Peptide Based Inhibitor Design for Interleukin 1 Beta (IL-1β)

By

Ece Bulut

A Thesis Submitted to the

Graduate School of Science and Engineering

In Partial Fulfilment of the Requirements for

The Degree of

Master of Science

In

Chemical and Biological Engineering

Koç University

October 2011

Koç University

Graduate School of Sciences and Engineering

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

Ece Bulut

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Burak Erman, Ph. D. (Advisor)

_____

Seda Keskin-Avcı, Ph. D.

_____

Rıza Kızılel, Ph. D

Date: _____

ACKNOWLEDGEMENTS

I am deeply thankful to my advisor Prof. Dr. Burak Erman for his encouragement, patience and guidance through the development of my thesis. It was a great opportunity to work with him and I accept it as an honor to being a member of his research group.

I owe special thanks to my thesis committee members Asst. Prof. Seda Keskin-Avcı and Dr. Rıza Kızılel for sparing their precious time and for their valuable comments.

I would like to thank the Koç University and the Vehbi Koç Foundation for their financial support during my master studies.

I am grateful to my colleagues Ümmühan Canan and Meriç Ataman for their support and lasting friendships, accompanying me in good and bad. In addition, I especially thank to my friends namely Aslı Çelebi, Büke Boşnak, Hüseyin Can Savaş, Selin Yenel, Sezin Tezcan and Özge Tuncel for their endless support and great friendships, I feel lucky to have them.

Special thanks to Dr.Evrim Besray Ünal for her friendship and constructive suggestions throughout my master studies. I thank to my officemates, Çiğdem Sevim Bayrak, Cemre Kocahakimoğlu and Mehmet Ali Öztürk for making the office such a nice environment.

Last but not least, I am indebted to my family. I would like to thank my parents Demet & Arcan Bulut and my sister Tuğçe Bulut for their continuous support all through my life. It would have been impossible to write this thesis without their love, help and guidance. I dedicate my thesis to them.

OCTOBER 2011

ÖZET

İnterlökin 1 beta (IL-1β), İnterlökin kesici enzim Kaspaz-1 (ICE) tarafından hücre içinde aktif hale getirilen önemli bir inflamatuar sitokindir. Aktivasyonla birlikte IL-1β hücre dışına salgılanır. IL-1β'nın yüksek dozda salınımları romatoid eklem iltihabı ve intestinal inflamasyonla ilişkilendirilmiştir. Bu yüzden L-1β'nın bloke edilmesi bu tarz hastalıkların kontrolünde önemli rol oynamaktadır. IL-1β'nın hücre zarına gömülü reseptörü (IL-1R) ve aksesuar proteini (IL-1RAcP) bulunmaktadır. IL-1β'nın reseptörüne ve aksesuar proteinine bağlanmasıyla bu proteinlerin hücre içindeki kısımları birbirine yaklaşır ve daha çok IL-1β salınımı için hazır hale gelir. Bu çalışmada aktif IL-1β'nın inhibisyonu hesaplamalı bağlama teknikleriyle incelenmiştir. IL-1β'nın reseptörüne bağlanırken aktif özellik gösteren kısımları bağlanma noktası olarak seçilmiştir. Uygun inhibitor adayı bulmak için yaygın olarak kullanılan Genetik Algoritma ve Saklı Markov Modeline dayalı Viterbi Algoritması bağlanma enerjileri ve yapısal özellikleri dikkate alınarak hedef protein üzerine uygulanmıştır. Sonuç olarak, Genetik Algoritma yardımıyla üçlü ve beşli peptid adayları ve Viterbi Algoritması yardımıyla otuz adet yedili peptid adayı bulunmuştur. En uygun adayı bulmak için seçilen peptid dizilimleri Moleküler Dinamik Simülasyonlarından hesaplanan bağlanma enerjileri yardımıyla tekrar elenmiştir. En muhtemel adayların IL-1α'yı da inhibe ettiği gözlemlenmiştir. Ligandların kopma süreçleri atom bazında incelenerek olasılık dağılımları gözlemlenmiştir.

ABSTRACT

Interleukin 1 beta (IL-1β) is a pro-inflammatory cytokine which is activated intracellularly by the cleavage of Caspase-1 (Interleukin converting enzyme, ICE). Upon activation, IL-1β is secreted to the extracellular region. Increased expression of IL-1β is associated with several diseases such as rheumatoid arthritis and intestinal inflammation. Thus in controlling these kinds of diseases, blocking IL-1β has great importance. IL-1β has an embedded receptor (IL-1R) and an accessory protein (IL-1RAcP) in the cell surface. IL-1β binds to its receptor and to the accessory protein which juxtaposes the intracellular domains of its receptors and causes more expression of IL-1β. In this study, inhibition of active IL-1β has been studied extensively by using computational docking tools and techniques. Binding residues of IL-1β to its receptor were accepted as docking regions. Genetic Algorithm (GA) and Viterbi algorithm (VA) based on the Hidden Markov Model were applied to obtain the most suitable inhibitor candidate by considering its binding free energy and secondary structure conformation. As a result, tripeptide and pentapeptide candidates from Genetic Algorithm and thirty heptapeptide candidates from the Viterbi Algorithm were obtained. Elimination of the candidates was achieved by Molecular Dynamics Simulations by calculating binding free energies of the ligands. Most potent inhibitor for IL-1β was observed to inhibit IL-1α as well. The unbinding process of ligands was investigated and their probability distribution graphs were examined.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1. INTRODUCTION

In recent years there has been a great interest in the usage of peptides as drugs due to their higher activity, specificity and lower toxicity than chemical drugs [1]. In this study a peptide inhibitor to IL-1β was designed by using computational docking tools and techniques.

IL-1β, a member of interleukin family, is a pro-inflammatory cytokine which is activated intracellularly by the cleavage of Caspase-1 (Interleukin converting enzyme, ICE). Upon activation, mature IL-1β is secreted to the extracellular region and then binds to its receptors which are embedded in the cell surface. This binding process juxtaposes the intracellular domains of its receptors and causes more expression of IL-1β. Expression of IL-1β is known to be enhanced by inflammatory diseases. Autoinflammatory diseases are defined as a subgroup of inflammatory diseases, which are mainly mediated by IL-1β. Thus in controlling these kinds of diseases, blocking IL-1β has great importance [2]. Chapter 2 gives detailed information about IL-1β and interleukin family diseases.

Identification of a specific peptide sequence as an inhibitor is a complex problem due to large number of possibilities. For even a tripeptide sequence there are 20^3 possibilities with 20 natural amino acids known. To eliminate these possibilities a rational computational approach was needed. In this study, two different algorithms are applied for finding the best candidate. The first one is the Genetic algorithm which commonly finds application in determination of a specific peptide sequence in peptide drug design [3, 4]. The Genetic Algorithm (GA) is a heuristic method that mimics nature and biological evolution. It starts with a set of solutions (chromosomes) which form a population. Then, these solutions are ranked according to the defined fitness function. Here, Autodock binding energies of tripeptides are selected as the ranking criteria. Best solutions from this set are taken iteratively until it finds the top scored candidate [4, 5]. Docking region of the target protein was selected according to an extensive literature overview and supported with the Hotpoint and GNM servers [6-14]. As a result, specific tripeptide and pentapeptide candidates were obtained for the target protein.

Secondly, Viterbi Algorithm based on the Hidden Markov Model was applied in order to find more convenient peptide sequence that binds to the IL-1β. Since inhibition of the target protein occurs in the extracellular region, Lipinski rules are no longer needed to hold. Thus, to increase specificity a longer peptide sequence was needed. Here, we designed an algorithm which can be applied to any protein with any peptide length. The output of the algorithm gives the most probable sequence by considering its binding free energy and secondary structure conformation. This algorithm generates peptide inhibitors by docking its residues pair by pair to a selected path, called the Viterbi path, along the protein surface. The algorithm is based on the Hidden Markov Model (HMM). In the Markov chains, outcome of any process affects the next process in a chainlike manner. The Hidden Markov Model is a specific type of Markov Chain in which the system is assumed to have unobserved (hidden) states. The state is not directly visible to the observer but visible according to output depending on the states [15]. Successful implementations of this model on biological data are well known [16-21]. In our model hidden states are the dipeptides while our observable states are the secondary structures. The algorithm consists of five steps: 1) determination of the binding path 2) partitioning the path into grids 3) docking to the grids and evaluating its binding free energies 4) Characterizing the ψ/φ propensities of the dipeptides 5) Using the outcome of these five steps in Viterbi decoding. To design a heptapeptide sequence the algorithm was implemented on IL-1β several times. Total of thirty different candidates were obtained. The algorithm was also tested on several proteins including IL-1α. These two algorithms were defined explicitly in Chapter 3.

In Chapter 3, elimination of the ligands by Molecular Dynamics Simulations (MD) was also explained in detail. Molecular Dynamics Simulations are one of the most preferred computational methods for understanding behavior of biological molecules. They give detailed information about thermodynamic properties of biological molecules, fluctuations and conformational changes of proteins [22]. Thus to understand binding mode of the preferred ligands more accurately, MD simulations were carried out. The unbinding energies of peptides examined in detail by using Jarzynski Equality in Steered Molecular Dynamics Simulations [23].

In the Chapter 4 two algorithm results with MD simulations were discussed. As a result, most convenient inhibitor was designed on IL-1β. A convenient ligand should possess low binding energies to the target protein and should structurally fit to the target site like a hand and glove.

Most of the peptides designed on the interleukin-1 family target the receptors embedded in the cell surface. We targeted active IL-1β in the extracellular region which will increase the specificity and will not interfere the inner cell activities of IL-1 receptor. By doing this we increased the specificity and side reactions occurring from the deactivation of interleukin receptor are prevented [12, 24].

In order to obtain the most convenient ligand from the remaining candidates, all unbinding process was investigated atom by atom. Due to their probability distribution graphs candidate with a sequence E-A-T-V-I-I-I (GLU-ALA-THR-VAL-ILE-ILE-ILE) gave greater binding and unbinding energies and selected as the most potent inhibitor. The ligand unbinds in a slow fashion, gives high binding free energies and specific to target protein.

CHAPTER 2

2. LITERATURE OVERVIEW

2.1. INTERLEUKIN FAMILY

Cytokines are small cell-signalling proteins that regulate immune responses and inflammatory reactions by a complex network. More than 50 cytokines have been identified including interleukin, chemokine, growth factor and interferon families [25].

Among all cytokines, Interleukin-1 family members are well defined as the most potent molecules of an innate immune system. All interleukin family members play a central role in the regulation of immune and inflammatory responses [25].

The Interleukin-1 family is comprised of eleven members which are listed in Table 1. IL-1β is the most studied one due to its mediator role in inflammatory response. In humans, blocking IL-1β has an important role in clinical medicine [2, 26].

Table 1 Interleukin -1 family [2]

New Name	Other Name	Property
IL-1F1	IL-1α	Agonist
IL-1F2	IL-1β	Agonist
IL-1F3	IL-1Ra	Receptor Agonist
IL-1F4	IL-18; IFN-γ-inducing factor	Agonist
IL-1F5	FIL1δ	Anti-inflammatory
IL-1F6	FIL-1ε	Agonist
IL-1F7	IL-1H4, IL-1ζ	Anti-inflammatory
IL-1F8	IL-1H2	Agonist
IL-1F9	IL-1ε	Agonist
IL-1F10	IL-1Hy2	Receptor Agonist
IL-1F11	IL-33	Agonist

2.1.1. INTERLEUKIN FAMILY RELATED DISEASES

Interleukin 1 is a multifunctional pro-inflammatory cytokine and has numerous effects in the immune system such as fever, increased acute phase response, increased lymphokine production, proliferation of muscle cells and mesangial cells and increased HIV-1 gene expression. After 1984, IL-1 was well identified and its family was described as the interleukin family. Among them, IL-1α and IL-1β are the most examined agonists [27].

Each of the IL-1α, IL-1β agonists and IL-1Ra antagonist (receptor agonist) binds the same receptor IL-1RI. Binding of IL-1α or IL-1β to the receptor is known as the early steps of signal transduction. Thus, blocking this binding interaction might be a useful target for the discovery of new drugs [8].

Table 2 lists specific cures for blocking interleukin-1 activities. It is evident that signalling of IL-1β only occurs in a trap which consists of receptor (IL1-R1) and co-receptor chain (IL-1RAcP) proteins [2]. The other ways of neutralizing IL-1β do not transmit a signal but deactivate its functions. Also the last row of Table 2 shows a specific peptide designed for the receptor which blocks all interleukin-1 family agonist and antagonist activities. Binding affinities for each member can be seen from the mechanism of the action column of the Table 2.

Anakinra and IL-1Ra are the receptor antagonists for both IL-1β and IL-1α. Anakinra is just a recombinant form of IL-1Ra. (Natural IL-1Ra is glycosylated while anakinra is not). IL-1Ra binds the receptor tightly and blocks the activity of either IL-1α or IL-1β, but does not transmit a signal. Anakinra is a natural activity blocker and has been approved for treating signs, symptoms and joint destruction of rheumatoid arthritis (autoimmune disease) for more than ten years. It is now the standard therapy for patients with systemic-onset juvenile idiopathic arthritis, refractory adult Still's disease and several systemic and local inflammatory diseases. Today, these chronic inflammatory diseases are known as autoinflammatory diseases. The term inflammatory comes from the inflammasome which means complex interacting intracellular proteins that initiate the autocatalysis of procaspase-1 into mature–active enzyme. Caspase-1 is known as the Interleukin converting enzyme (ICE), which cleaves precursor forms of IL-1β and IL-18 and activates them [2].

Table 2 Specific therapies for blocking IL-1 activities [2]

Reagent	Composition	Mechanism of action	Specificity
Anakinra	**IL-1Ra**	**Binds to IL-1RI> IL-1α> IL-1β**	**IL-1α and IL-1β**
IL-1Ra	**IL-1Ra**	**Binds to IL-1RI> IL-1α> IL-1β**	**IL-1α and IL-1β**
IL-1 Trap	**IL-1RI + IL-1RAcP**	**Neutralizes IL-1β>IL-1α>IL-1Ra**	**IL-1β (IL-1α)**
Soluble IL-1RII	**IL-1RII**	**Neutralizes IL-1β>IL-1α**	**IL-1β**
Anti-IL-1β	**Monoclonal antibody**	**Neutralizes IL-1β**	**IL-1β**
Anti-IL-1RI	**Monoclonal antibody**	**Blocks IL-1RI**	**IL-1α and IL-1β**
Anti-IL-1RAcP	**Monoclonal antibody**	**Blocks IL-1RacP**	**IL-1α and IL-1β, IL-33**
Peptide Antagonist	**RYTVELA**	**Blocks IL-1RI**	**IL-1α and IL-1β**

Autoinflammatory diseases are known as a subgroup of inflammatory diseases, which are mainly mediated by IL-1β. Thus in controlling these kinds of diseases, blocking IL-1β has a great importance. Most of the autoinflammatory diseases occur due to the mutations in proteins that comprise the inflammasome and result in increased secretion of IL-1β [2].

The Trap in Table 2 has been approved for the treatment of autoinflammatory diseases such as Muckle-Wells syndrome and familial cold-induced autoinflammatory syndrome (FCAS) listed in Table 3. Also, destruction of the insulin producing beta cell appears to be an IL-1β-mediated autoinflammatory disease [2].

The difference of autoinflammatory diseases from autoimmune diseases comes from their periodicity, strong associations with exogenous triggering events and lack of associations with class II MHC haplotypes. Another significant difference is that autoinflammatory diseases are responsive to IL-1β blockade, while autoimmune diseases are responsive to TNF-α neutralization. All autoinflammatory diseases are listed in Table 3 [2].

Table 3 Autoinflammatory Diseases [2]

Autoinflammatory diseases
Familial Mediterranean fever
Familial cold autoinflammatory syndrome (FCAS)
Muckle-Wells syndrome
Neonatal-onset multi-inflammatory disease (NOMID)
Mevalonic aciduria
Hyper IgD syndrome
Adult-onset Still's disease
Systemic-onset juvenile idiopathic arthritis
Schnitzler's syndrome
Anti-synthetase syndrome
TNF receptor–associated periodic syndrome
Macrophage activation syndrome 2
Behcet's syndrome
Normocomplementemic urticarial vasculitis
Pericarditis
PAPA syndrome
Blau's syndrome
Sweet's syndrome
Urate crystal arthritis (gout)
Type 2 diabetes

2.1.2. RECEPTOR TYPES OF INTERLEUKIN FAMILY

The interleukin-1 receptor family has ten receptor members, some of which are shown in Figure 1 and all are listed in Table 4. The most important receptor types for IL-1β are known as IL-1RI, IL1-RII and IL1-RIII (IL-1 RAcP). IL-1β binding to its receptors occurs in vitro which causes more IL-1β production within the cell region [2, 26].

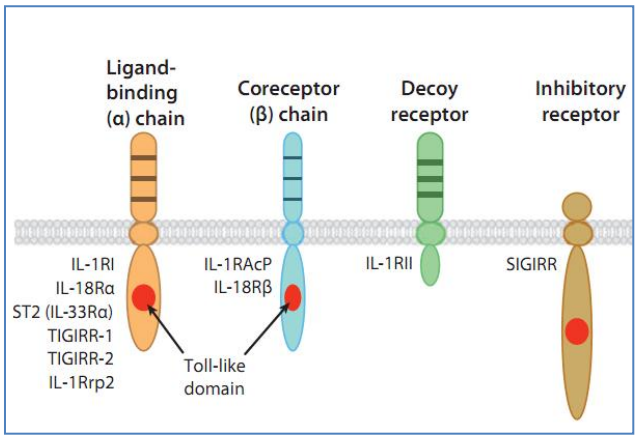

Figure 1 IL-1 family of receptors [2].

For IL-1β, the most important receptors are IL-1RI and its co-receptor IL-1RacP. They contain toll-like domains (TIR) which are crucial members in the initialization of signalling [2, 26].

Table 4 Interleukin -1 receptor family [2]

Name	Designation	Ligands	Co-receptor
IL-1RI	**IL-1RI**	**IL-1α, IL-1β, IL-1Ra**	**IL-1RAcP (IL-1R3)**
IL-1RII	**IL-1R2**	**IL-1β,IL-1β precursor**	**IL-1RAcP (IL-1R3)**
ST2/Fit-1	**IL-1 R4 (IL-33Ra)**	**IL-33**	**IL-1RAcP (IL-1R3)**
IL-18Rα	**IL-1R5**	**IL-18,IL-1F7**	**IL-18Rβ (IL-1R7)**
IL-1Rrp-2	**IL-1R6**	**IL-1F6,IL-1F8,IL-1F9**	**IL-RAcP (IL-1R3)**
TIGIRR-2/IL-1RAPL	**IL-1R8**	**unknown**	**unknown**
TIGIRR-1	**IL-1R9**	**unknown**	**unknown**
SIGIRR	**TIR8**	**unknown**	**unknown**

2.1.2.1. IL-1β AND ITS RECEPTORS

Active IL-1β can form five different complexes with the receptors but only one complex will transmit a signal. These five complexes are numbered and can be seen in Figure 2 as 1, 2, 13, 14, 15 and 16. Binding of IL1-RI and IL1RacP with IL-1β is similar to the binding of IL1-RII and IL1RacP with IL-1β (numbered as 15). But IL1-RII lacks the intracellular TIR domain necessary for signalling. The formation of the heterodimer complex with its receptors induces signalling since the juxtaposition of two TIR domains enables the recruitment of myeloid differentiation primary response protein 88 (mYD88), IL-1RI associated kinase 4 (IRAK4), TNFR associated factor 6 (TRAF6) which are indicated as the important signalling intermediates. Activation of the nuclear factor κB (NF-κB) is also demonstrated in Figure 2 [2, 13] .

IL-1RII blocks IL-1β responses in two complementary ways. First, it captures IL-1β and prevents its interaction with IL-1RI. Secondly, it sequesters IL-1RAcP necessary for signalling [13].

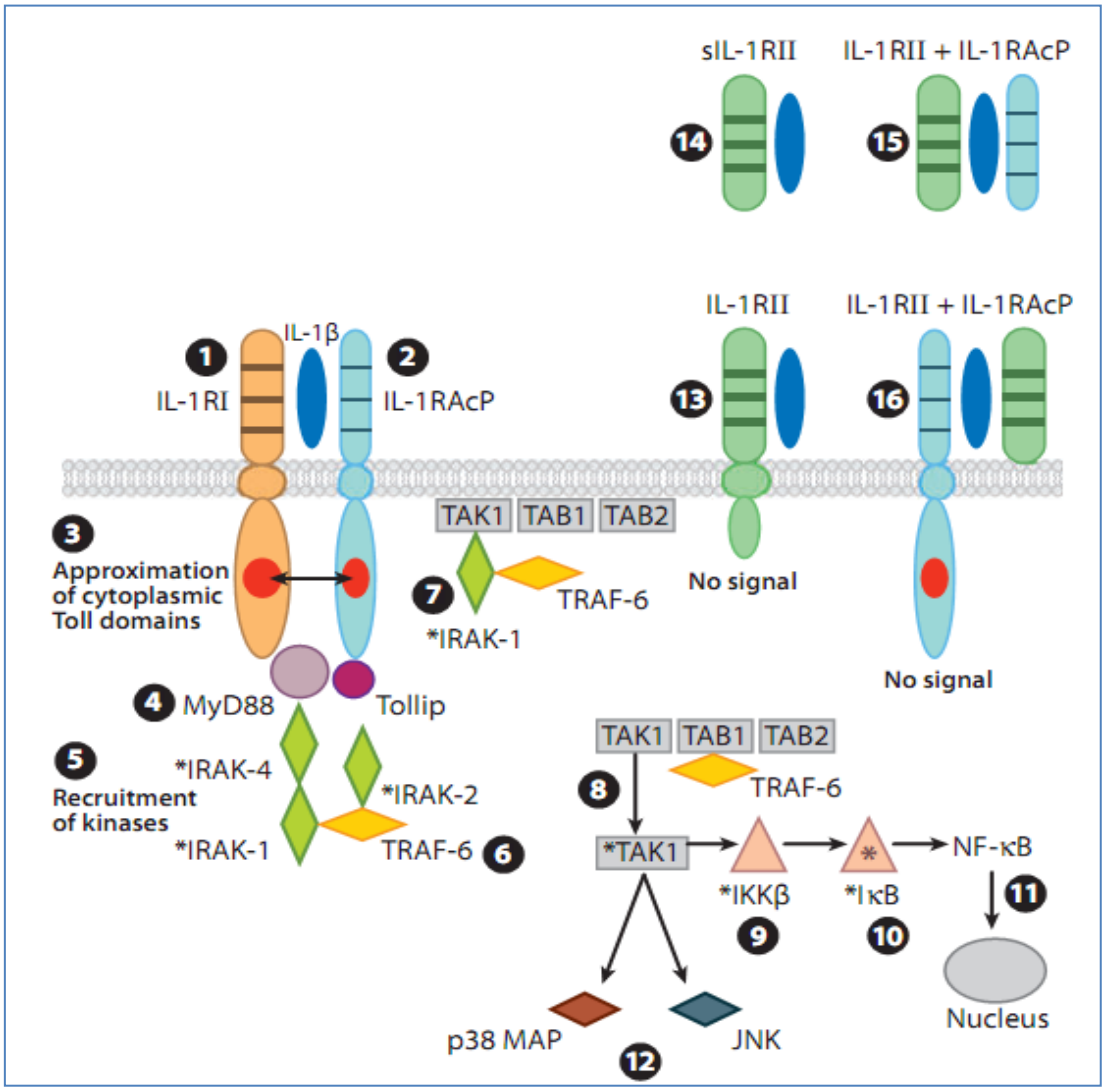

Figure 2 IL-1 signal transduction and decoy receptors [2].

