, vol. 532, p. 343, 2013.

[159] J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker, "Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations," *Journal of Molecular Biology*, vol. 331, no. 1, pp. 281–299, 2003.

[160] S. Hwang, M. Blanco, and W. A. Goddard, "Atomistic simulations of corrosion inhibitors adsorbed on calcite surfaces I. Force field parameters for calcite," *The Journal of Physical Chemistry B*, vol. 105, no. 44, pp. 10746–10752, 2001.

[161] E. Stöckelmann and R. Hentschke, "Adsorption isotherms of water vapor on calcite: A molecular dynamics-Monte Carlo hybrid simulation using a polarizable water model," *Langmuir*, vol. 15, no. 15, pp. 5141–5149, 1999.

[162] S. Kerisit and S. C. Parker, "Free energy of adsorption of water and calcium on the {10 1 [combining macron] 4} calcite surface," *Chemical Communications*, no. 1, pp. 52–53, 2004.

[163] G. Collier, N. A. Vellore, J. A. Yancey, S. J. Stuart, and R. A. Latour, "Comparison between empirical protein force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces," *Biointerphases*, vol. 7, no. 1, p. 24, 2012.

[164] A. D. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. a. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The Journal of Physical Chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[165] N. A. Vellore, J. A. Yancey, G. Collier, R. A. Latour, and S. J. Stuart, "Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions," *Langmuir*, vol. 26, no. 10, pp. 7396–7404, 2010.

[166] C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker, "Protein structure prediction using rosetta," *Methods in Enzymology*, vol. 383, pp. 66–93, 2004.

[167] T. Lazaridis and M. Karplus, "Effective energy function for proteins in solution," *Proteins: Structure, Function, and Bioinformatics*, vol. 35, no. 2, pp. 133–152, 1999.

[168] T. Kortemme, A. V. Morozov, and D. Baker, "An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes," *Journal of molecular biology*, vol. 326, no. 4, pp. 1239–1259, 2003.

[169] A. Leaver-Fay, M. J. OMeara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, *et al.*, "Scientific benchmarks for guiding macromolecular energy function improvement," *Methods in Enzymology*, vol. 523, p. 109, 2013.

[170] K. P. Kilambi and J. J. Gray, "Rapid calculation of protein pK a values using rosetta," *Biophysical Journal*, vol. 103, no. 3, pp. 587–595, 2012.

[171] C. Mao, C. E. Flynn, A. Hayhurst, R. Sweeney, J. Qi, G. Georgiou, B. Iverson, and A. M. Belcher, "Viral assembly of oriented quantum dot nanowires," *Proceedings of the National Academy of Sciences*, vol. 100, no. 12, pp. 6946–6951, 2003.

[172] S. Mann, D. D. Archibald, J. M. Didymus, T. Douglas, B. R. Heywood, F. C. Meldrum, and N. J. Reeves, "Crystallization at inorganic-organic interfaces: Biominerals and biomimetic synthesis," *Science*, vol. 261, no. 5126, pp. 1286–1292, 1993.

[173] J. V. Barth, J. Weckesser, C. Cai, P. Günter, L. Bürgi, O. Jeandupeux, and K. Kern, "Building supramolecular nanostructures at surfaces by hydrogen bonding," *Angewandte Chemie International Edition*, vol. 39, no. 7, pp. 1230–1234, 2000.

[174] H. Churchill, H. Teng, and R. M. Hazen, "Correlation of pH-dependent surface interaction forces to amino acid adsorption: Implications for the origin of life," *American Mineralogist*, vol. 89, no. 7, pp. 1048–1055, 2004.

[175] J. Bebié and M. A. Schoonen, "Pyrite surface interaction with selected organic aqueous species under anoxic conditions," *Geochemical Transactions*, vol. 1, no. 8, p. 47, 2000.

[176] A. D. Roddick-Lanzilotta, P. A. Connor, and A. J. McQuillan, "An in situ infrared spectroscopic study of the adsorption of lysine to TiO2 from an aqueous solution," *Langmuir*, vol. 14, no. 22, pp. 6479–6484, 1998.

[177] J.-F. Lambert, "Adsorption and polymerization of amino acids on mineral surfaces: A review," *Origins of Life and Evolution of Biospheres*, vol. 38, no. 3, pp. 211–242, 2008.

[178] C. Lomenech, G. Bery, D. Costa, L. Stievano, and J.-F. Lambert, "Theoretical and experimental study of the adsorption of neutral glycine on silica from the gas phase," *ChemPhysChem*, vol. 6, no. 6, pp. 1061–1070, 2005.

[179] D. Costa, C. Lomenech, M. Meng, L. Stievano, and J.-F. Lambert, "Microsolvation of glycine by silanol ligands: A DFT study," *Journal of Molecular Structure: Theochem*, vol. 806, no. 1, pp. 253–259, 2007.

[180] A. Rimola, M. Sodupe, S. Tosoni, B. Civalleri, and P. Ugliengo, "Interaction of glycine with isolated hydroxyl groups at the silica surface: First principles B3LYP periodic simulation," *Langmuir*, vol. 22, no. 15, pp. 6593–6604, 2006.

[181] G. L. Gambino, A. Grassi, and G. Marletta, "Molecular modeling of interactions between L-lysine and functionalized quartz surfaces," *The Journal of Physical Chemistry B*, vol. 110, no. 10, pp. 4836–4845, 2006.

[182] G. L. Gambino, G. M. Lombardo, A. Grassi, and G. Marletta, "Molecular modeling of interactions between L-lysine and a hydroxylated quartz surface," *The Journal of Physical Chemistry B*, vol. 108, no. 8, pp. 2600–2607, 2004.