The first complex is numbered as 1 and 2. (1, 2) IL-1β binding to its receptors causes a heterodimeric complex. This complex causes the Toll domains to approximate in the intracellular region (3). MyD88 and Tollip proteins are recruited with this approximation (4). Step 4 triggers phosphorylations of kinases and recruitment of TRAF-6 protein (5, 6). Phosphorylated proteins migrate to membrane and associate with TAK1 (TGF-β activated kinase 1), TAB1 (TAK1 binding protein) and TAB2 complexes (7). These complexes then migrate to cytosol and TAK1 is phosphorylated following the ubiquitination of TRAF-6 (8). Phosphorylated TAK1 activates IKKβ (9), and phosphorylated IKKβ phosphorylates IκB (10). Phosphorylated IκB degrades, releasing NF-κB, which enters the nucleus (11). In addition to the phosphorylation of IKKβ, TAK1 also activates mitogen-activated protein kinase (MAPK) p38 and JNK (12). On the surface of the cell, IL-1RII, a decoy receptor, may also bind IL-1β (13), but this complex does not recruit IL-1RAcP, and there is no signal.

Extracellular domains (soluble or sIL-1RII) of the IL-1RII bind IL-1β and neutralize its activity (14). sIL-1RII can also bind IL-1β and form a complex with soluble IL-1RAcP (15) or cell-bound IL-1RAcP (16). In the last two complexes, IL-1β is not able to bind to IL-1RI and therefore cannot transmit a signal [2].

2.1.3. STRUCTURAL EXAMINATION OF IL-1β AND IMPORTANT RESIDUES IN BINDING

2.1.3.1. SYNTHESIS OF IL-1β

Synthesis of IL-1β occurs in a cycle which starts by the formation of its heterodimeric complex with its receptors. It then continues with its production into pro- IL-1β which is cleaved by Caspase-1 (interleukin cleaving enzyme) into active IL-1β. Active IL-1β is secreted to the extracellular region and again mainly interacts with its receptors [2]. The synthesis process is explained in detail in Figure 3. Other extracellular interactions of IL-1β will be discussed in the next section.

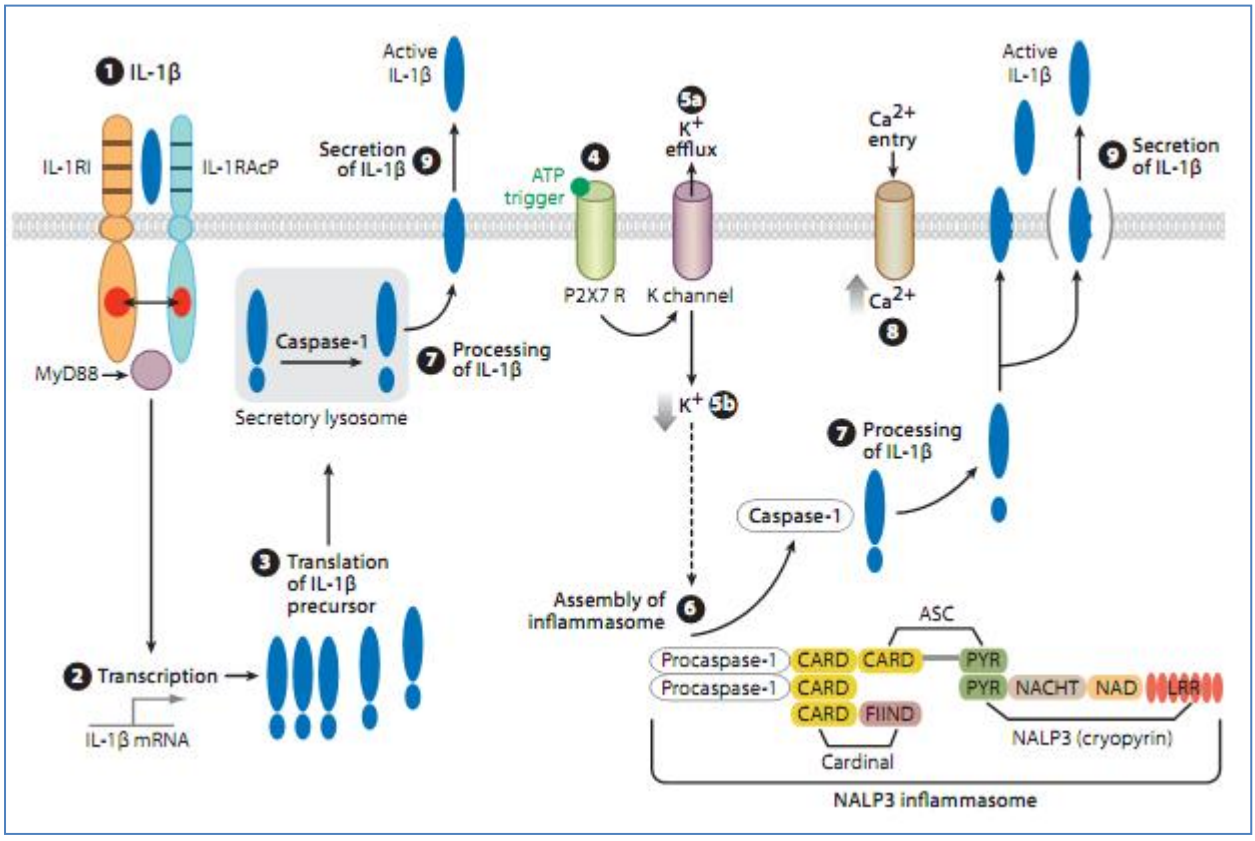

Figure 3 Synthesis of IL-1β [2].

Primary blood monocytes or tissue macrophages are activated by IL-1β. In step 1, activated IL-1β forms a complex with its receptors. This complex enables the receptors embedded parts to approximate and recruitment of MyD88 protein. Transcription and translation occurs afterwards (step 2 and 3). Translation takes place in cytosol, not in the endoplasmic reticulum. The activated monocyte/macrophage releases ATP into the extracellular space. Activation of P2X7 by ATP (step 4) causes efflux of potassium from cell (step 5a), decreasing intracellular levels of potassium (step 5b). The reduction in potassium levels results of assembling the components of NALP3 inflammasome (step 6). These assembled components of the inflammasome initiate the activation of procaspase 1 to caspase-1. Active caspase-1 cleaves the IL-1β precursor and activates it either in cytosol or in secretory lysosome (step 7). Releasing mature IL-1β from the cell occurs with the increment of the calcium levels (step 8 and 9). The rise in intracellular calcium activates phosphatidylcholine-specific phospholipase C and calcium-dependent phospholipase A, which facilitate the secretion of IL-1β (step 9) with exocytosis of the lysosomal contents.

2.1.3.2. COMPARISON OF IL-1β AND IL1-α

IL-1β and IL-1α are similar forms of IL-1. They are synthesized as 31-kD precursors and cleaved by proteases to their 17-kD mature form [2, 28]. These members carry similar functional and structural properties. For instance, they share 25% amino acid homology and bind to the same receptor (IL-1RI).

IL-1β

The molecule contains twelve antiparallel β strands where six of them (1, 4, 5, 8, 9, 12) are antiparallel barrels (Figure 4a). The molecule contains three similar fragments each containing two pairs of β strands (Figure 4b). These three pairs form the six stranded barrel. The inner surface of the protein consists of 24 hydrophobic side chains, whereas both ends of the barrel have polar residues. The N and C termini of the molecule are close to each other at the open end of the barrel. Two of β hairpins among five are located in the open end. β bulge loop is located between strands 4 and 5 which plays an important role in receptor binding. The IL-1Ra antagonist lacks in β bulge region and its interactions.

IL-1α

The molecule is very similar to IL-1β with the same β barrel properties. The major difference is an NH_2 terminal extension of 14 residues beyond the NH_2 terminus of IL-1β. As indicated in Table 2, the receptor trap prefers IL-1β over IL-1α [28, 29].

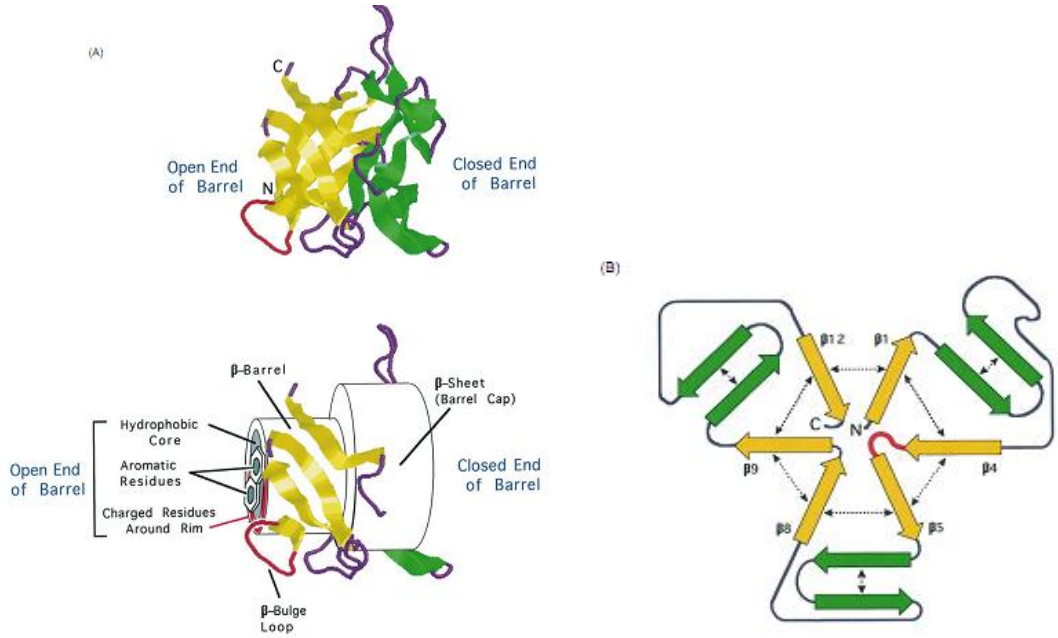

Figure 4 Human type of IL-1β structure in β strands [27].

a) Six of the barrels: The other end of the barrel is "open (yellow)" consists of hydrophobic core, charged and aromatic residues. b) 3-fold pseudo symmetry of β strands [27].

2.1.3.3. BINDING OF IL-1β TO IL-1RI IN COMPARISON WITH IL1Ra

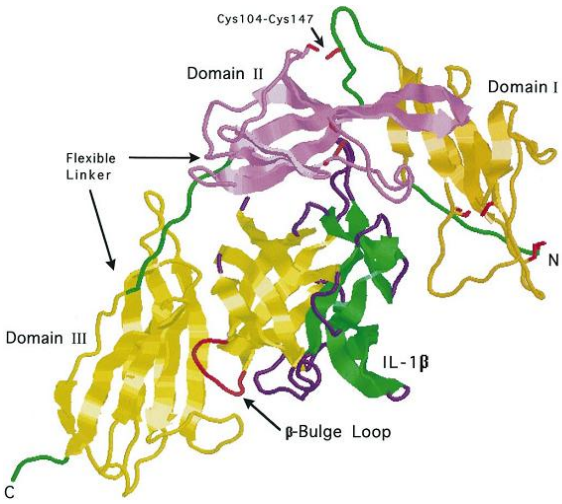

Figure 5 Binding of IL-1β to IL-1RI with three domains and β bulge [27].

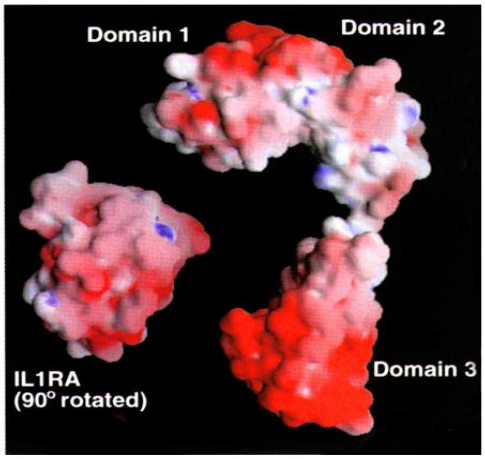

Figure 6 Solvent accessible surface of IL-1R with three domains and IL1Ra rotated for binding site clearance [14].

The IL-1 Receptor consists of three main domains and was found by direct mutagenesis that IL-1β has two binding regions (A, B) (Figures 5, 6). Both sites contact with the receptor. Site A consists of domains 1, 2 whereas site B is domain 3 [8].

IL-1β interacts with the IL-1RI such that the uncapped end of the barrel is exposed to domain 3 of the receptor in a naturally complementary way (Figure 5). 2.088 Å of ligand surface is buried in the receptor domain by the interaction. This interaction is reported to be mediated by

13 salt bridges and 7 H bonds. The remaining domains also contribute with 10 salt bridges and 7 H bonds resulting in 1.001Å more buried area [27].

The important residues for binding on site A were identified as: 11, 13-15, 20-22, 27, 29-36, 38, 126-131, 147 and 149. The most critical ones **Arg 11** and **Gln 15** contact domain 2 while **His 30** and **Gln 32** contact with the domain 1-2 junction. The authors [8] claim that an important part of binding energy for this site comes from van der Waals forces. Site B is formed by residues 1-4, 6, 46, 48, 51, 53-54, 56, 92-94, 103, 105-106, 108, 109, 150 and 152. Site B only makes contact with domain 3.The authors also claim that for site B hydrophilic (4, 48, 51, 93, 53, 105, 108) and hydrophobic residues (6, 46, 56, 150) form most of the binding energy [8].

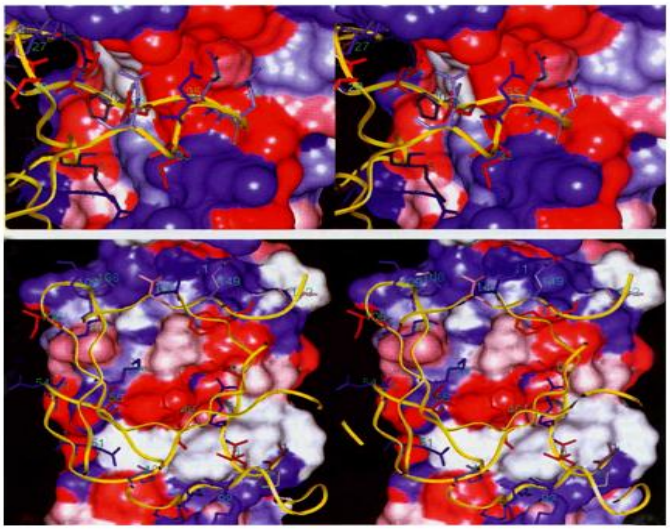

Figure 7 Stereoscopic surface representations of s-IL-1R and ribbon diagrams of IL-1β.

Top: Site A Bottom: Site B. IL-1β ribbon is yellow, hydrophobic residues are coloured red and hydrophilic residues are coloured magenta [8].

Figure 7 represents defined sites for binding (sites A, B). The receptor is coloured according to its hydrophobic and hydrophilic regions which are claimed to play an important role in binding energy.

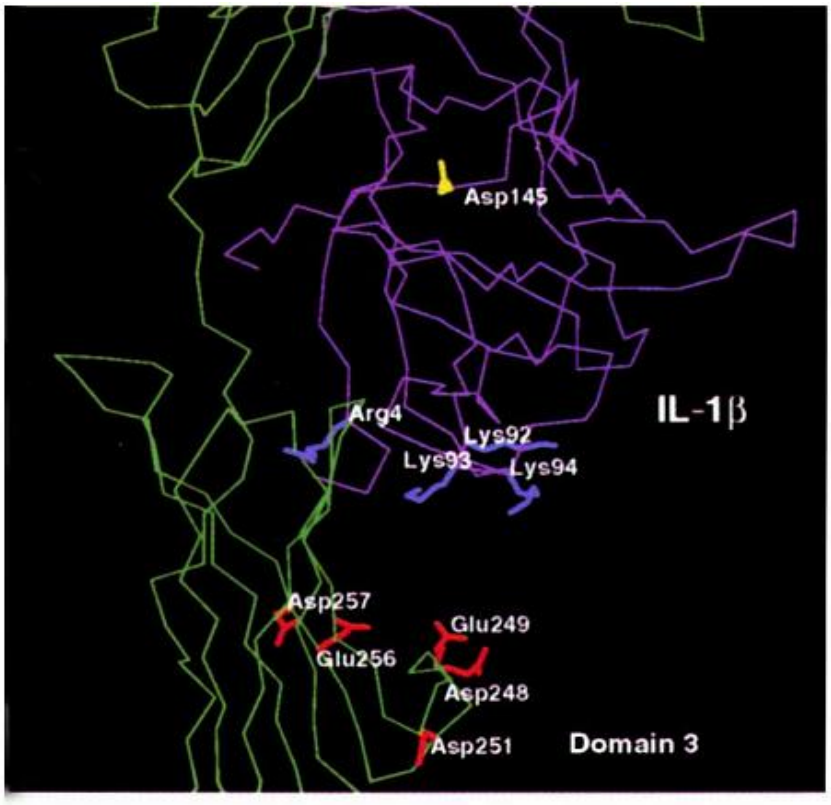

Figure 8 IL-1β with its receptor [14].

In Figure 8, IL-1β is indicated as magenta and receptor is green. A region of positively charged residues on IL-1β and negatively charged residues on receptors are indicated. These two regions interact directly [14].

Figure 8 shows the interaction between negatively charged residues on receptor and positively charged residues on IL-1β. Indicated residues show a high affinity to each other upon binding. Asp 145 residue is claimed to have an important role for binding to accessory protein (IL-1 RAcP) which does not interact with either domain 3 or 1-2 [14].

Similar to IL-1β, IL1–RI and IL-Ra are also important to determine hot points for binding. Figure 9 and Figure 10 indicate similarities between antagonist protein (IL1Ra) and IL-1β proteins. Important residues of IL1-Ra are shown in Figure 10.

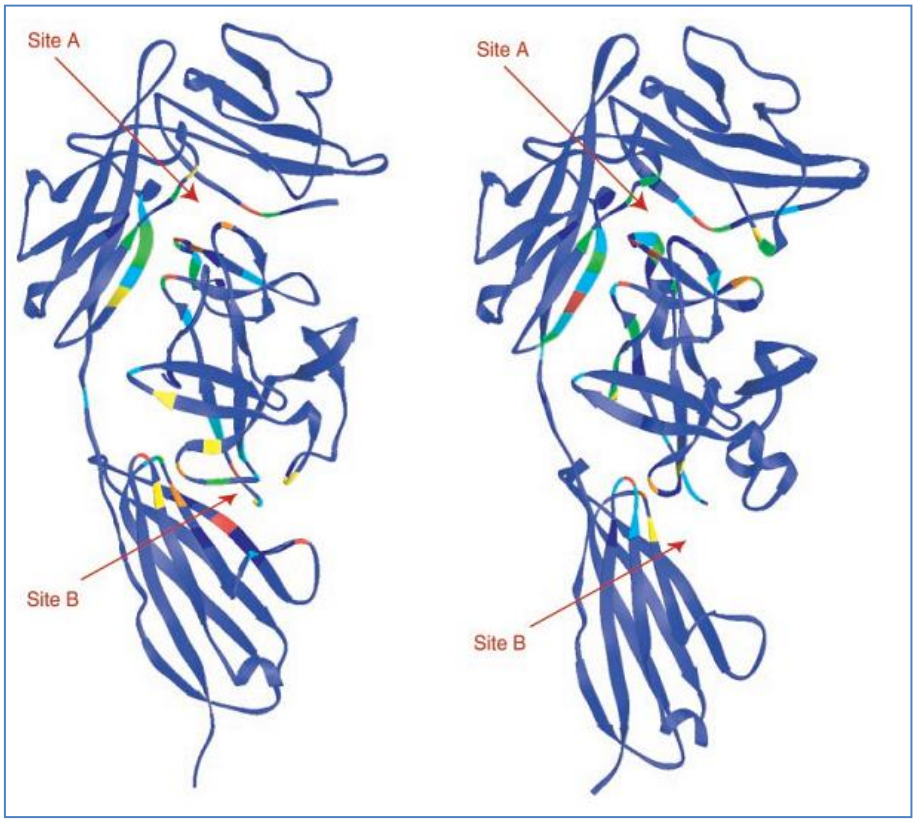

Figure 9 Structures of IL-1RI&IL-1β and IL-1RI&IL-1Ra [10].

In Figure 9, IL-1RI & IL-1β complex (left) and IL-1RI & IL-1ra complex (right) look very similar. The difference comes from the third domain in IL-1Ra complex which swings 20° away from the ligand. If presented, cell membrane would be at the bottom of the Figure. The amino terminus of the receptor is at the top right, and the carboxyl terminus at the bottom. Binding Site A lies between the first and second domains whereas Site B, which is utilized by IL-1β, but not by IL-1Ra, lies on the face of the third domain [10].

Figure 9 shows structures of IL-1RI&IL-1β and IL-1RI&IL-1Ra with binding sites indicated as A and B. Conformational differences between IL-1β and IL-1Ra can be observed from this Figure.

The importance of domain 3 was investigated by superimposing IL-1β and IL-1Ra. It was well indicated by Auron et al. that domain 1-2 interactions were the same for both IL-1β and IL-1Ra. The difference comes from the domain 3 reactions. As indicated in Figure 8, domain 3 has highly negative residues. A rotation of flexible domain 3 (20°) would bring these sites together. IL-1Ra binds to domain 3 with only 1 salt bridge and 4 H bonds which results in 1.774 Å buried area. The total buried surface area reduced from the 3.089 to 1.774 Å in IL-

1Ra binding. The reduction of interaction is closely related to the absence β bulge loop region found in IL-1β. Moreover, IL-1Ra lacks several interactions with domain 2 [14, 27].

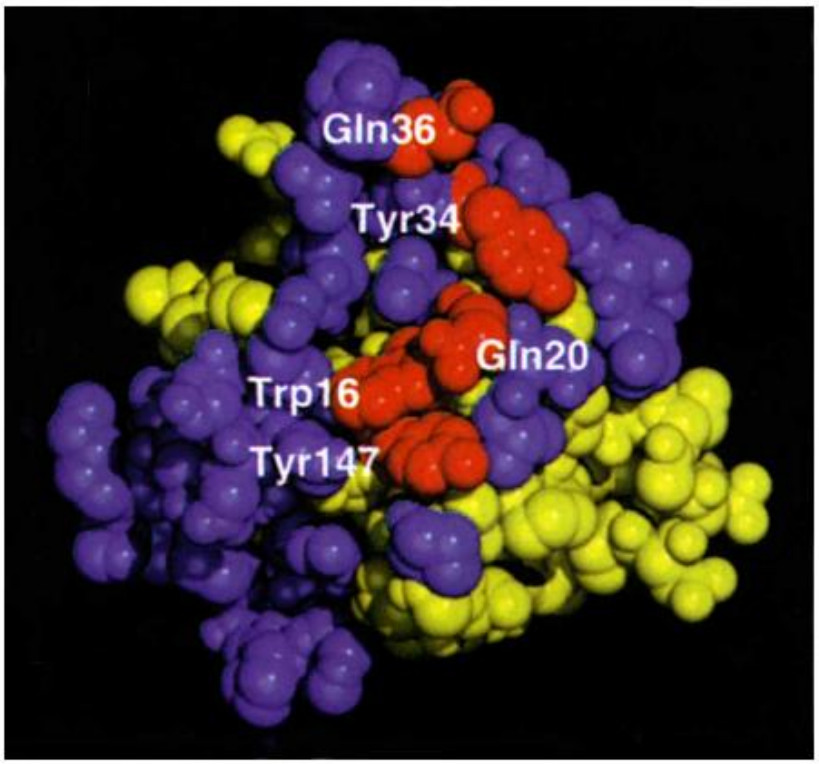

Figure 10 Space filling model of bound IL1-Ra [14]

The red residues shown in Figure 10 are important for interaction, other residues that are buried by the receptor are in blue and those that do not interact are given as yellow [14].

Also, Figure 10 demonstrates important residues determined by site-directed mutagenesis for binding of IL-1Ra (antagonist) to the receptor (IL-1RI). It was found by removing domain 3 from receptor that this domain is crucial for receptor-agonist binding since its absence reduce binding energy more than receptor-antagonist energy. Thus domain 3 is crucial for binding agonists (IL-1β and IL-1α), not antagonist [14].

2.1.3.4. A NONSIGNALLING COMPLEX (IL-1β bound to IL-1RII and IL-1RAcP)

IL-1β bound to IL-1RII and IL-1RAcP does not transmit a signal. To understand how signal transmission occurs to intracellular region, several studies have been undertaken.

Similar to IL-1RI, both IL-1RII and IL-1RAcP consists of three immunoglobulin-like domains (D1–D3) that resemble a question mark. The D1 domain of IL-1RAcP was observed to make no interaction with IL-1β–IL-1RII. IL-1RI and IL-1RII had similar interactions with

the D1D2 domain which acts as a single module. The D3 domain is believed to have flexible linkers for both which curls around IL-1β. For the trimer complex, important residues for binding were indicated as Val 214-Asn 219 from the linker, Tyr 130, Ile 131, Tyr 133 from D1D2 module and Phe 252 and Thr 287 from D3 which can be seen in Figure 11 [13].

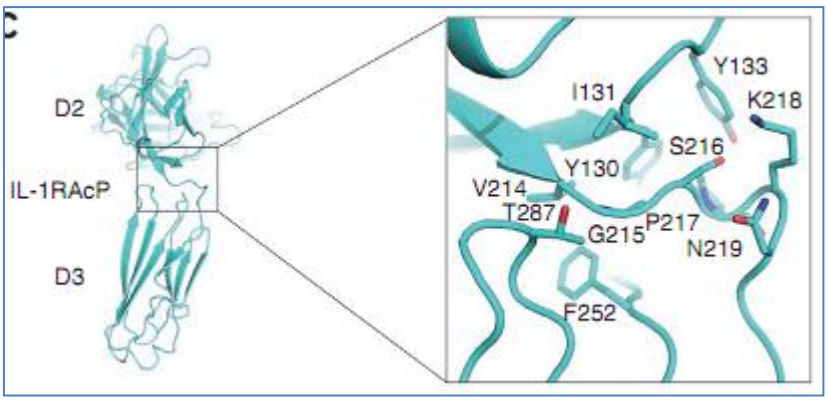

Figure 11 Important residues of IL-1RAcP for binding [13].

Similar residues such as **Arg 11, Gln 15** and **Gln 32** were observed to have interactions with IL1-RII as in IL1-RI. Both IL-1RI and IL-1RII were found to bind IL-1β at sites A and B, whereas IL-1RI binds IL-1Ra only at site A since site B in the IL-1Ra–IL-1RI complex is absent. However, IL-1RI was observed to bind IL-1Ra with higher affinity than it binds to IL-1β. This information indicates that site A determines most of the binding energy. But in case of the IL-1RII - IL-1β complex the preference of ligand changed. It was observed to bind IL-1β with higher affinity than IL-1Ra. Consistent with that, site I between IL-1β and IL-1RII had 19 hydrogen bonds, whereas there were only 7 hydrogen bonds in the modelled site A between IL-1Ra and IL-1RII [13].

As can be seen in Figure 12, IL-1β–IL-1RII complex was observed to interact with IL-1RAcP through a composite binding surface, which can be divided into site III between IL-1β and IL-1RAcP and site IV between IL-1RII and IL-1RAcP. Figure 12 also demonstrates important residues for binding.

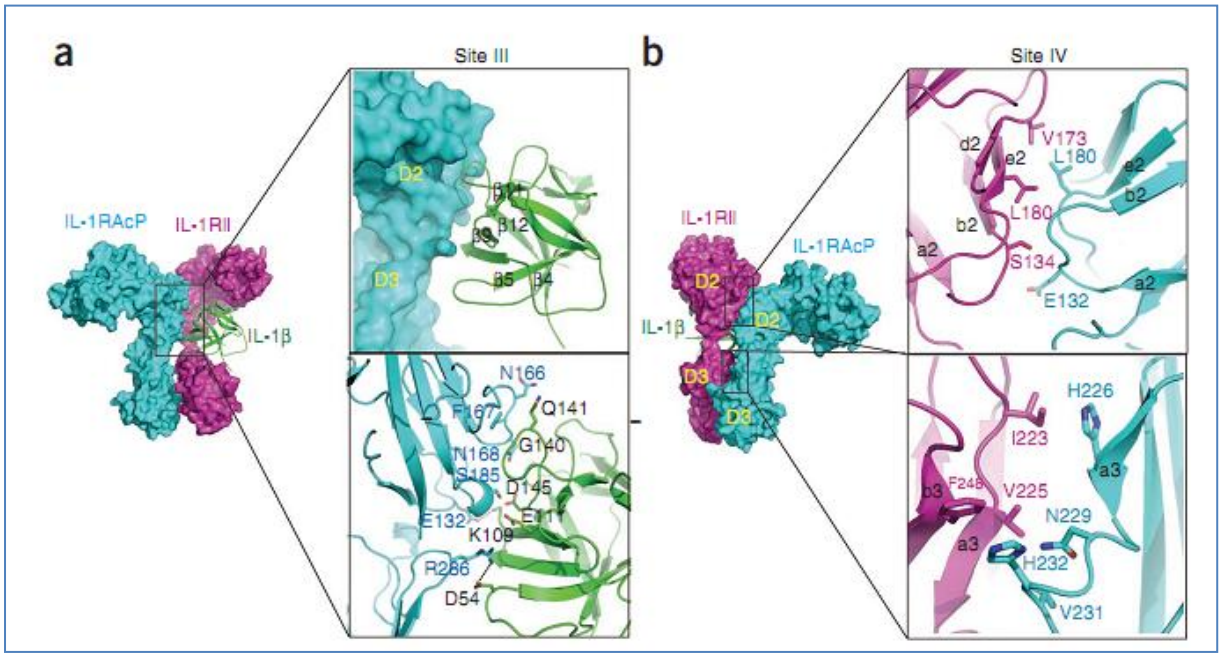

Figure 12 Binding interface between IL-1β–IL-1RII and IL-1RAcP [13].

Figure 12: (a) Site III between IL-1β and the D2 and D3 domains of IL-1RAcP. (b) Site IV between IL-1RII and IL-1RAcP composed of D2–D2 and D3–D3 interactions [13].

Residue Asp54 from IL-1β formed a salt bridge with Arg286 of IL-1RAcP. Residue Lys109 formed a salt bridge with Glu132 of IL-1RAcP, and Glu111 had hydrophilic interaction with Ser185 of IL-1RAcP. Van der Waals interactions between Gly140 and Gln141 of IL-1β and Asn166, Phe167 and Asn168 of IL1RAcP were observed. Hydrophilic interaction between Asp145 of IL-1β and Ser185 of IL-1RAcP noted [13].

At the trimer complex it was demonstrated that D3 domains of IL-1RI and IL-1RII facing IL-1RAcP was not exactly identical which shows the difference between IL-1β–IL-1RI and IL-1β–IL-1RII in their binding affinities for IL-1RAcP (6.38 nM and 2.36 μM). It was indicated that in the signalling IL-1β–IL-1RI–IL-1RAcP complex, IL-1β and the D2 and D3 domains of IL-1RI would form a most suitable composite surface to recruit IL-1RAcP in a nearly perpendicular direction, bringing the D3 domains of both receptors closer. Then, the intracellular TIR domains of both receptors would approximate to transmit signal [13].

Also it was concluded that the absence of Site B between IL-1Ra and the D3 domain of IL-1RI, prevents binding of the IL-1RAcP to IL-1Ra- IL-1RI complex [13].

2.1.4. EXTRACELLULAR INTERACTIONS OF IL-1β WITH OTHER PROTEINS

Mature or active IL-1β is secreted to the extracellular region after it is cleaved by Caspase-1. To understand its behaviour in extracellular region it is also important to consider interactions with other proteins. By this way, any possible side-reactions can be determined from inhibition of IL-1β.

To understand the extracellular interactions of IL-1β, String database [30] and Cytoscape pathway program were used [31]. Fifty neighbouring proteins considered as a limiting parameter for the String database to understand all interactions of IL-1β. Figure 13 shows nine important interactions for IL-1β. Caspase 1 and Caspase 4 will not be considered for being intracellular proteins. The receptor proteins (IL1-RII, IL1-Ra and IL1-RI) were discussed in detail within the previous section and will not be considered in this part.

Figure 14 and Figure 15 show important residues from the String database. Figure 15 shows binding proteins with blue arrows. Intracellular interactions with proteins will not be considered such as: IRAK-1, MYD88, TRAF6, MAP14, CASP3 and NFKB1.

ADRB2 protein is known as b2-adrenergic receptor whose pathway of signalling is related to IL-1β. It was found that IL-1β enhances b2-adrenergic receptor expression in human airway epithelial cells by activating PKC (protein kinase C). Since it is an intracellular interaction, this pathway will not be considered either [32].

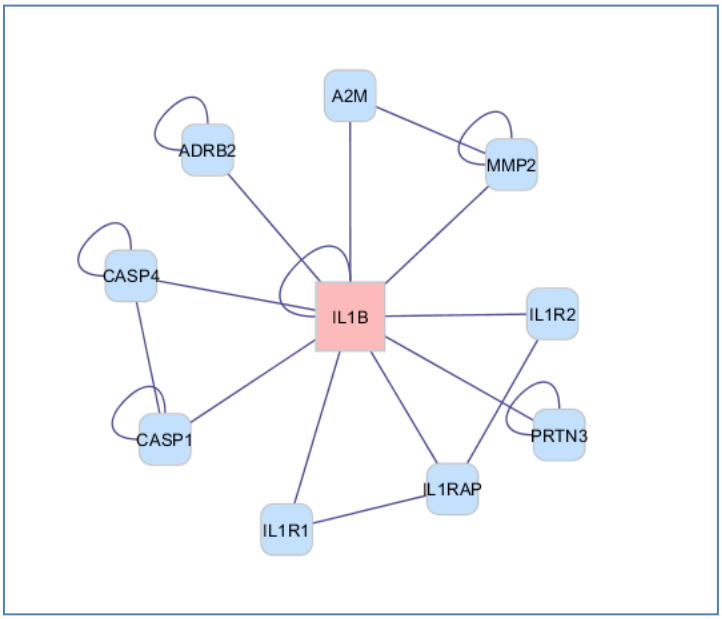

Figure 13 Important proteins for IL-1β from the Cytoscape program [31].

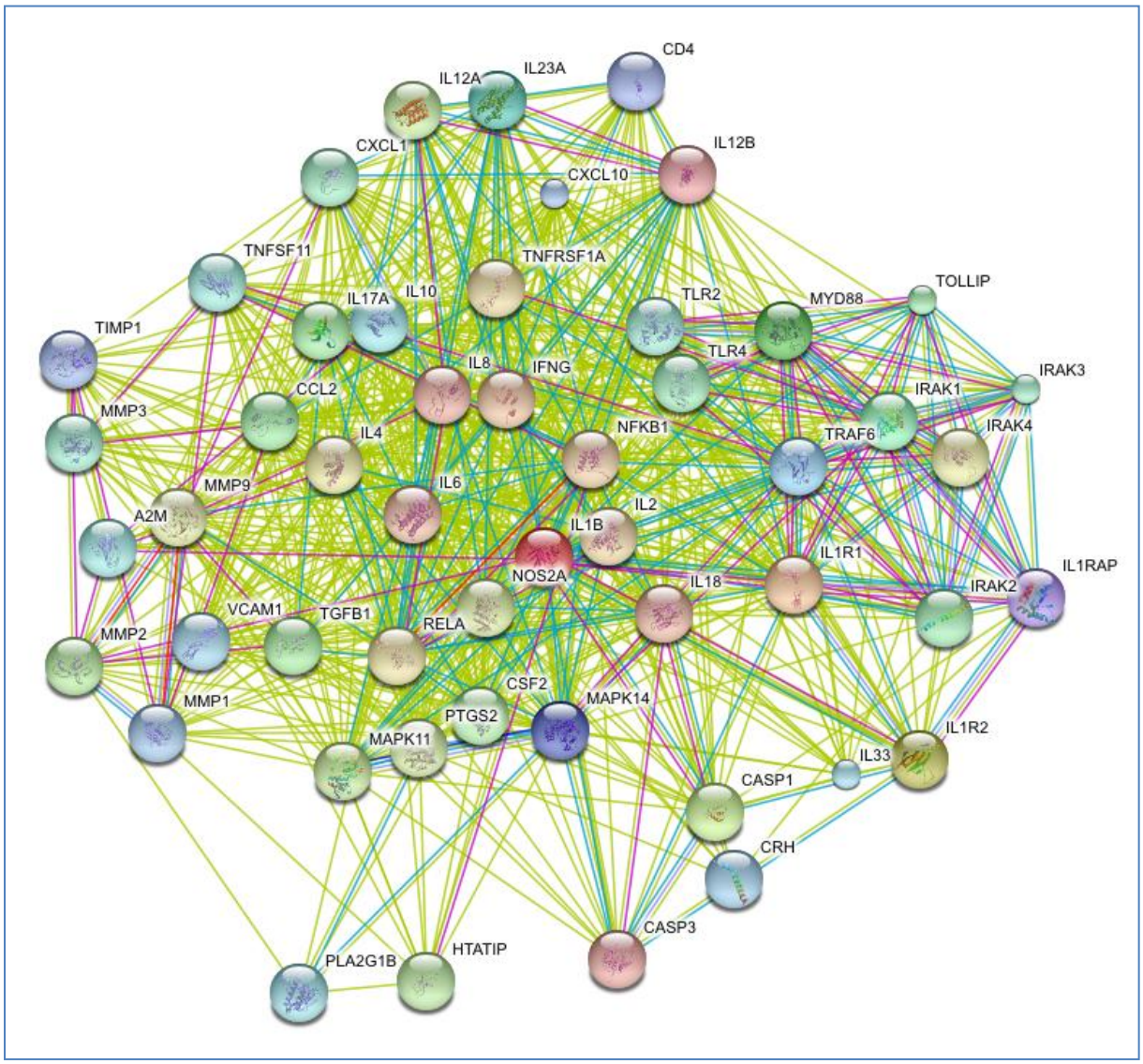

Figure 14 Important proteins for IL-1β from String database (50 neighbouring proteins) [30].

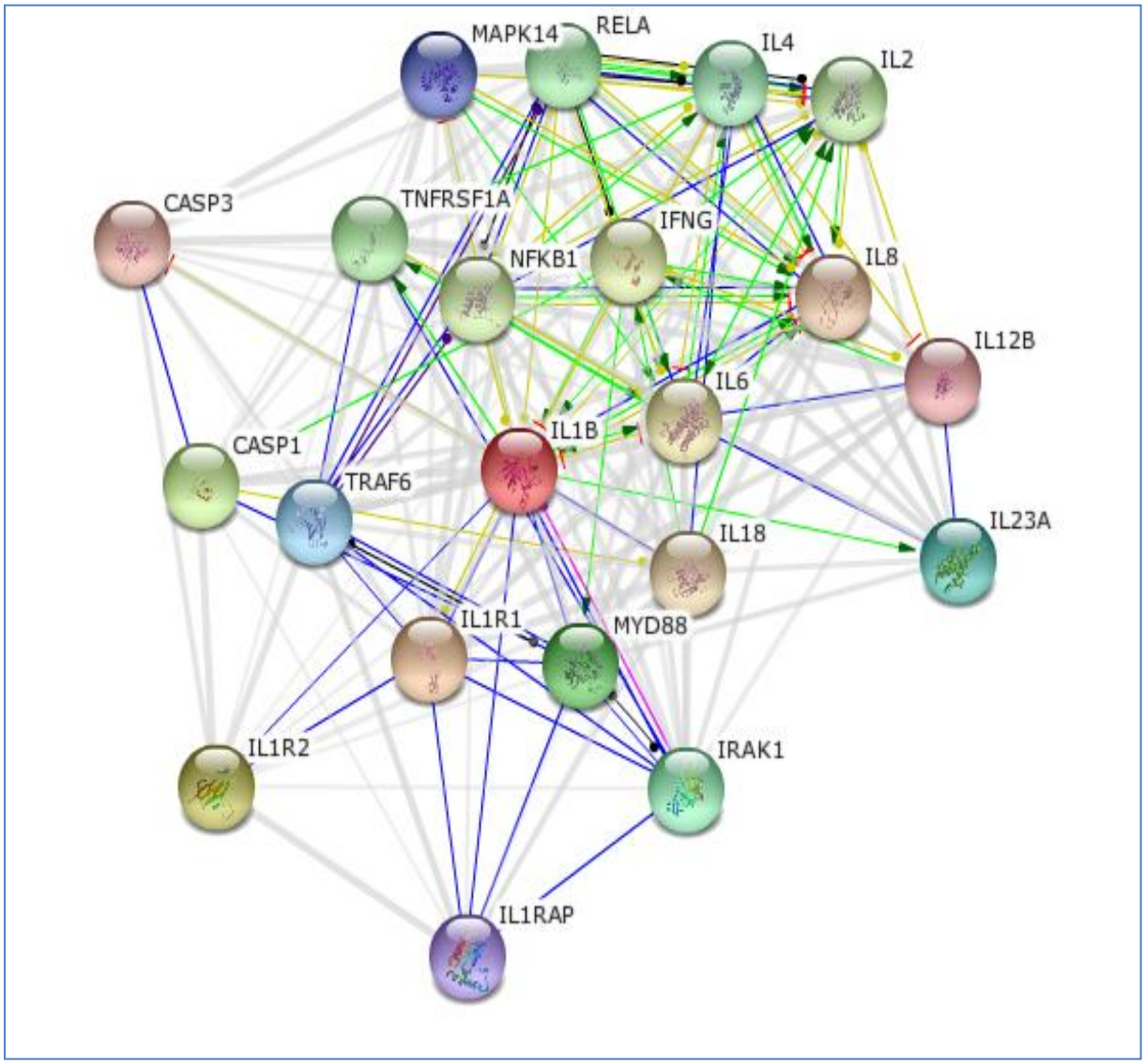

Figure 15 Important proteins for IL-1β from String database (20 neighbouring proteins)

The blue lines show binding. The green arrows show activation, and the yellow lines stand for expression [30].

IL-1β can be processed to become an active cytokine not only by intracellular interactions but also with neutrophil protease proteinase-3 (**PRTN3**). Proteinase -3 is also known to contribute to IL-18 processing. Other proteases such as elastase, matrix metalloprotease 9 (**MM9**) and granzyme A process the IL-1β precursor in the extracellular region. In addition, a mast cell chymase generates active IL-1β [2]. While granzyme A was reported as an extracellular converting enzyme, another study showed that both granzyme A and B act into intracellular space [33, 34].

A2M is a major plasma proteinase inhibitor which binds all proteinases. Also it is found that Alpha-2-macroglobulin precursor (**A2M**) binds through free thiol groups in A2M to IL-1β and inhibits its reactions in the extracellular region [35]. It was not clear that A2M inhibits either active IL-1 β or pIL-1 β.

Matrix metalloproteinases (**MMPs**) degrade extracellular matrix macromolecules and play important roles in many biological processes. These processes can be listed as morphogenesis, ovulation, embryo implantation, cell migration, tissue involution, angiogenesis, and wound healing.

Fewer MMP activities have been detected in steady state tissues but a synthesis of many of them has been known to be regulated with the cytokines, hormones and growth factors. Degradation of IL-1β by **MMP-1** (interstitial collagenase), **MMP-2** (gelatinase A), **MMP-3** (stromelysin 1), and **MMP-9** (gelatinase B) was observed in the extracellular space in case of inflammation. This degradation was effectively blocked by issue inhibitor of metalloproteinases **(TIMP)-1.** MMP treated IL-1β lost its ability to enhance the synthesis of prostaglandin E2 and pro-MMP-3 in human fibroblasts. The primary cleavage site of IL-1β by MMP-2 was identified at the **Glu25 - Leu26** bond while Caspase1-an intracellular enzyme-cleaves at **Asp116 - Ala117** [36, 37].

Another study showed that Streptococcal pyrogenic exotoxin B (SPE B), conserved extracellular cysteine protease expressed by the human pathogenic bacterium Streptococcus pyogenes, can cleave inactive human IL-1β precursor (pIL-1β) to produce biologically active IL-1β in extracellular space [38].

Also it is known that IL-1β induces IL-17, IL-17 in turn stimulates the production and release of IL-1β from primary human blood monocytes [2].

IL-18 (IFNγ inducing factor) shares similar inflammatory properties to IL-1β. IFNγ is known to potentiate IL-1β release from human cells [39].

Extracellular interactions of IL-1β mainly consist of the cleavage of IL-1β. As it was discussed in intracellular reactions, the main cleavage was mediated by the Caspase-1enzyme. Thus, either by Caspase-1 or Matrix metalloproteinases (**MMPs**) active IL-1β is released into the extracellular region. Our main goal is to inhibit active IL-1β, which makes the most important reaction with its receptors. Hence, it was understood that no side-reactions will occur during inhibition since it makes reactions only for activation in the extracellular region.

2.2. PEPTIDE DRUGS

Peptides are short amino acid sequences that consist of fewer than 50 amino acids or 5000Da molecular weight. In recent years there has been a great interest in the usage of peptides as drugs due to their advantages [1].

Human diseases including osteoporosis (calcitonin), diabetes (insulin), prostate cancer and endometriosis (gonadotropin-releasing hormone), acromegaly and ulcers (somatostatin) and hypothyroidism (thyrotropin-releasing hormone (TRH) are treated with peptide based drugs today [40].

Peptide drugs show higher activity, specificity and lower toxicity than chemical drugs. Their accumulation in tissues is less. They can be easily modified to avoid degradation and improve bioavailability. The disadvantages of peptide drugs such as short-life, lack in vivo stability due to the protease degradation and difficulty in passing membranes are the problems that are still need to be solved in the therapeutic usage of peptides. But there are promising strategies to overcome these limitations in peptide drug design [40].

There are also known peptide drugs designed on the IL-1 receptor which will be explained in the following section. In this study we designed an inhibitor to active IL-1β to control related diseases. Most of the inhibitors are designed for the receptor which affects all kinds of agonists and intracellular interactions. We targeted active IL-1β in the extracellular region which will increase the specificity and will not interfere the inner cell activities of IL-1 receptor. By doing this we increased the specificity and side reactions occurring from the deactivation of interleukin receptor are prevented [12, 24].

2.2.1. DESIGNED PEPTIDES ON THE INTERLEUKIN-1 RECEPTOR

2.2.1.1. PEPTIDE DESIGN ON IL1 RECEPTOR: R-Y-T-V-E-L-A [12]

A specific peptide was designed from accessory protein to inhibit activities of IL-1 receptor. Figure shows the regions for effective peptide derivation. The most effective one was found to be RYTVELA which was named as 101.10. It is indicated as dark blue in the Figure 16 part B.

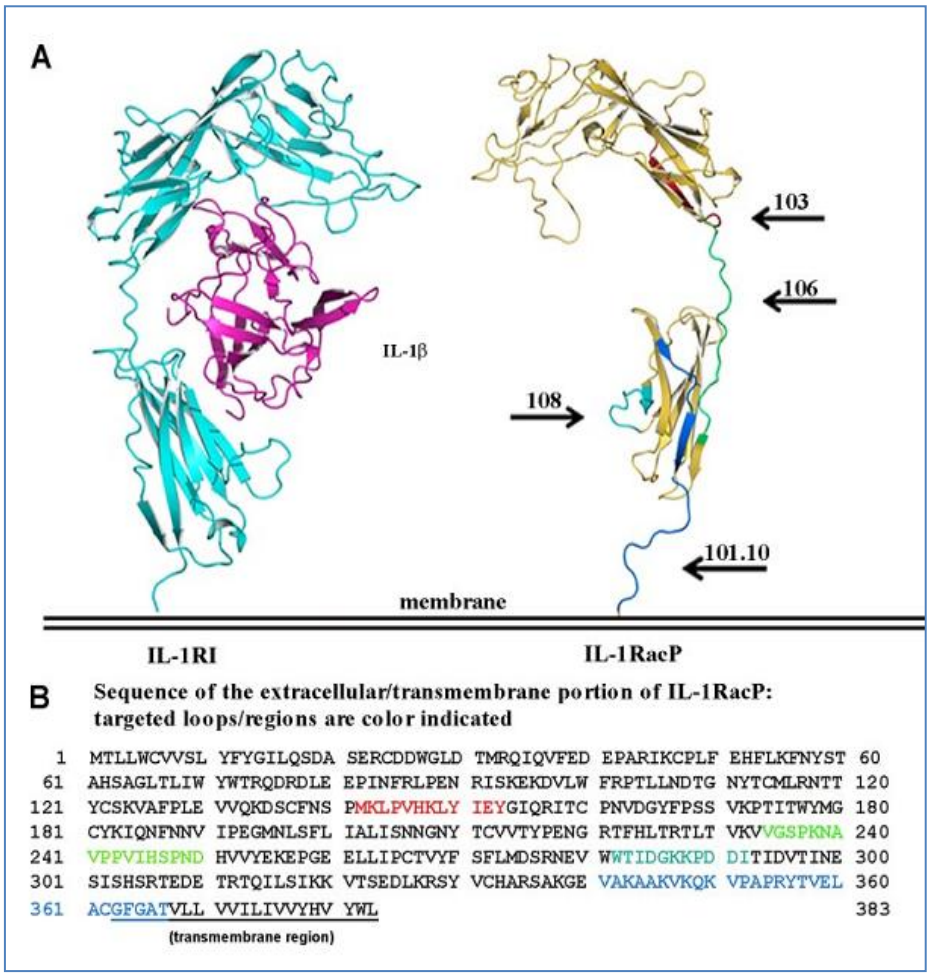

Figure 16 IL-1β with receptor and accessory protein.

A) Ribbon-like model of IL-1RI (PDB I.D:1ITB), IL-1, and IL-1RacP and identification of regions of derived effective peptides. B) Primary sequence of the IL-1RacP. Coloured sequences refer to corresponding loops indicated on A; blue, 101.10, turquoise: 108, green, 106, red, 103 [12].

2.2.1.2. ANTISENSE PEPTIDE DESIGN ON IL-1R 1 USING BORASHCI LOOP OF IL-1β

Antisense peptide design plays an important role in the inhibition of IL-1β. A sense peptide is one whose sequence is coded or by the nucleotide sequence of the sense strand of DNA (read 3'->5'). Conversely, the antisense or complementary peptide is coded for the nucleotide sequence (read 5'->3') of the complementary strand of DNA [24, 41, 42].

Several studies had been made to inhibit IL-1RI binding by using Borashci loop of human IL-1β.

BORASHCI LOOP:A nine amino acid-long peptide (VQGEESNDK), corresponding the exposed b-bulge in position 163–171 (47–55 mature sequence between the fourth and the fifth b-strand of the IL-1β structure (Boraschi loop), can mimic the immunostimulatory effects of IL-1β in vivo being devoid of inflammatory effects [24, 41, 42].

A seven amino acid residue peptides V-I-T-F-F-S-L and L-I-T-V-L-N-I were designed as inhibitors of IL-1R1 in vitro [41, 42]. Also in-vivo experiments for F-V-I-T-F-F-S-L-Y resulted efficiently for inhibition of IL-1R 1 [24].

CHAPTER 3

3. METHODS

3.1. MODELS OF PROTEIN

For all computational methods IL-1β with PDB I.D:1IOB, IL-1β bound to receptor with PDB I.D:1ITB and IL-1α with PDB ID: 2KKI were used. The following figures indicate IL-1β and IL-1α proteins. The structures of IL-1β and IL-1α are closely related and structural differences discussed in Chapter 2.

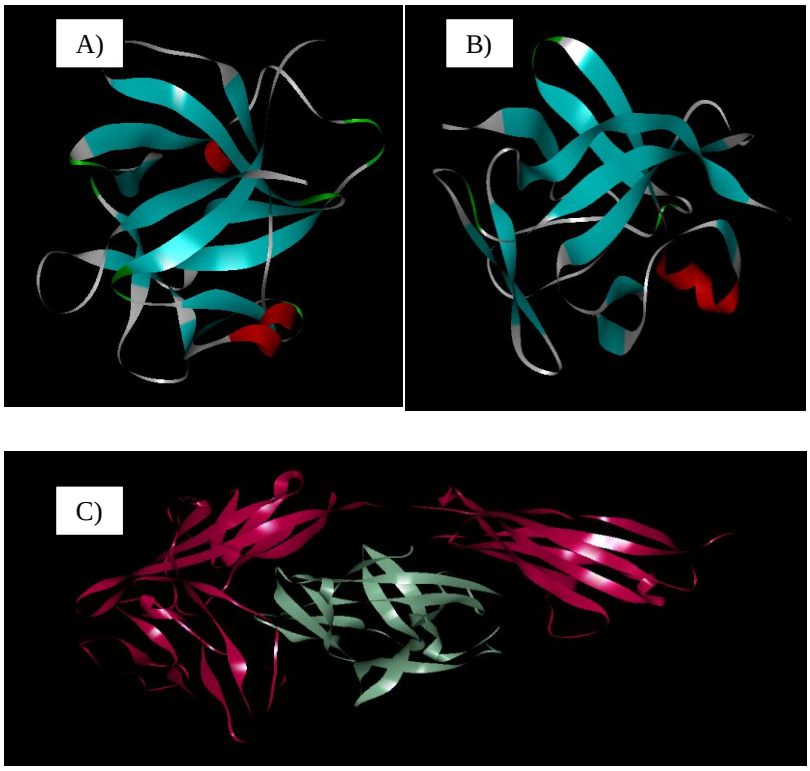

Figure 17 Snapshots from IL-1β and IL-1α.

A) IL-1β with PDB I.D:1IOB B) IL-1β bound to receptor with PDB I.D:1ITB C) IL-1α with PDB ID: 2KKI

3.2. MOLECULAR DOCKING

For molecular docking the most cited programs were selected among commonly used docking programs. Initial docking studies are carried out with Autodock 4.0 [43] since it is suitable by using Python Programming Language 2.7. Rechecking of selected ligands were carried out with GOLD (Genetic Optimization for Ligand Docking) 4.1 program from Cambridge Crystallographic Data Centre, UK [44].

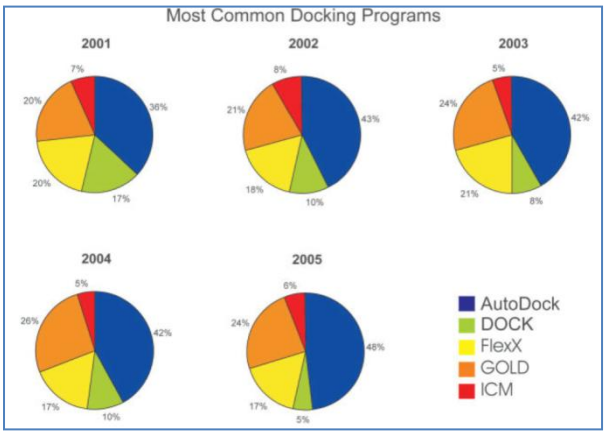

Figure 18 Most common docking programs [45].

3.2.1. AUTODOCK 4.2[43]

Autodock 4.2 docking program was used for preliminary docking studies within this study. It is the most cited docking program among all docking programs [45].

Autodock calculates the free energy of binding by subtracting unbound ligands' and proteins' energy separately from the bound state of ligand-protein complex.

The energies were calculated by evaluating the intramolecular energetics of the transition from the unbound state to the bound conformation for both protein and ligand separately, and then by evaluating the intermolecular energetics of bringing the protein and ligand together into the bound complex. The formula is given below for the binding free energy calculation:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad (1)$$

Here the first two terms indicate intramolecular energies of bound and unbound states of ligand while the following two terms indicate intramoleculer energies of protein for bound and unbound states. The change in intermolecular energy was calculated in the third parentheses where ΔSconf is the entropy loss upon binding. In unbound state, since two molecules are distant from each other, $V^{(P-L)}{}_{bound}$ was accepted as zero.

The pair-wise atomic terms (indicated as V) include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation given in the following formula. Here W indicates weighting constants determined from experiments [46].

Conformational loss in entropy is directly related to the number of rotatable bonds in the molecule (Ntors):

$$\Delta S_{conf} = W_{conf} N_{tors} \quad (2)$$

While docking, all torsion angles were set as flexible and Gasteiger charges were added by Autodock Tools (ADT). Polar hydrogens were added and unit atom approximation was applied. As a binding site, the most potential binding site was selected and taken as a grid centre. The maximum length of each amino acid, dipeptides, tripeptides and others were measured with DS program; which were used for grid-box size determination. Grid spacing between grid points was chosen as 0.4 Angstrom. The Lamarckian Genetic Algorithm was applied as the docking search parameter. For the Viterbi algorithm used in this study, to save computational time and effort, the lowest docking parameters were applied for ranking the ligands. The following settings for docking were applied; population size=150, runs = 50, maximum number of energy evaluations = 250 000; number of generations = 27 000.

After obtaining the last peptide sequences the parameters were updated to obtain more accurate results. The updated parameters are; population size=150, runs = 100; maximum number of energy evaluations = 2 500 000, number of generations = 27 000. Most probable conformation with the lowest binding free energy was selected as a result.

For genetic algorithm the updated parameters were used from the beginning to assess binding energies in detail.

Autodock also converts binding free energy to inhibition constant, Ki by using following equations:

$$\text{Equilibrium}: E + I \rightarrow EI \tag{3}$$

$$\text{Binding}: E + I \rightarrow EI \qquad K \text{ binding}, Kb \tag{4}$$

$$\text{Dissociation}: EI \rightarrow E + I \qquad K \text{ dissociation}, Kd \tag{5}$$

$$K \ binding = \frac{1}{K(dissociation)} \tag{6}$$

$$\ln K \text{ binding} = -\ln K \text{ dissociation} \tag{7}$$

$$\ln Kb = -\ln Kd \tag{8}$$

$$Ki = \text{dissociation constant of the enzyme} - \text{inhibitor complex} = Kd \tag{9}$$

$$Ki = \frac{E\ [I]}{[EI]} \tag{10}$$

$$\ln Kb = -\ln Ki \tag{11}$$

$$\Delta G \text{ binding} = -R * T * \ln Kb \tag{12}$$

$$\Delta G \text{ inhibiton} = -R * T * \ln Ki \tag{13}$$

Binding and inhibition occur in opposite direction, so the minus sign will drop.

$$\Delta G = R*T*\ln Ki\ not - R*T*\ln Ki \tag{14}$$

$$\frac{\Delta G}{R*T} = lnKi \tag{15}$$

$$\exp\left(\frac{\Delta G}{R*T}\right) = Ki \tag{16}$$

3.2.2. GOLD/HERMES [44]

Like Autodock, GOLD also uses a genetic algorithm for ranking flexible ligands by docking. The binding free energy of ligands was predicted with a GOLD score and a ChemScore. Total of 20 Å area was selected as the surface of binding with the determined hot points. All other parameters were kept at their default values.

Gold scoring function uses the following equation where $S_{hb\text{-}ext}$,S_{wdw_ext}, S_{hb_int}, S_{wdw_int} represents protein–ligand hydrogen-bond score, protein-ligand van der Waals score, contribution to the Fitness due to intramolecular hydrogen bonds in the ligand and contribution due to intramolecular strain in the ligand respectively.

$$GOLDfitness = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} \tag{17}$$

The Chemscore function uses the following formula to calculate binding free energy. S_{hbond}, S_{metal} and S_{lipo} indicate hydrogen-bonding, acceptor-metal, and lipophilic interactions, respectively. H_{rot} represents the loss of conformational energy upon binding. ΔG terms are the coefficients obtained from experimental data [47].

$$\Delta G_{binding} = \Delta G_o + \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + \Delta G_{rot} H_{rot} \tag{18}$$

3.2.3. HYPERCHEM [48] and DISCOVERY STUDIO VISUALIZER [49]

All .pdb and .mol2 files used in this study were prepared with Hyperchem by using script commands. The best binding conformation of peptides are analyzed in detail by Accelrys Discovery Studio.

3.2.4. VISUAL MOLECULAR DYNAMICS (VMD) [50]

For the regeneration of .pdb files during genetic algorithm implementation, VMD mutate command was used since it is applicable to the python scripts.

3.3. HOT POINT DETERMINATION

3.3.1. HOT POINT DETERMINATION BY GNM [6, 7, 51]

From the literature review, some hot points stand out with their higher affinity for binding. The Gaussion Network Model (GNM) was applied in order to confirm target residues. These residues gave greater picks in the GNM graph [6, 7, 51].

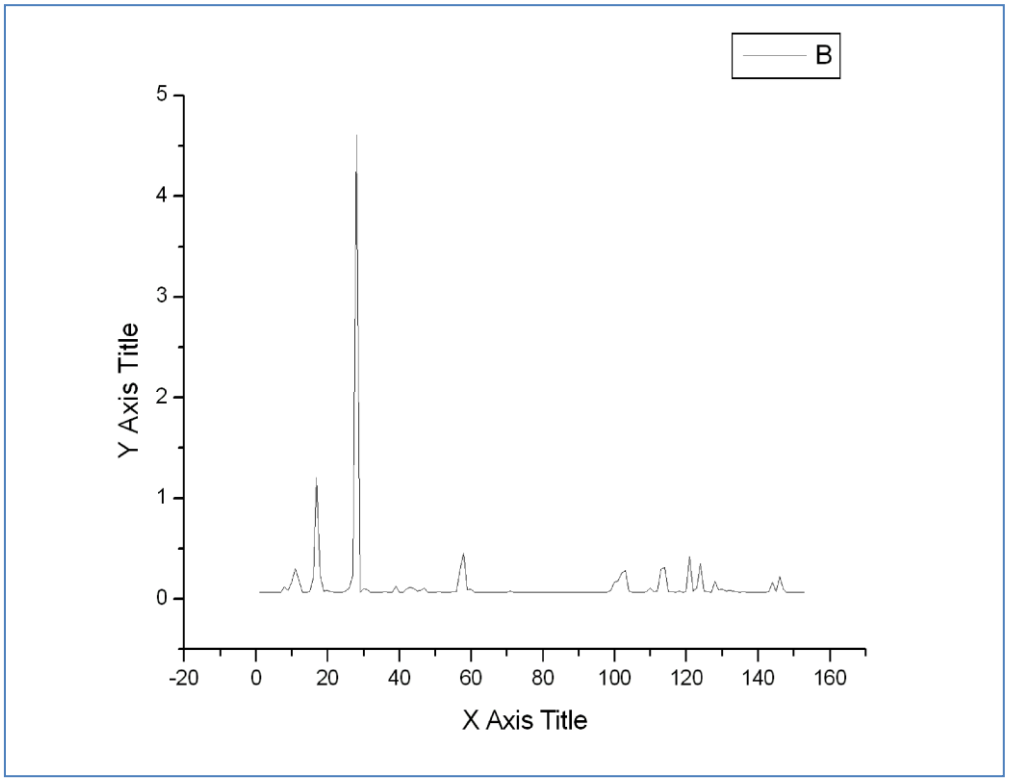

Figure 19 Hot point determination graph by GNM.

Higher picks shows a higher binding affinity [7, 51].

As can be seen from Figure 19, that there are four picks up to residue 40, which might correspond to residues **His 30, Arg 11, Arg 4** and **Gln 15** mentioned in the previous section. There are more residues with smaller picks in the later part of the graph which might correspond to **Lys 92, Asn 108, Glu 128.**

3.3.2. HOT POINT DETERMINATION BY HOTPOINT SERVER

A Hot point server was also used to determine important residues for binding [9]. Hot points for binding are indicated darker in Table 5. Chain A is IL-1β while Chain B is IL1-RI. As expected, residues 11, 15, 30, 32 and 128 gave higher results. Some new residues 6, 31, 46, 56 also gave higher affinity. Residue 56 also can be the corresponding pick seen around 60 in the GNM graph.

Table 5 Hot-point prediction results from HotPoint server [9]

Prediction Results						
Residue Number	**Residue Name**	**Chain**	**RelCompASA**	**RelMonomerASA**	**Potential**	**Prediction**
1	A	A	53.94	123.91	9.19	NH
2	P	A	65.56	90.46	1.24	NH
3	V	A	5.63	9.24	14.78	NH
4	R	A	8.46	62.72	17.31	NH
6	L	A	11.83	29.89	28.70	H
11	R	A	9.07	22.57	22.29	H
14	Q	A	14.95	65.02	16.28	NH
15	Q	A	12.49	90.78	26.18	H

25	E	A	28.25	50.69	8.40	NH
27	K	A	11.78	23.06	8.56	NH
30	H	A	3.26	50.09	25.96	H
31	L	A	3.07	20.79	32.58	H
32	Q	A	2.54	86.42	30.65	H
33	G	A	29.79	102.03	4.82	NH
34	Q	A	46.48	83.83	13.59	NH
35	D	A	8.57	34.35	13.55	NH
38	Q	A	50.91	80.84	6.85	NH
46	F	A	5.64	24.15	31.41	H
48	Q	A	26.59	57.84	24.82	NH
51	E	A	38.20	58.21	11.72	NH
53	N	A	45.60	97.66	5.05	NH
56	I	A	3.83	40.53	21.45	H
92	K	A	33.88	42.70	2.83	NH
93	K	A	3.50	39.54	14.42	NH
94	K	A	24.46	61.34	11.37	NH
105	E	A	24.51	53.94	12.27	NH
108	N	A	40.50	75.77	3.00	NH
127	A	A	21.70	39.16	4.60	NH
128	E	A	7.35	51.37	22.60	H
149	Q	A	34.64	59.89	15.49	NH
150	F	A	47.77	60.51	15.47	NH

152	S	A	75.76	96.87	6.69	NH
11	E	B	57.40	76.23	5.26	NH
12	K	B	60.46	87.07	5.03	NH
13	I	B	16.23	63.05	31.49	H
14	I	B	34.07	52.84	8.11	NH
15	L	B	32.02	35.02	17.68	NH
16	V	B	8.35	21.45	18.92	H
30	N	B	59.27	86.11	4.56	NH
31	P	B	90.96	104.62	2.53	NH
109	A	B	16.33	21.27	11.37	NH
110	I	B	14.41	30.17	7.49	NH
111	F	B	7.85	31.97	28.25	H
112	K	B	31.89	66.80	12.22	NH
113	Q	B	1.64	26.10	25.88	H
114	K	B	21.95	57.64	17.35	NH
115	L	B	12.01	15.44	32.78	H
122	G	B	5.13	20.37	15.79	NH
124	V	B	0.63	35.47	29.30	H
126	P	B	9.67	22.13	12.72	NH
127	Y	B	19.77	47.80	28.54	H
129	E	B	32.27	68.16	9.30	NH
163	R	B	28.64	42.16	16.03	NH
201	L	B	20.97	46.50	17.36	NH

204	N	B	62.51	92.64	7.42	NH
237	L	B	19.58	39.20	32.12	H
238	S	B	37.63	83.03	10.04	NH
239	D	B	16.35	17.51	9.40	NH
240	I	B	9.95	58.38	23.60	H
249	V	B	35.60	70.98	17.26	NH
250	I	B	17.60	26.86	17.04	NH
251	D	B	46.06	60.50	2.47	NH
252	E	B	39.00	101.02	5.66	NH
259	E	B	22.01	34.73	20.16	NH
260	D	B	35.01	48.76	7.26	NH
261	Y	B	6.67	66.78	15.26	NH
263	S	B	36.47	55.28	15.14	NH
265	E	B	54.95	69.22	12.32	NH
275	L	B	2.34	5.25	24.34	H
298	K	B	9.67	47.74	14.53	NH
300	T	B	30.64	95.00	7.59	NH
303	I	B	55.85	71.51	6.30	NH

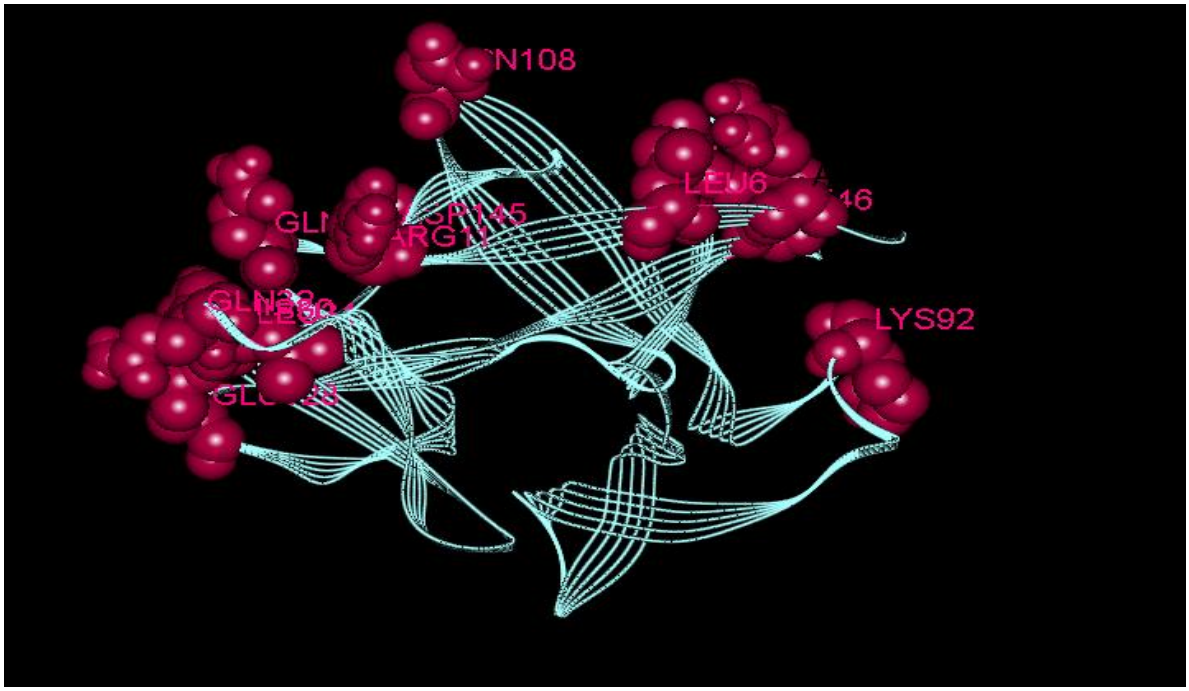

Figure 20 Important residues for binding of IL-1β.

Hot points are given in CPK representation (1IOB).

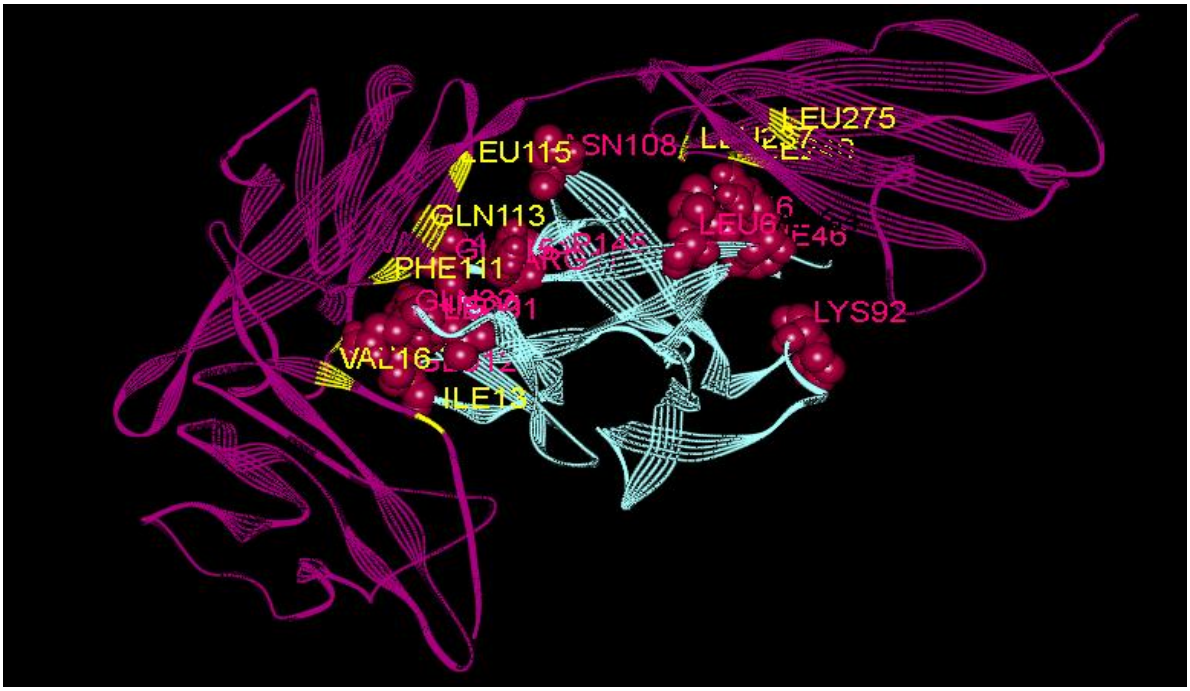

Figure 21 Bound conformation of IL-1β with IL1-RI.

The purple colour stands for the receptor with yellow hot points. The light green stands for IL-1β with pink labelled hot points. (.pdb I.D:1ITB).

Figure 20 and Figure 21 show hot points for IL-1β and bound conformation of IL-1β with its receptor respectively. These points were taken from the literature review, GNM and also the Hotpoint server.

3.4. GENETIC ALGORITHM

Identification of a specific peptide sequence, as inhibitor, to target protein is a complex problem. For even a tripeptide sequence there are 20^3 possibilities with 20 natural amino acids known. Trying all these possibilities will be time consuming. In order to eliminate these possibilities a computational time-saver method was applied. A Genetic algorithm (GA) is a commonly used successful method in determination of a specific peptide sequence in peptide drug design [3, 4]. Unal et al. [3] applied this algorithm to NF-Kb where a potential heptapeptide inhibitor was designed with successful binding free energy. Kamphausen et al. [4] implemented GA for RNA folding and Thrombin inhibitor to peptide design. The efficiency of the algorithm also can be measured by the number of cycles or the time consumed during evolution. Kamphausen et al. designed an inhibitor in five cycles, which is an outstanding result for GA. In another implementation of GA, Budin et. al. found conserved motifs to caspase family [52].

By referencing those examples, firstly the Genetic algorithm was implemented to IL-1β to design a tripeptide sequence in this study.

The Genetic Algorithm (GA) is a heuristic method that mimics the nature and biological evolution. The GA was first proposed by Holland in the 1970's. It basically follows Darwin's principles for natural selection where solutions evolve under a definite selection criteria. The GA algorithm starts with set of solutions (chromosomes) which form a population. The solutions are ranked according to some fitness function determined. Best solutions from this set are taken to a new population, which makes the new population better. To create the new population the GA uses mechanisms such as mutation and crossover. This is repeated until the solution converges to one solution in each generation [4, 5].

Elitism

Elitism means showing precedence to top scored population members.[4] In our study peptides with the lowest energy calculated via Autodock will have priority.

Mutation

In genetic algorithms the mutation operator is generally used to implement genetic diversity into a population. By generating variety this important operator prevents stagnating the GA solutions in the local minima. But if the rate of mutation is too high then the solution may diverge instead of converging [4].

Crossover

Crossover is another operator for increasing variety. Two parents need to be selected to implement this operator. From parents, new offsprings will be produced by exchanging their genes. If the crossover rate is too high, then the GA solutions may converge too soon, which results in missing better solutions [4]. An example of crossover and mutation can be seen in Figure 22.

In our study, the peptide sequences represent chromosomes while 20 natural amino acids represent genes. Each chromosome consists of 3 genes since the aim is to design a tripeptide sequence to target protein. The input of the algorithm is the first population which comprises randomly selected three peptide sequences in .pdb format. Our fitness function is the docking results which give binding free energy of each peptide sequence. The hot spot of the target protein for the GA was selected as HISTIDINE 30 residue which gave the highest peak in the GNM graph among known hot points from the literature. Only the mutation operator was implemented in this study. Since tripeptide is a small molecule, crossover will not show any important contribution. Mutation of the tripeptides was achieved by single residue change as can be seen in the following Figure 22.

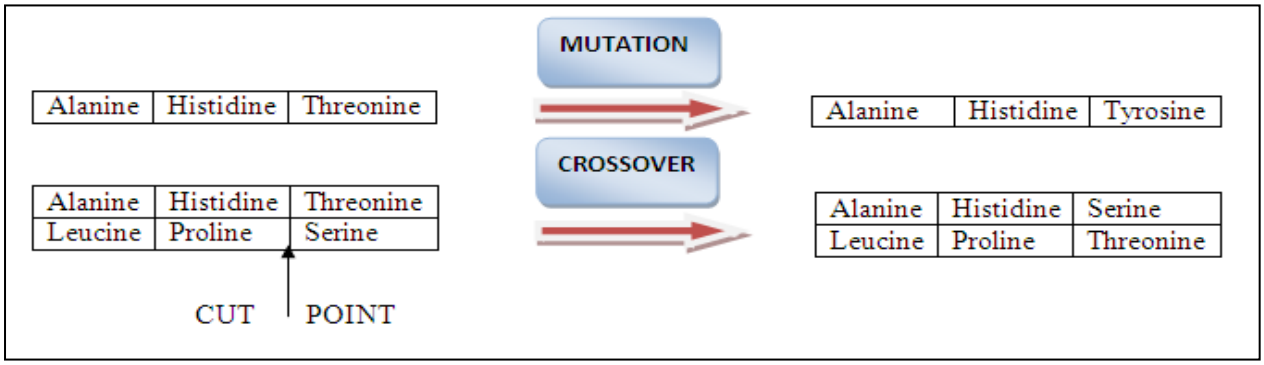

Figure 22 Simple examples of Mutation and Crossover via amino acids.

The following steps are implemented in the GA:

1) To the targeted hot spot, randomly selected first generation with 50 tripeptides docked via Autodock. The parameters selected for Autodock are given in the previous section. First generation was ranked according to their binding affinities from docking results.

2) Five parents with the highest binding affinity directly passed to the next generation by Elitism.[4]

3) All remaining members are mutated by single residue mutation. The new generation was formed by using only mutation to increase variety.

4) The new generation docked and ranked according to their binding free energies and again the steps 1 to 4 were implemented.

5) After ranking a sufficient number of generations, our peptide sequences started to converge to the same solution.

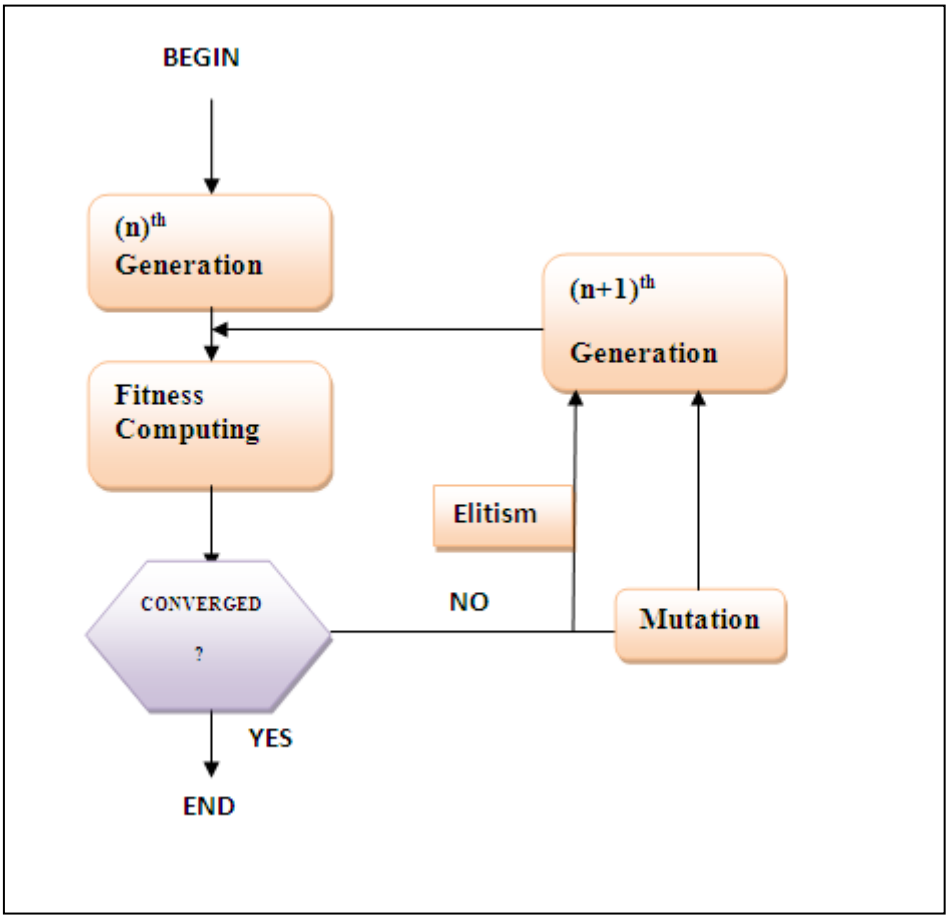

Figure 23 Schematic diagram of Genetic Algorithm system used in this study.

3.5. VITERBI ALGORITHM

In order to design a potential inhibitor to IL-1β, another algorithm was implemented called the Viterbi Algorithm. Like the Genetic Algorithm, this algorithm can also be applied to any protein with any peptide sequence length. This algorithm generates peptide inhibitors by docking its residues pair by pair to a selected path along the protein surface. The algorithm is based on the Hidden Markov Model (HMM). Successful implementations of this model on biological data are well known. Unal et al. implemented the HMM in a de-novo peptide design where successful inhibitors for NF-κB, HIV-1 and Spg 40 proteins were designed. [3, 16] The HMM was applied by Karplus et al. to obtain remote homologs of protein sequences.[20] Additionally, the HMM finds important position in the analysis of protein structures [53], RNA and DNA sequences [18], multiple sequence alignments in comparative gene finding, detection of conserved elements [54] and in de-novo peptide design. [3, 16] Also Kaboyashi et al. described a computational tool based on HMM in order to screen binding peptides to MHC class II proteins [21] while Krogh et al. developed a web-server for predicting transmembrane protein topology [55].

3.5.1. MARKOV CHAIN

The Markov chain is an important type of chance process where the outcome of any process affects the next process in a chainlike manner. A. Markov developed the model in 1907, and the model has found application in a wide range of areas from music to chemistry in problem solving [15].

3.5.1.1. DEFINITION

We have a set of states represented as S= [*s1, s2, s3*…..]. The Markov model goes from one state to another over defined probabilities. Consider a process that starts from state s_i and then moves to state s_j with a probability of p_{ij}. The probability p_{ij} is called the transition probability and does not depend on the previous steps that the process is in.

A generalized equation shows the relation between number of states and transition probabilities.

$$p_{ij} = \sum_{k=1}^{r} p_{ik} p_{kj} \tag{19}$$

3.5.1.2. THE HIDDEN MARKOV MODEL

Hidden Markov Model is a specific type of Markov Chain in which the system assumed is to have unobserved (hidden) states. The state is not directly visible to the observer but visible according to output depending on the states.

Example: URN and BALL model

Assume there are N large glass urns full of M distinct colourful balls in a divided room with a curtain. According to some random process one can choose an initial urn without telling the second person which urn was selected. From that urn a ball is chosen at random and recorded as an observation. Then the ball is replaced to the urn selected. A new urn will be selected according to the random selection process associated with the current urn and the ball selection will be repeated. The entire process will be the finite observations of colours where the selection of the urn is a state. In each state the colour probability is defined. The choice of urns will be represented as the state transition matrix. Here second person only sees the ball colours in sequence without any knowledge of chosen urn sequences. Since the choice of the urn does not directly depend on further or previous urns, this model is called a Markov process. This simple example is also explained with the Figure 24 where O and P represent the observations and probabilities respectively [56].

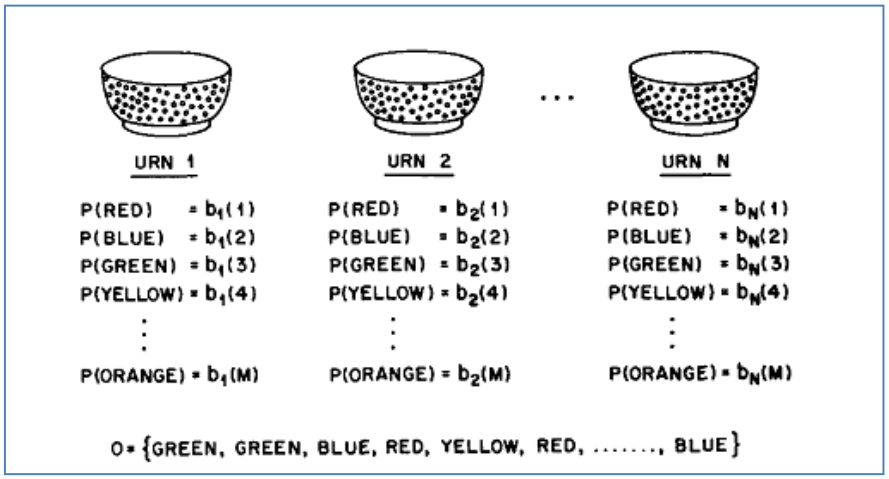

Figure 24 Urn and Ball model example for the Hidden Markov Models [56].

3.5.2. VITERBI ALGORITHM

The Viterbi algorithm (VA) was first proposed by A.Viterbi as a method of decoding convolutional codes. Speech recognition, computational linguistics, satellite systems and bioinformatics are the most common application areas of this algorithm [57, 58].

The Viterbi algorithm is a Hidden Markov Model based algorithm which finds the most likely solution from the sequence of hidden states. These hidden states form the most probable path called the *Viterbi Path*. It differs from the HMM by having forward and backward algorithms.

Important applications of this algorithm in bioinformatics are known. Krogh et al. used the VA in the alignment of sequences to model protein families and domains [17]. Unal et al. used the VA in de-novo peptide inhibitor design to various proteins [16]. Moreover, Brejova et al improved the VA by implementing a new labelling system to the Viterbi paths [59]. Fischer at al. developed a new sequence predicting model based on the VA where the observable outputs are the mass peaks in mass spectroscopy and the hidden states are the amino acid sequences [60]. Mirabeau et al. designed a computational bioinformatics tool to identify novel biologically active peptides in the human proteome by using the VA [19].

In line with these references the Viterbi Algorithm was applied to design a seven amino acid long inhibitor to IL-1β in our study. The algorithm was written via Python Programming Language and applicable to any protein with any peptide length. The algorithm generates a peptide sequence as a solution of the problem where the torsion angles were assumed to be observable variables of the problem. The first step was the determination of the Viterbi path

which mainly consists of hot points along the surface of the protein. Then the algorithm tries all possible sequences by docking dipeptides pair by pair along the path. After determining probabilities of the most possible pairs the Markov based partition function using the Ising model matrix multiplication scheme was formed [61]. The transition probabilities based on this partition function were evaluated. Observable variables were formed by ψ and φ angles obtained from the Ramachandran potentials [62]. Thus the conformational suitability of the designed peptide was satisfied. The Ramachandran torsion angles were assumed to obey the Markov model by which a given angle depends only to the preceding torsion angle. A schematic explanation of the algorithm is represented in Figure 25. In our model hidden states (S1, S2 …) are the dipeptides while our observable states (O1, O2 ...) are the secondary structures. Transition probabilities (a12, a23 ...) are the probabilities that an amino acid occupying the $t+1^{st}$ position with the knowledge of the amino acid at t^{th} position. These probabilities were obtained from binding free energies by docking. Output probabilities (b12, b23, b34 ...) are the angle preference probabilities of amino acid types given the preceding angle. Emission probabilities were obtained from selected data banks.

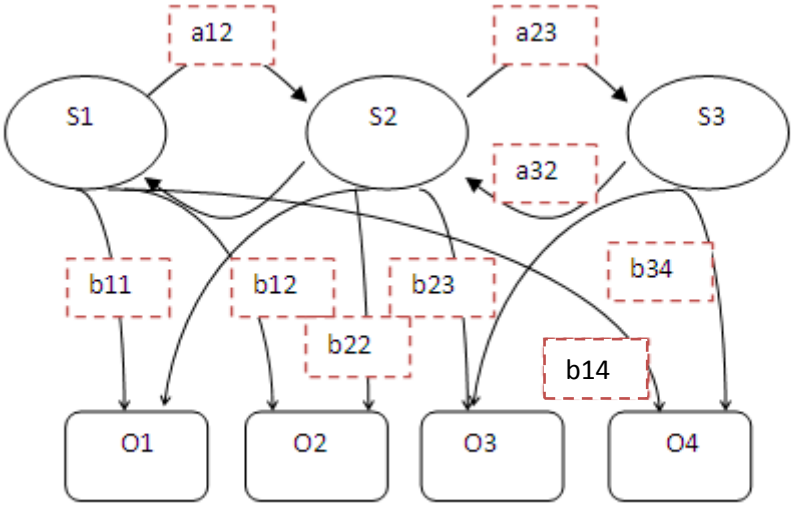

Figure 25 Schematic representation of the algorithm.

Here S1, S2, S3 represents the states, O1, O2, O3, O4 represents observable variables. Transition probabilities are represented with small a, while emission probabilities are represented with b.

3.5.2.1. DETERMINATION OF THE VITERBI PATH & DOCKING

By referencing previously determined hot-points four different paths along the receptor binding sites were determined. For a seven residue long peptide, seven neighbouring hot points were indicated and seven grid boxes were constructed. All selected paths and residues can be seen from Table 6 and Figure 27. Mid points of two close chiral carbon coordinates were given as the grid centre of the grid box. To provide enough flexibility to the path, grid boxes were chosen such that all dipeptides can rotate with freedom in the box. This freedom gave enough flexibility to obtain the most probable conformation within the selected neighbourhood of hot points. The first grid contains the first amino acid, the first and second grids contain the first and second amino acids. The t_{th} and $t+1_{st}$ grids contain the t_{th} and $t+1_{st}$ residues.

For each centre, 400 possible dipeptides were docked via Autodock except the first centre. To the first centre 20 amino acids were docked and all binding energies were evaluated from the most probable conformation. Python scripts were used to automate the docking process. The Autodock's output of binding energy is in kcal/mol. To the first grid, the first chiral carbons' coordinates were given as the docking centre. For the rest, the mid point of chiral carbons discussed above, was selected to coincide with the centre of the dipeptide in docking process. This pair wise docking continued until the last residue. ACE caps were added to dipeptides' N termini for the calculation of torsion angles. A schematic example of docking grids is given in Figure 26.

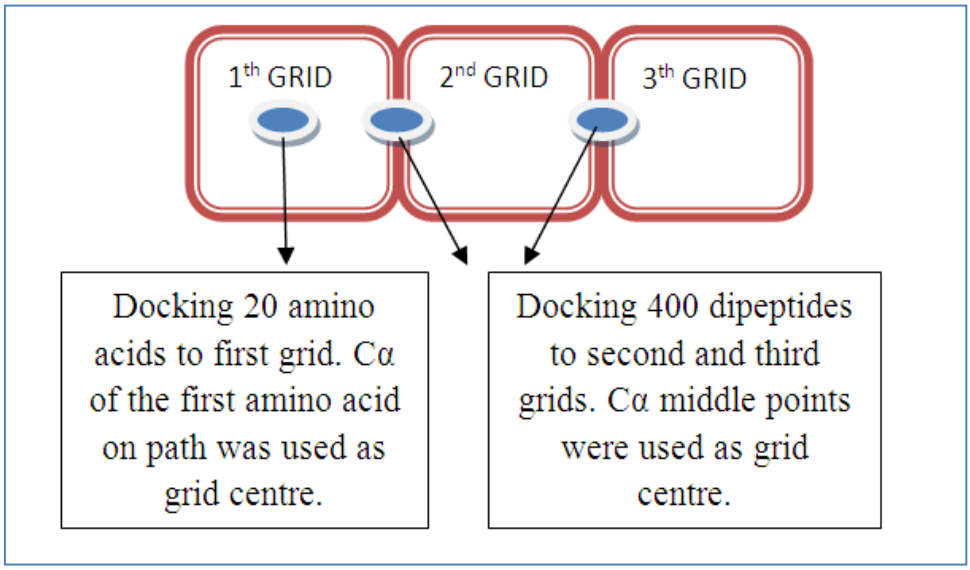

Figure 26 Schematic diagram of docking to grid boxes for a tripeptide.

Table 6 Selected Viterbi Paths for IL-1β

	1.GRID	2.GRID	3.GRID	4.GRID	5.GRID	6.GRID	7.GRID
1.PATH	ASN-108	GLU-105	LEU-110	THR-147	MET-148	GLN-149	PHE-150
2.PATH	GLN-32	LEU-31	HIS-30	GLU-128	GLN-15	ARG-11	ASP-145
3.PATH	LEU-6	ARG-4	PHE-46	LYS-55	ILE-56	PRO-57	LYS-103
4.PATH	LEU-6	ARG-4	PHE-46	LYS-92	LYS-93	LYS-94	MET-95

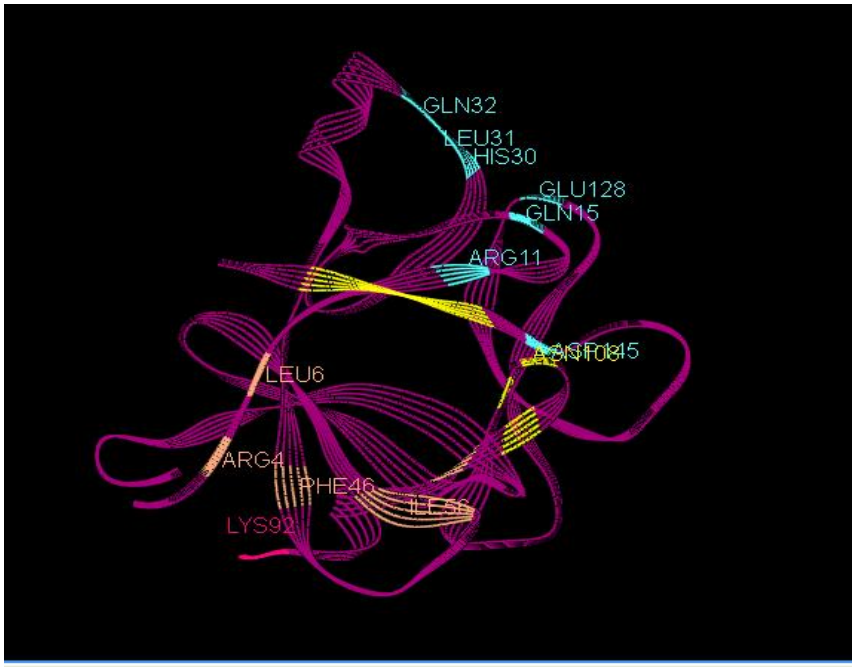

Figure 27 Selected Viterbi paths for IL-1β.

Selected paths are represented in Figure 27. The 1st path is represented in yellow and the 2nd path is in light blue. The camel stands for the 3rd path and the different parts of the 4th path from the 3rd are shown in pink.

3.5.2.2. CALCULATION OF TRANSITION PROBABILITIES

Twenty types of amino acids form states in our Markov model. Due to the features of the Markov model transition probabilities were determined. Here transition probabilities were formed by the probability of an amino acid occupying the t+1st position while the amino acid is at the tth position**.** Autodock binding energies were used as the statistical weights for determining the transition probabilities. The Rotational Isomeric States (RIS) approach was implemented in matrix multiplication scheme to calculate probabilities from the energies obtained [61, 63, 64].

For the first amino acid bound to the 1st grid box the statistical weight matrix (U1) was determined according to the following equation.

$$U_1 = \exp(\beta E_{1;i}) \tag{20}$$

Here β represents 1/ kT. k and T stands for the Boltzmann constant, Temperature respectively. The subscript i denotes any amino acid within 20. For the first grid 20 binding energies were available while for the rest 400 binding energies were evaluated. Temperature was chosen as 310K (body temperature).

For the dipeptide located between the t^{th} and $t+1^{st}$ residues along the peptide a statistical weight matrix was formed by using the following equation.

$$U_{t+1} = \exp(\beta E_{t,t+1;i,j}) \tag{21}$$

Here t represents the location of the grid box within n number of grids. i and j stand for the amino acid types among 20.

The partition function (Z) was obtained according to the following formula [61]:

$$Z = J^* \prod_{t=2}^{n} U_t J \tag{22}$$

Where J*=U1 and J= column [1 1 1… 1]. The J matrix has 20 members due to the 20 amino acids. First member was selected as Alanine, thus the J* matrix had to be changed from the original version to provide equal probability to each amino acid. The original J* matrix in the Flory notation was J*= [1 0 0..] which will locate the Alanine directly to the first grid.

The bound probabilities were calculated from the following equation where $p_{t,t+1;i,j}$ stands for the probability of having the residue i at the t^{th} position and j at the $t+1^{th}$ position.

$$p_{t,t+1;i,j} = \frac{J^* U_2 .... U_t U'_{t+1} ..... U_n J}{Z} \tag{23}$$

Here U'_{t+1} is a kind of matrix where all members are zero except the i, j^{th}. This formula was used to calculate all transition probabilities for each grid. For each grid, except the first, 20x20 matrices were obtained. For example, in the 1^{st} column and 1^{st} row the probability of Alanine-Alanine pair was recorded according to weight matrices. At the end, for a seven residue long peptide, six 20x20 matrices were obtained as transition probabilities. For the first grid 1x20 column matrix was obtained.

All multiplication within the equations includes dot product. Equation 23 was also used to calculate emission probabilities.

3.5.2.3. CALCULATION OF EMISSION PROBABILITIES

Emission probabilities are formed by the torsion angle probabilities of the peptides. A proper candidate should have convenient φ-ψ angles such that the torsions are energetically favourable during binding. The conformation of the peptide can be measured by two sets of probabilities:

1) Probabilities of torsion states determined by the angles ψ_t-φ_t of the residue at the t^{th} grid.
2) Probabilities for the torsion angles ψ_t-φ_{t+1} of the dipeptide formed by residues at the t^{th} and $t+1^{st}$ grids.

Firstly, each amino acid's torsion angles were measured and its torsion state was determined. Secondly dipeptide angles were measured and identified torsion states were converted to probabilities. Then these probabilities were combined in order to obtain most favourable conformation of the peptide. Torsion states were identified by the Ramachandran map shown in Figure 29 and Figure 30. An appropriate candidate should choose its conformation from the defined states. Each dipeptide was capped at its N-termninus by acetyl cap in order to obtain φ_t angle. Defined torsion angles are shown in Figure 28.

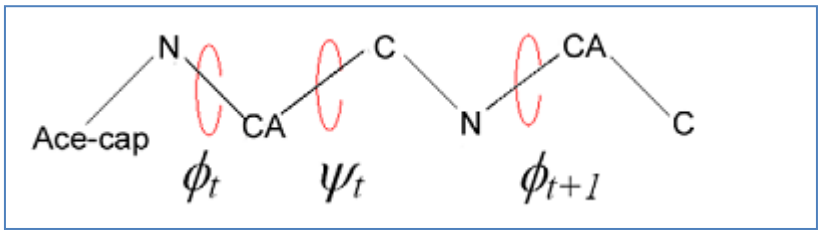

Figure 28 Schematic representation of φ-ψ angles [16].

Ramachandran plot was formed by the ψ-φ angle preferences of each amino acid. Due to the attractive and repulsive interactions between atoms the ψ-φ angle of a residue cannot have a random value. Depending on the residue type the torsion angles show preferences for different regions located on the Ramachandran map. There is a clear correlation between the residue type and the frequency of appearance of regions for adjoining members [16, 63, 64].

To obtain these regions, first a database was selected. Probability distributions of residues among the regions depend on the selected library. The coil library which includes the denatured conformations of peptides was downloaded from the website:

http://www.roselab.jhu.edu/coil/. The selection criteria of the library listed as: less than 20% sequence identity, better than 1.6 angstrom resolution, and a refinement factor of 0.25 or better.

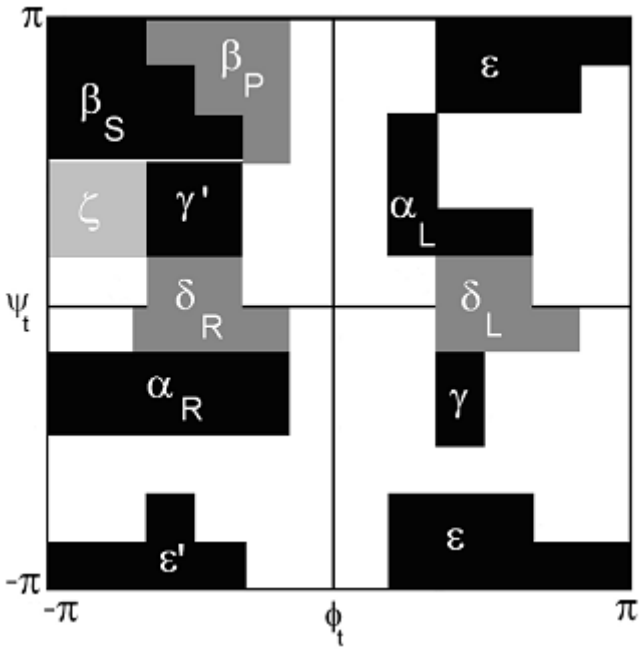

Figure 29 Ramachandran map with 11 states [16].

PDB Select Database of native proteins was used in order to draw the Ramachandran map in Figure 29. Unal et al also defined the given states for the Ramachandran plot. But in our study, we enlarged the state definitions as illustrated in Figure 30. In Figure 29, most of the plot was formed by the white areas which represent the zero state. Zero state means the peptide is outside of the defined region. While applying the VA, we realized that most of the potent inhibitors were in zero state. The idea of defining new states leads us to scan a database which mainly consists of the peptide-protein complexes. PDB Select database and PEP_X databases were selected. All peptide-protein complexes were examined and the emission probabilities were defined from those databases. Number of peptides and their percentages that fell into the zero state region were shown in Table 7, Table 8, Table 9 and Table 10 for amino acids and dipeptides. All probabilities were defined from PEP_X database and checked with PDB Select database. The new states were defined according to the occurrence of amino acid and dipeptide types. Figure 30 shows the new Ramachandran map for defined states.

Table 7 State distributions of amino acids from PDB Select Database within 793836 complexes

STATES	STATE 0	STATE 1	STATE 2	STATE 3	STATE 4	STATE 5	STATE 6	STATE 7	STATE 8	STATE 9	STATE 10	STATE 11
NUMBER OF PEPTIDES	134806	15568	82207	153975	15441	80383	50432	7837	16367	61734	62856	112230
PERCENTAGE OF PEPTIDES	%16	%2	%10	%20	%2	%10	%7	%1	%2	%8	%8	%14

Table 8 State distributions of dipeptides from PDB Select Database within 793836 complexes

STATES	STATE 0	STATE 1	STATE 2	STATE 3	STATE 4	STATE 5	STATE 6	STATE 7	STATE 8	STATE 9	STATE 10	STATE 11
NUMBER OF PEPTIDES	176699	22071	80959	146244	20239	69995	41597	9029	23838	52159	55055	95951
PERCENTAGE OF PEPTIDES	%22	%3	%11	%19	%3	%8	%5	%1	%3	%6	%6	%12

Table 9 State distributions of amino acids from PEP_X Database within 13627 complexes

STATES	STATE 0	STATE 1	STATE 2	STATE 3	STATE 4	STATE 5	STATE 6	STATE 7	STATE 8	STATE 9	STATE 10	STATE 11
NUMBER OF PEPTIDES	1783	185	1563	2659	185	1472	704	161	219	1016	1875	1805
PERCENTAGE OF PEPTIDES	%13	%1	%11	%21	%1	%10	%6	%1	%2	%8	%13	%13

Table 10 State distributions of dipeptides from PEP_X Database within 15055 complexes

STATES	STATE 0	STATE 1	STATE 2	STATE 3	STATE 4	STATE 5	STATE 6	STATE 7	STATE 8	STATE 9	STATE 10	STATE 11
NUMBER OF PEPTIDES	3394	395	1537	2585	357	1262	592	218	399	943	1712	1661
PERCENTAGE OF PEPTIDES	%23	%2	%11	%18	%2	%8	%3	%1	%2	%6	%12	%12

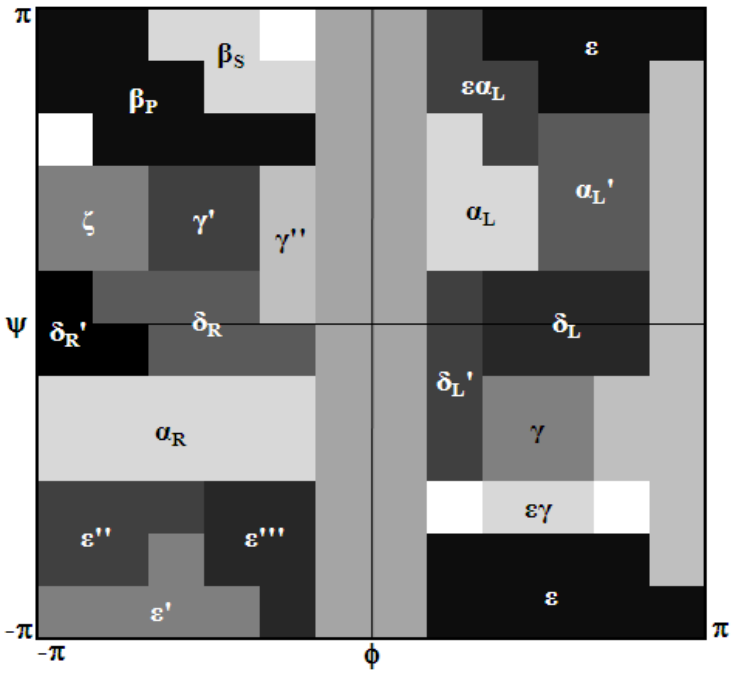

Figure 30 New states defined from the PDB Select Database.

As illustrated earlier there are two types of emission probabilities. The occurrence of a type i residue (among 20 amino acids) in torsion state m (among defined states) was calculated according to the following equation. Detailed explanation of states is represented in Table 11.

$$b_{i,m} = \frac{f(i,m)}{\sum_{m'} f(i,m')} \tag{24}$$

Here *f(i,m)* represents the normalized frequency of occurrence of ψ_t-φ_t in torsion-states m for type i residue. The singlet probability is denoted by $b_{i,m}$.

Second probability is $b_{i,j,m}$ where the preceding residue type i and the residue type j is in state m of ψ_t-φ_{t+1} space. For a dipeptide (with type i and j) the probability of $b_{i,j,m}$ is a column vector with 21 elements.

$$b_{i,j,m} = \frac{f(i,j,m)}{\sum_{m'} f(i,j,m')} \tag{25}$$

Product of two probabilities was used as emission matrices. Then equation 23 was applied in order to calculate emission probabilities.

3.5.3. IMPLEMENTATION OF THE VITERBI ALGORITHM

After defining emission probabilities from ψ-φ angle preferences and transition probabilities from binding free energies, we implemented VA in order to find most probable sequence.

While implementing the VA we followed the notation in References [16, 65]. The following definitions were used.

n: Number of residues of the peptide.

t: Grid number, $1 \leq t \leq n$

m: Index identifying the torsion state, $1 \leq m \leq 21$

$S=[S_1,S_2, \ldots ,S_{20}]$ The 20 natural amino acids set.

$A=[\ \alpha_1, \alpha_2, \ldots , \alpha_{21}\ ]$ The twenty one torsion angle regions.

q_t: The state of the t^{th} grid. For example $q_t=S_i$ means that the state of the t^{th} grid is the amino acid S_i.

α_t: The torsion-state of the amino acid at the t^{th} grid.

$p = (p_{i,j})$: The transition probability, $p_{ij}=Pr\ (q_{t+1}=S_j \mid q_t=S_i)$.

$b = (b_{i,j})$: The emission probability, $b_{i,j}=Pr\ (\alpha_{t+1}=A_j \mid \alpha_t=A_i)$.

$\pi= (\pi_i)$: The initial distribution vector, where $\pi_i=Pr\ (q_1=S_i)$.

The algorithm consists of forward and backward algorithms. Forward algorithm governs emission and transition probabilities while backward algorithm backtracks the find most probable sequences.

In the forward algorithm the algorithm will define **$\max Pr\ (q_1, q_2, \ldots, q_n \mid \alpha_1, \alpha_2, \ldots ,\alpha_n)$** by using emission and transition probabilities. This statement means that the algorithm finds the most preferable torsion angles in the protein binding site. For example **$Pr\ (q_1 \mid \alpha_1)$** states the probability of torsion state of an amino acid at the 1^{st} grid.

For amino acid type i and with position t:

$$\boldsymbol{\delta}_{t}\ \mathbf{i} = \max_{q1,q2....qn} \mathbf{Pr}\ \ \mathbf{q_1}, \mathbf{q_2}, ..., \mathbf{q_t} = \mathbf{S_i and}\ \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ... , \boldsymbol{\alpha}_t \quad (26)$$

$\boldsymbol{\delta}_{t}\ \mathbf{i}$ is the maximum probability of all paths with state S_i at grid t having torsion states $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ... , \boldsymbol{\alpha}_t$. Thus $\boldsymbol{\delta}_{1}\ \mathbf{i}$ can be defined as:

$$\boldsymbol{\delta}_{1}\ \mathbf{i} = \mathbf{Pr}\ \ \mathbf{q_1} = \mathbf{S_1 and}\ \boldsymbol{\alpha}_1 \quad (27)$$

Then, our aim is to determine maximum probability which means high affinity peptide with preferable torsion angles.

$$\max_{q1,q2....qn} \mathbf{Pr}\ \ \mathbf{q_1}, \mathbf{q_2}, ..., \mathbf{q_t and}\ \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ... , \boldsymbol{\alpha}_t \quad (28)$$

3.5.3.1. INITIALIZATION STEP

$$\boldsymbol{\delta}_{1}\ \mathbf{i} = \pi_i \mathrm{b}_i\ \boldsymbol{\alpha}_1\ , 1 \le i \le 20 \quad (29)$$

For the first grid box, a special condition for all twenty types of amino acids was applied. All emission (b) and initial (π) probabilities were chosen as unity to give an equal chance to each amino acid at the beginning.

3.5.3.2. INDUCTION STEP

$$\delta_{t+1}(j) = \max_{1 \le i \le 20} \delta_t(i) p_{ij} b_{ij}(a_{t+1} \mid a_t), 1 \le t \le n-1, 1 \le i, j \le 20 \quad (30)$$

δ_{t+1}, which is called the delta matrix, is a 1x20 array consists of the multiplication of binding affinities and torsion angle preference probabilities. For each grid box, transition and emission probabilities were multiplied row by row and the maximum of the results are kept in 1x20 arrays. For each loop, current delta matrix ($\delta_t(i)$) was recalculated to become new current matrix. At the beginning, first matrix of the transition probabilities was taken as the current matrix. This process is represented in Figure 31.

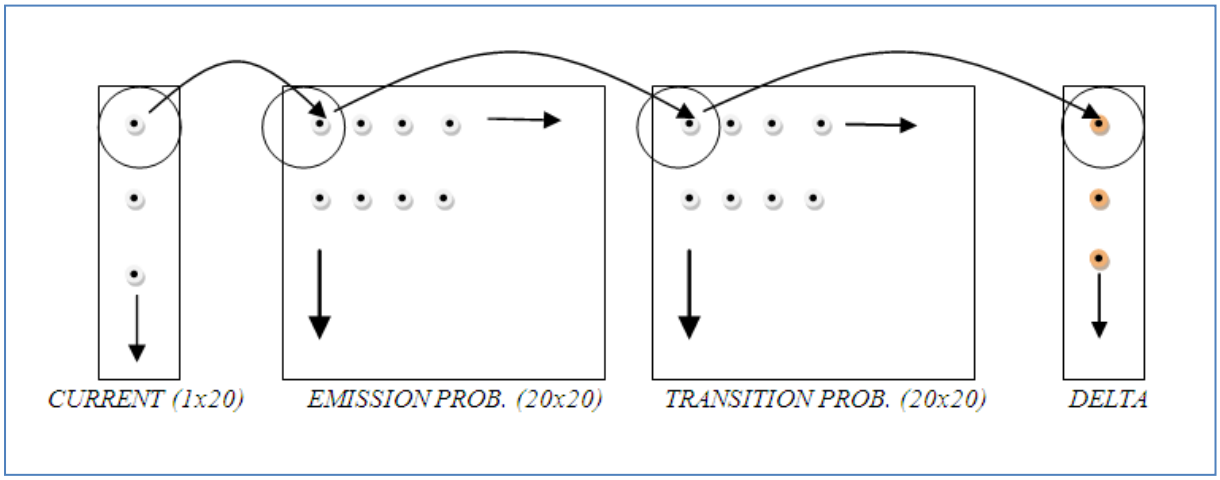

Figure 31 Calculation of delta matrices.

3.5.3.3. BACKTRACKING

Backtracking of delta matrices is the last part of the algorithm and it results in a high affinity peptide sequence to binding site. Initially the maximum of the last delta matrix was chosen. The index of the maximum represents the last amino acid of the sequence directly. We let

$$\Xi_n = \arg\max_{1\le i\le 20} \delta_n(i) \tag{31}$$

and choose $q_n = \delta_{\Xi n}$. Here q_n is the final state of the last residue. The remaining types were determined recursively via the following equation:

$$\Xi_t = \arg\max_{1\le i\le 20} \delta_t(i)\, p_{i\Xi_{t+1}} \tag{32}$$

Then putting $q_t = \delta_{\Xi t}$ we determined the remaining part of sequence. By choosing the maximum value in the last delta matrix, key value for the start was determined. The key value represents the column to be chosen in the last transition matrix. Each value in this column then multiplied with the matching residue in delta column. Again the maximum value of the multiplication was obtained and used as the next key value. At the end a new peptide sequence was obtained.

Table 11 Explanation for the defined states

STATE	NOTATION	DESCRIPTION	ANGLES
1	ε’	Mirror image of extended region ε	(-180<=phi<=-60 and -180<=psi<=-150) or (-120<=phi<=-90 and -150<=psi<=-120)
2	ε	The extended regions	(30<=phi<=180 and -180<=psi<=-150) or (30<=phi<=150 and -150<=psi<=-120) or (60<=phi<=180 and 150<=psi<=180) or (90<=phi<=150 and 120<=psi<=150)
3	α_R	Right handed alpha helix	(-180<=phi<=-30 and -90<=psi<=-30)
4	γ	Tight turn region	(60<=phi<=120 and -90<=psi<=-30)
5	δ_R	The right handed bridge region between two b-strands	(-120<=phi<=-30 and -30<=psi<=0) or (-150<=phi<=-60 and 0<=psi<=30)
6	δ_L	Mirror image of the δ_R region	(60<=phi<=150 and -30<=psi<=30)
7	ζ	Region observed mostly in residues preceding PRO	(-180<=phi<=-120 and 30<=psi<=90)
8	γ’	Inverse tight turn region	(-120<=phi<=-60 and 30<=psi<=90)
9	α_L	Mirror image of α_R	(30<=phi<=90 and 30<=psi<=90) or (30<=phi<=60 and 90<=psi<=120)
10	β_S	Extended beta sheet forming region	(-90<=phi<=-30 and 120<=psi<=150) or (-120<=phi<=-60 and 150<=psi<=180)
11	β_P	Region with extended polyproline-like helices	(-150<=phi<=-30 and 90<=psi<=120) or (-180<=phi<=-90 and 120<=psi<=150) or (-180<=phi<=-120 and 150<=psi<=180)
12	δ_R ‘		(-180<=phi<=-120 and -30<=psi<=0) or (-180<=phi<=-150 and 0<=psi<=30)

13	ε’’		(-180<=phi<=-120 and -150<=psi<=-90) or (-120<=phi<=-90 and -120<=psi<=-90)
14	ε’’’		(-90<=phi<=-30 and -150<=psi<=-90) or (-60<=phi<=-30 and -180<=psi<=-150)
15	γ’’		(-60<=phi<=-30 and 0<=psi<=90)
16			(-30<=phi<=30 and -180<=psi<=180)
17	$\varepsilon\alpha_L$		(30<=phi<=60 and 120<=psi<=180) or (60<=phi<=90 and 90<=psi<=150)
18	α_L’		(90<=phi<=150 and 30<=psi<=120)
19			(180<=phi<=150 and -150<=psi<=150) or (120<=phi<=150 and -90<=psi<=-30)
20	δ_L’		(30<=phi<=60 and -90<=psi<=30)
21	Eγ		(20<=phi<=60 and -120<=psi<=-90)

3.6. MOLECULAR DYNAMICS

Molecular Dynamics Simulations are one of the most applied computational methods for understanding behaviour of biological molecules. MD simulations provide detailed information on the fluctuations and conformational changes of proteins, nucleic acids and give detailed information about structures and thermodynamic properties of biological molecules. Since the very first simulation of bovine pancreatic trypsin inhibitor; (58 residues and 450 atoms) in 1977, MD usage developed rapidly. Although its being alternative to experiments still a controversial issue, it is widely used in the refinement and understanding of the experimental data. Considerable increase in computer power enables the simulations of more realistic model these days. Parameters used to describe the model are adjusted in various force fields. For biological molecules the most popular force fields are CHARMM, GROMOS and AMBER force fields [22].

MD Simulations are generally used to optimize the structure for protein flexibility, to include solvent effects and for the refinement of docked complexes in drug design area. There are several studies stating the value of MD simulations after docking process [22, 66-68]. Since docking is generally preferred for the elimination of large datasets, remaining individuals should be tested with more accurate and time-consuming methods such as MD.

MD basically solves the Newton's equation of motion for each particle as shown in equation 33. In equation 33 m_i defines the mass of particle while the f_i states the forces acting on atoms. These forces can be calculated from the derivation of a potential energy (U_{tot}) as represented in equation 34. r^N s(r_1, r2, r3…r_N) is the position of each particle in three dimension. Here N represents the number of atoms in the system. U_{tot} is compromised of bonded and non-bonded interactions as given in Equation 35.

$$m_i \ddot{r}_i = f_i \tag{33}$$

$$f_i = -\frac{\partial}{\partial r_i} U \tag{34}$$

$$U_{tot} = U_{bond} + U_{angle} + U_{dihedrals} + U_{Coulomb} + U_{VdW} \tag{35}$$

Here, first three terms represent intramolecular interactions and given in detail within the following equations. K_b and K_θ are the constants for bonds and bond angles. V, φ, γ states a force constant, dihedral angle and the phase angle respectively.

$$U_{bond} = \sum_{bonds} K_b (b - b_{eq})^2 \tag{36}$$

$$U_{angle} = \sum_{angle} K_\theta (\theta - \theta_{eq})^2 \tag{37}$$

$$U_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2} 1 + \cos(n\varphi - \gamma) \tag{38}$$

Non-bonded interactions are given as:

$$v^{Coulomb}(r) = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r} \tag{39}$$

$$v^{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{40}$$

For the electrostatic changes Coulomb Potential is calculated via Equation 39. And for the Leonard Jones Potential Equation 40 is adjusted into force fields. Here σ, ε states the diameter and the dielectric constant. Q_1 and Q_2 are partial charges of atoms 1 and 2 [69, 70].

MD Simulations can be carried out in various conditions. NVT simulation, also known as canonical ensemble, the number of particles (N), the volume (V), of each system in the ensemble is kept constant. This is the appropriate choice when conformational searches of molecules are carried out in vacuum without periodic boundary conditions. Without periodic boundary conditions, volume, pressure, and density are not defined. Constant-pressure dynamics cannot be carried out.

At NPT simulations, also known as isothermal–isobaric ensemble, temperature (T), pressure (P) and number of particles (N) are kept constant. This is the ensemble of choice when the correct pressure, volume, and densities are important in the simulation. It can also be used during equilibration to achieve the desired temperature and pressure before changing to NVT simulation where data collection starts [71].

3.6.1. SIMULATION PARAMETERS

In our study, MD simulations were performed in explicit solvent (water) with NAMD 2.6 [70] CHARMM27 [72] force field. In order to have an appropriate space for pulling the ligand for SMD simulations, the water box was created such that there stays more water in pulling direction. Available topology files within force field were used and necessary coordinate files were created by VMD Tk console. Simulations were performed at 310 K temperature and 1 bar pressure. Ions were added to simulation box in order to neutralize excess charge. Langevin dynamics were used to control the system temperature and pressure. A time step of 1fs was used. Non-bonded and electrostatic forces were evaluated each time step. Rigid bonds were stated as "none" in order to keep flexibility. All energy, pressure and coordinate files were updated each 1000^{th} time step of simulation. Rest of the parameters needed for simulation were applied as recommended in Namd tutorial papers.

3.6.2. SIMULATION STEPS

In this study five stepped MD simulations were followed as recommended from tutorial papers.

3.6.2.1. MINIMIZATION BY NPT ENSEMBLE

Total of 20000 time step (0.02ns) minimization with NPT ensemble were done in order to search for a local minima, places in which the molecule is relaxed. Minimization search is done by systematically varying the atoms positions and calculating the energy in system. In all NPT simulations, protein molecule was fixed and only water molecules in the system were allowed to relax. Fixing the protein in NPT ensemble allows the water to respond to forces faster than with relaxed protein. Thus it saves computational effort.

3.6.2.2. EQUILIBRATION BY NPT ENSEMBLE

Equilibration involves molecular dynamics which solves Newton's law for each atom to dictate their trajectories. Total of 500000 (0.5ns) NPT equilibration was carried out. All files

from the first minimization were used. Periodic Boundary Conditions were calculated via Tk console and applied in all NPT simulations.

3.6.2.3. MINIMIZATION BY NVT ENSEMBLE

Total of 20000 time step (0.02ns) simulation was carried out without Periodic Boundary Conditions. The volume is kept constant throughout the run. Conformational searches of molecules are carried out in vacuum. Without periodic boundary conditions, volume, pressure, and density are not defined. Constant-pressure dynamics cannot be carried out.

3.6.2.4. EQUILIBRATION BY NVT ENSEMBLE

Total of 4000000 time step (4ns) simulation was carried out without Periodic Boundary Conditions. In this simulation, our main concern was to convergence of Root Mean Square Deviation (RMSD). RMSD is the numerical measure of the difference between two structures. In our simulation, RMSD was calculated against the structure of first time step configuration of the protein. The formula is defined below:

$$RMSD\alpha = \sqrt{\frac{\sum_{j=1}^{Nt}\sum_{\alpha=1}^{N\alpha}(r\vec{\alpha}(tj) - < r\vec{a} >)^2}{N\alpha}} \tag{41}$$

Here Nα is the number of atoms; Nt is the number of time steps for which atomic positions are compared. $r\vec{\alpha}(tj)$ is the position of atom α at time tj and <rα> is the average value of the position of atom α to which the positions $r\vec{\alpha}(tj)$ are being compared. It is defined as:

$$< r\vec{\alpha} >= \frac{1}{Nt}\sum_{j=1}^{Nt} r\vec{\alpha}(tj) \tag{42}$$

3.6.2.5. STEERED MOLECULAR DYNAMICS

Steered Molecular Dynamics is a widely used method to measure unbinding energy of ligand and to understand the mechanism of unfolding process. The basic idea behind the SMD is to pull the ligand by applying external force to one or more atoms in it. Ligand is forced to move in chosen direction which is called the unbinding path. This applied external force can be calculated from the following formula.

$$F = k(x_0 + vt - x) \tag{43}$$

In equation 41, k, v, x_o, x stand for the spring constant, velocity, initial position and current position respectively.

$$W = F \times v \times \Delta t \tag{44}$$

With the help of classical work equation (42) one can calculate the work done on the system. Here Δt represents the time step [70].

The general work concept of thermodynamics is closely related with the external parameters of the system. For a system in contact with its thermal reservoir when some external parameters are changed infinitely slowly from an initial point A to the final point B, then the work performed on the system is equal to the Helmholtz free energy- the energy difference between the initial and final configurations.

$$\mathrm{W} = \Delta \mathrm{F} \equiv \mathrm{F_B} - \mathrm{F_A} \tag{45}$$

By contrast when the parameters were are changed in a finite rate, then work will depend on the microscopic initial conditions of system and reservoir. From the second law of thermodynamics, average of work cannot be smaller than the difference of initial and final configuration. This is the case where system identified as irreversible process.

$$\Delta F \leq \bar{W} \tag{46}$$

An outstanding contribution to MD achieved by Jarzynski in 1997. By relating the free energy difference between two equilibrium states of a system to the average of the work over all irreversible for going from one state to the other, Jarzynski created the famous equation:

$$\exp\ -\beta \Delta \mathrm{F}\ =\ <\ \exp\ -\beta \mathrm{W}\ > \tag{47}$$

Jarzynski proposed this equation for computationally easier way to calculate free energies in molecular simulations [23, 73].

In this study, the constant velocity 10^{-5} Å/fs was used and spring constant was taken as 7kcal/molÅ^2. For each candidate SMD pulling direction was calculated via VMD Tk console.

Pulling atom was chosen as the most proper one closer to the outer surface of the protein while the fixed atom was chosen as the residue facing toward to the pulling atom. Pulling pathway was chosen as 20 Å since it is suitable for the size of the water box. Greater unbinding pathways may get too close to the mirror image in simulation box which will lead faulty results in SMD. Different snapshots from NVT simulation were taken as starting points of SMD. Jarzynski Equality was applied in order to measure the unbinding energies of ligands. Energy, pressure and coordinate file frequencies were decreased to taken from each 500^{th} time step in order to obtain more accurate results.

CHAPTER 4

4. RESULTS

4.1. GENETIC ALGORITHM RESULTS

Genetic algorithm was applied to IL-1β to design a tripeptide inhibitor. After 13 generations the same sequence (T-S-W) started to appear within the five top scored peptides. At the end of 20 generations Threonine-Serine-Tryptophan (T-S-W) converged. The top scored five peptides and their binding free energies in kcal/mol can be seen in Table 12. The difference in the energies for same sequence comes from the Autodock features, which calculates the same sequence's binding each time. We preferred to recalculate the binding energies to understand if the sequence binds with low energy each time. Autodock gives 50 different conformations for each calculation and the lowest ones can be seen in Table 12. Thus for only T-S-W 7*50 conformations were obtained after the 13th generation. Since each generation contains 50 tripeptides the most potent sequence with lowest binding energy can be obtained at the end of 1000 trials instead of trying all 8000 tripeptides. This algorithm might have converged earlier if the energies had been kept constant. Thus recalculating energies every time costs computational effort but increases accuracy.

Table 12 Genetic algorithm Results after 13th generation

NUMBER OF GENERATIONS	1	2	3	4	5
14	TSW	WLF	ITP	FWW	SFN
ENERGY(kcal/mol)	-7.25	-6.85	-6.36	-6.24	-5.77
15	TSW	WLF	FWW	AFW	VCY
ENERGY(kcal/mol)	-7.86	-6.9	-6.59	-6.37	-6.1
16	AFW	GID	CHW	TSW	WLF
ENERGY(kcal/mol)	-6.63	-6.52	-6.49	-6.43	-6.23
17	TSW	VCF	SMK	AFW	KHG

ENERGY(kcal/mol)	-7.84	-6.76	-6.61	-6.6	-6.57
18	TSW	KHG	TTW	SWK	RVY
ENERGY(kcal/mol)	-7.47	-6.93	-6.88	-6.7	-6.65
19	DGW	TSW	TTW	SFA	EWM
ENERGY(kcal/mol)	-7.58	-6.98	-6.59	-6.52	-6.49
20	TSW	FNF	TTW	DDH	QIW
ENERGY(kcal/mol)	-7.1	-6.66	-6.38	-6.32	-6.19

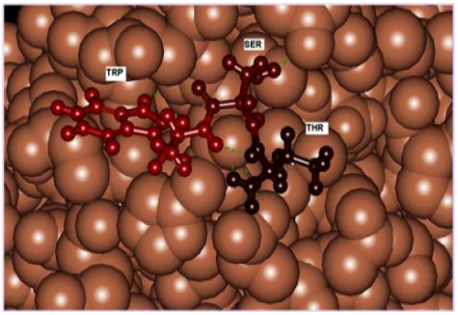

Figure 32 Docked conformation of IL-1β with TSW.

Docked conformation of T-S-W to IL-1β is shown with a CPK representation of the surface in Figure 32. After obtaining the most potent sequence four and five amino acid long peptides were designed since a tripeptide sequence might not be specific to the surface that is targeted. Secondly Il-1β is an extracellular protein, thus there is no need to consider molecular weight for peptide sequences. To design a tetrapeptide, the obtained amino acids were kept in the middle as fixed. 20 amino acids were docked to the beginning of the T-S-W sequence. From 20 results the three top scored ones were selected and can be seen in Table 13**.** Tyrosine – Threonine Serine - Tryptophan, Glycine – Threonine – Serine - Tryptophan and Aspartic acid –Threonine - Serine - Tryptophan are represented as **YTSW, GTSW, and DTSW** respectively**.**

Table 13 Tetrapeptide inhibitor design results

	X-T-S-W	
YTSW	GTSW	DTSW
-7.43 kcal/mol	-6.95kcal/mol	-6.84 kcal/mol

A total of 20 natural amino acids docked once to design pentapeptides to the end of the selected tetrapeptides. Most potent results can be seen in Table 14. Without docking step by step, all 400 combinations were docked at once to the selected tripeptide. Results of trying all 400 combinations are given in Table 15. Aspartic acid – Threonine – Serine – Tryptophan - Asparagine, Tyrosine – Threonine – Serine – Tryptophan - Histidine and Tyrosine -Threonine – Serine – Tryptophan - Phenylalanine are shown in one-letter representation as **DTSWN, YTSWH** and **YTSWF.**

Table 14 Pentapeptide inhibitor design results

	(D-Y-G)-T-S-W-X	
DTSWN	YTSWH	YTSWF
-7.97kcal/mol	-7.93 kcal/mol	-7.73 kcal/mol

Tryptophan - Threonine – Serine – Tryptophan - Arginine, Cysteine – Threonine –Serine - Tryptophan - Serine and Tryptophan – Threonine – Serine – Tryptophan - Serine are shown in one-letter representation as **WTSWR, CTSWS** and **WTSWS**.

Table 15 Pentapeptide inhibitor design results

	X-T-S-W-X	
WTSWR	CTSWS	WTSWS
-8.37 kcal/mol	-8.36 kcal/mol	-8.31 kcal/mol

According to the given results those containing tryptophan show a higher affinity to the target protein surface. The last pentapeptides were also controlled with the GOLD docking program. The Gold score and Chemscore fitness values can be seen in Table 16.

Table 16 Gold score and Chemscore results for pentapeptide inhibitors

	WTSWS	CTSWS	WTSWR
GOLD score	29.7	28.43	25.07
Chemscore DG (kcal/mol)	-6.7	-4.6	-10.5

According to both docking programs the pentapeptide with sequence W-T-S-W-R showed the most consistent results and was suggested as a potent inhibitor to IL-1β. All docked conformations are represented as CPK in the following figures.

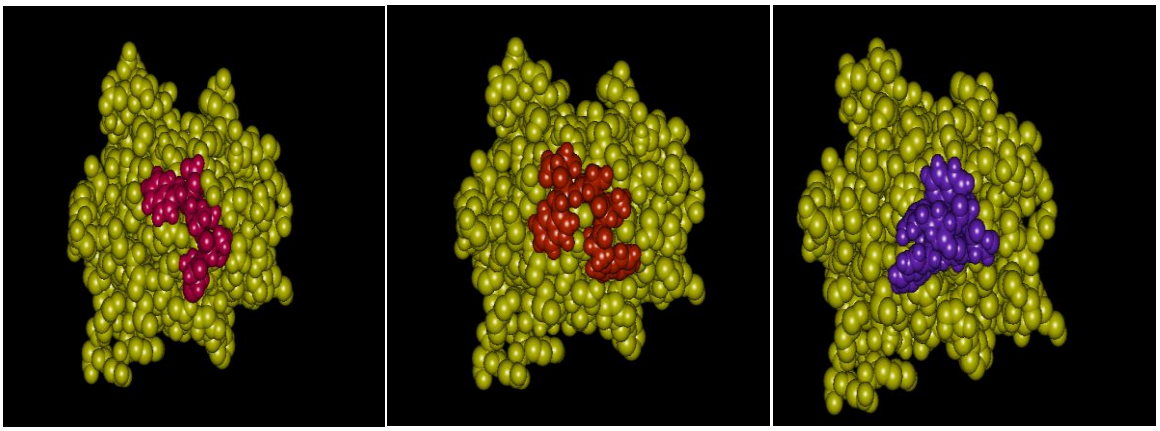

Figure 33 Docked conformation of IL-1β with predicted pentapeptides:

WTSWS is represented in orange, CTSWS in pink and WTSWR in purple.

As can be seen from Figure 33, all predicted pentapeptide inhibitors fit their binding cavity like a hand in glove. Besides having a great binding free energy, another important feature of a potential drug is the stability of the candidate in a cavity surface. Thus all three pentapeptides suggested as potent inhibitors to IL-1β.

4.2. VITERBI ALGORITHM RESULTS

4.2.1. TEST OF ALGORITHM

The algorithm was firstly tried on known complexes in order to verify its accuracy. Hot points of the complexes were identified from literature [8, 16, 74]. Total of seven different complexes were tried in order to design a tripeptide inhibitor. Table 17 summarizes the docking results in kcal/mol for both known inhibitor and designed one. Also the hot points which were used as the Viterbi path are represented.

The algorithm gives successful results in comparison with the known tripeptide inhibitors. In case of Scytalidocarboxyl peptidase B protein (1S2K) the algorithm gives a higher affinity tripeptide with similar properties. All amino acid types except Isoleucine (non-polar) in the known inhibitor are basic. Another similar result was obtained from SPG 40 protein where the known and designed inhibitor only differs in the middle residue. The difference can be juxtaposed in a logical result since both Proline and Phenylalanine are non-polar amino acids. In case of H.I.V protein the docking results were similar in terms of keeping the two charged and one non-polar residue. Complexes without known prior tripeptide inhibitor were also tested. The algorithm again gives rational results by means of binding free energy.

Table 17 Test Results of the Viterbi Algorithm

PROTEIN-PDB I.D	HOT POINTS	KNOWN TRIPEPTIDE INHIBITOR	VITERBI ALGORITHM RESULT
Interleukin 1-β (1ITB)	Chain A : Residues 29, 30, 31	-	Y-R-Y TYROSINE-ARGININE-TYROSINE -7.91 (kcal/mol)
Caspase 1 (2HBQ)	Chain A : Residues 237, 238, 285	-	Y-E-Y TYROSINE-ARGININE-TYROSINE -7.13 (kcal/mol)
H.I.V (1A30)	Chain A : Residues 29, 30, 31	E-D-L GLUTAMIC ACID-ASPARTIC ACID-LYSINE -7.53 (kcal/mol)	K-W-K LYSINE-TRYPTOPHAN-LYSINE -7.95 (kcal/mol)

SPG 40 (1ZBW)	Chain A : Residues 10, 14, 79	W-P-W TRYPTOPHAN-PROLINE-TRYPTOPHAN -8.7 (kcal/mol)	W-F-W TRYPTOPHAN-PHENYLALANINE-TRYPTOPHAN -8.84 (kcal/mol)
Proteinase K (1P7V)	Chain A : Residues 132, 133, 134	-	C-G-G CYSTEINE-GLYCINE-GLYCINE -5.89 (kcal/mol)
(1OGT)	Chain A : Residues 138, 139, 140	-	P-W-S PROLINE-TRYPTOPHAN-SERINE -6.28 (kcal/mol)
Scytalidocarboxyl peptidase B protein (1S2K)	Chain A : Residues 152, 155, 156	A-I-H ALANINE-ISELEUCINE-HISTIDINE -5.33 (kcal/mol)	K-R-R LYSINE-ARGININE-ARGININE -11.16 (kcal/mol)

4.2.2. RESULTS FOR INTERLEUKIN 1 BETA

Total of thirty heptapeptides were designed as potent inhibitors using defined Viterbi paths. All peptide sequences with their binding free energies are represented in Table 18.

Table 18 Viterbi Algorithm Results for IL-1β

NUMBER	SEQUENCE	SEQUENCE (ONE LETTER REPRESENTATION)	PATH NUMBER	BINDING FREE ENERGY (Kcal/mol)
1	TRP-ALA-LYS-LEU-LEU-LEU-GLU	WAKLLLE	1	-8.1
2	ASP-TYR-CYS-TYR-THR-VAL-VAL	DYCYTVV	1	-9.2

3	TRP-TYR-TYR-PHE-THR-TRP-TRP	WYYFTWW	1	-9.7
4	GLY-VAL-GLY-GLY-VAL-GLU-GLU	GVGGVEE	1	-6.3
5	GLN-ALA-LEU-ALA-CYS-CYS-CYS	QALACCC	1	-7.2
6	HIS-SER-PRO-PRO-PRO-PRO-PRO	HSPPPPP	1	-8.72
7	LYS-SER-SER-THR-THR-GLY-PRO	KSSTTGP	1	-9.02
8	GLN-ALA-ILE-MET-CYS-CYS-CYS	QAIMCCC	1	-7.4
9	HIS-PHE-CYS-CYS-CYS-CYS-CYS	HFCCCCC	1	-8.96
10	LEU-LEU-CYS-CYS-CYS-CYS-CYS	LLCCCCC	1	-6.2
11	TRP-GLU-SER-SER-TRP-PRO-SER	WESSWPS	2	-9.0
12	LYS-VAL-VAL-PRO-VAL-VAL-VAL	KVVPVVV	2	-6.1
13	PRO-GLU-LEU-PRO-PRO-ILE-PRO	PELPPIP	2	-7.0

14	TRP-TRP-PRO-ARG-VAL-PRO-ARG	WWPRVPR	2	-9.2
15	TRP-CYS-PRO-CYS-CYS-ASP-ASP	WCPCCDD	2	-7.06
16	ASN-PRO-PRO-PRO-PRO-PRO-PRO	NPPPPPP	2	-7.1
17	ASN-PRO-PRO-VAL-VAL-ILE-ILE	DPPVVII	2	-6.0
18	TRP-PHE-PHE-CYS-ASP-PRO-ASP	WFFCDPD	3	-8.75
19	TRP-GLN-SER-TRP-LEU-LEU-LEU	WQSQLLL	3	-9.0
20	GLU-ALA-THR-VAL-ILE-ILE-ILE	EATVIII	3	-10.35
21	GLU-ALA-PRO-PRO-PRO-PRO-PRO	EAPPPPP	3	-8.9
22	THR-SER-LEU-GLN-GLN-SER-THR	TSLQQST	3	-8.5
23	TRP-PHE-TRP-LEU-ASP-ASP-ASP	WFWLDDD	3	-10.81
24	TRP-PRO-ILE-ILE-ILE-LYS-PRO	WPIIIKP	3	-8.8

25	HIS-PRO-CYS-CYS-PRO-PRO-PRO	HPCCPPP	3	-8.8
26	HIS-PRO-CYS-CYS-HIS-PRO-PRO	HPCCHPP	3	-8.0
27	ALA-LEU-SER-SER-LEU-SER-SER	ALSSLSS	3	-7.04
28	HIS-PRO-PRO-PRO-PRO-PRO-PRO	HPPPPPP	4	-8.3
29	TYR-LEU-LEU-LEU-LEU-LEU-LYS	YLLLLLK	4	-6.9
30	PRO-GLU-GLU-PRO-PRO-PRO-PRO	PEEPPPP	4	-9.1

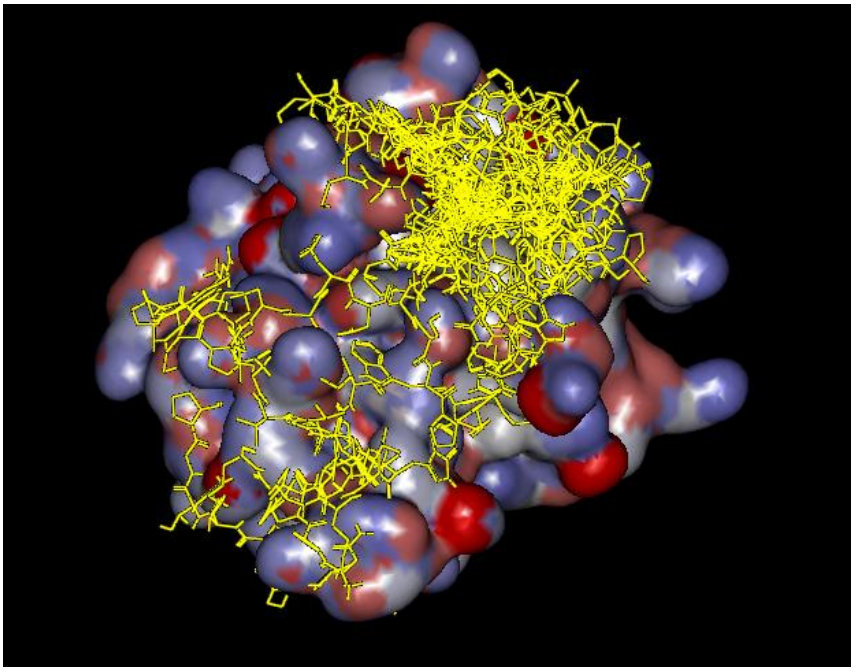

Figure 34 Surface representation of IL-1β with potent inhibitors (yellow) from the Viterbi Algorithm.

As can be seen from Figure 34, most of the potent inhibitors gathered in one cavity. This cavity includes most of the hot points including ILE 56 from the Boraschi Loop. As represented earlier the Boraschi Loop (β bulge loop) plays an important role in binding to receptor. Thus this grouping in a cavity occasion shows that our results are consistent with the literature. As illustrated earlier the Viterbi algorithm gives the most potent path to the ligand without restricting it. Hence the potential ligand finds the suitable cavity and conformation for itself.

By referencing the similarities between IL-1β and IL-1α, we tested all potent inhibitors on IL-1α. Results can be seen in Table 19. Six different binding regions to the receptor were selected as docking points. As expected, binding affinities are high meaning that our peptides inhibit both α and β.

Table 19 Docking Results of Potent Inhibitors to IL-1α and IL-1β

NUMBER	SEQUENCE	SEQUENCE (ONE LETTER REPRESENTATION)	BINDING FREE ENERGY TO IL-1β (Kcal/mol)	BINDING FREE ENERGY TO IL-1α (Kcal/mol)
1	TRP-ALA-LYS-LEU-LEU-LEU-GLU	WAKLLLE	-8.1	-7.81
2	ASP-TYR-CYS-TYR-THR-VAL-VAL	DYCYTVV	-9.2	-8.75
3	TRP-TYR-TYR-PHE-THR-TRP-TRP	WYYFTWW	-9.7	-9.06
4	GLY-VAL-GLY-GLY-VAL-GLU-GLU	GVGGVEE	-6.3	-7.12
5	GLN-ALA-LEU-ALA-CYS-CYS-CYS	QALACCC	-7.2	-7.618
6	HIS-SER-PRO-PRO-PRO-PRO-PRO	HSPPPPP	-8.72	-7.89
7	LYS-SER-SER-THR-THR-GLY-PRO	KSSTTGP	-9.02	-10.36
8	GLN-ALA-ILE-MET-	QAIMCCC	-7.4	-7.81

	CYS-CYS-CYS			
9	HIS-PHE-CYS-CYS-CYS-CYS-CYS	HFCCCCC	-8.96	-8.885
10	LEU-LEU-CYS-CYS-CYS-CYS-CYS	LLCCCCC	-6.2	-9.57
11	TRP-GLU-SER-SER-TRP-PRO-SER	WESSWPS	-9.0	-8.184
12	LYS-VAL-VAL-PRO-VAL-VAL-VAL	KVVPVVV	-6.1	-8.33
13	PRO-GLU-LEU-PRO-PRO-ILE-PRO	PELPPIP	-7.0	-8.021
14	TRP-TRP-PRO-ARG-VAL-PRO-ARG	WWPRVPR	-9.2	-11.55
15	TRP-CYS-PRO-CYS-CYS-ASP-ASP	WCPCCDD	-7.06	-8.86
16	ASN-PRO-PRO-PRO-PRO-PRO-PRO	NPPPPPP	-7.1	-9.80
17	ASN-PRO-PRO-VAL-VAL-ILE-ILE	DPPVVII	-6.0	-8.211
18	TRP-PHE-PHE-CYS-ASP-PRO-ASP	WFFCDPD	-8.75	-9.49

19	TRP-GLN-SER-TRP-LEU-LEU-LEU	WQSQLLL	-9.0	-6.53
20	GLU-ALA-THR-VAL-ILE-ILE-ILE	EATVIII	-10.35	-9.4
21	GLU-ALA-PRO-PRO-PRO-PRO-PRO	EAPPPPP	-8.9	-8.20
22	THR-SER-LEU-GLN-GLN-SER-THR	TSLQQST	-8.5	-7.06
23	TRP-PHE-TRP-LEU-ASP-ASP-ASP	WFWLDDD	-10.81	-8.23
24	TRP-PRO-ILE-ILE-ILE-LYS-PRO	WPIIIKP	-8.8	-9.83
25	HIS-PRO-CYS-CYS-PRO-PRO-PRO	HPCCPPP	-8.8	-7.36
26	HIS-PRO-CYS-CYS-HIS-PRO-PRO	HPCCHPP	-8.0	-6.70
27	ALA-LEU-SER-SER-LEU-SER-SER	ALSSLSS	-7.04	-7.97
28	HIS-PRO-PRO-PRO-PRO-PRO-PRO	HPPPPPP	-8.3	-8.08
29	TYR-LEU-LEU-LEU-LEU-LEU-LYS	YLLLLLK	-6.9	-6.72

30	PRO-GLU-GLU-PRO-PRO-PRO-PRO	PEEPPPP	-9.1	-9.31

4.2.3. SELECTION OF LIGANDS

In order to achieve one solution, a rational elimination was needed among 30 potential inhibitors. Another docking program (GOLD [47]) was used to recalculate binding free energies to the target protein. A potential inhibitor should give correlate solutions in both docking programs. To eliminate ligands by GOLD, same docking places in Autodock were used. As expected GOLD and AutoDock programs gave correlated results mostly. Due to time conserving process of Molecular Dynamics Simulations 30 inhibitors were decided to decline to four. After Molecular Dynamics Simulations were completed, these potential inhibitors were reduced to one solution.

Five criteria were considered during elimination. 1) Recalculation of binding free energies by GOLD. 2) Autodock binding free energies and its correlation with GOLD results. 3) Having similar binding free energies for IL-1α. 4) Having non tryptophan including peptide sequences. 5) Having non recursive amino acids in peptide sequences.

As tryptophan gives lower binding energy results in each docking, ligands that including tryptophan were eliminated. By doing this, we increased the specificity of the peptide to the target surface. Rest of the ligands were eliminated due to the uncorrelated results with GOLD. Also the natural peptide motif formation was considered. Occurrence of recursive amino acid sequences is rare in nature. Thus recurring amino acid sequences were eliminated such as LLCCCCC (number 10). These occurrences were checked from the website http://cssp2.sookmyung.ac.kr [75]. Eliminated ligands are shown in dark colour in Table 20.

Table 20 Binding Free Energy Comparison with GOLD

NUMBER	SEQUENCE	SEQUENCE (ONE LETTER REPRESENTATION)	BINDING FREE ENERGY TO IL-1β (Kcal/mol)	BINDING FREE ENERGY TO IL-1α (Kcal/mol)	BINDING FREE ENERGY TO IL-1β by GOLD (Kcal/mol)
1	TRP-ALA-LYS-LEU-LEU-LEU-GLU	WAKLLLE	-8.1	-7.81	-5.10
2	ASP-TYR-	DYCYTVV	-9.2	-8.75	-6.28

	CYS-TYR-THR-VAL-VAL				
3	TRP-TYR-TYR-PHE-THR-TRP-TRP	WYYFTWW	-9.7	-9.06	-10.9
4	GLY-VAL-GLY-GLY-VAL-GLU-GLU	GVGGVEE	-6.3	-7.12	-6.02
5	GLN-ALA-LEU-ALA-CYS-CYS-CYS	QALACCC	-7.2	-7.618	-3.26
6	HIS-SER-PRO-PRO-PRO-PRO-PRO	HSPPPPP	-8.72	-7.89	-3.68
7	LYS-SER-SER-THR-THR-GLY-PRO	KSSTTGP	-9.02	-10.36	-0.96
8	GLN-ALA-ILE-MET-CYS-CYS-CYS	QAIMCCC	-7.4	-7.81	-3.01
9	HIS-PHE-CYS-CYS-CYS-CYS-CYS	HFCCCCC	-8.96	-8.885	-3.90
10	LEU-LEU-CYS-CYS-CYS-CYS-CYS	LLCCCCC	-6.2	-9.57	-4.20
11	TRP-GLU-	WESSWPS	-9.0	-8.184	-3.68

	SER-SER-TRP-PRO-SER				
12	LYS-VAL-VAL-PRO-VAL-VAL-VAL	KVVPVVV	-6.1	-8.33	-6.75
13	PRO-GLU-LEU-PRO-PRO-ILE-PRO	PELPPIP	-7.0	-8.021	-5.89
14	TRP-TRP-PRO-ARG-VAL-PRO-ARG	WWPRVPR	-9.2	-11.55	-10.77
15	TRP-CYS-PRO-CYS-CYS-ASP-ASP	WCPCCDD	-7.06	-8.86	-4.14
16	ASN-PRO-PRO-PRO-PRO-PRO-PRO	NPPPPPP	-7.1	-9.80	-6.13
17	ASN-PRO-PRO-VAL-VAL-ILE-ILE	DPPVVII	-6.0	-8.211	-5.42
18	TRP-PHE-PHE-CYS-ASP-PRO-ASP	WFFCDPD	-8.75	-9.49	-6.83
19	TRP-GLN-SER-TRP-LEU-LEU-LEU	WQSQLLL	-9.0	-6.53	-6.57
20	GLU-ALA-THR-VAL-	EATVIII	-10.35	-9.4	-6.60

	ILE-ILE-ILE				
21	GLU-ALA-PRO-PRO-PRO-PRO-PRO	EAPPPPP	-8.9	-8.20	-6.28
22	THR-SER-LEU-GLN-GLN-SER-THR	TSLQQST	-8.5	-7.06	-0.2
23	TRP-PHE-TRP-LEU-ASP-ASP-ASP	WFWLDDD	-10.81	-8.23	-5.60
24	TRP-PRO-ILE-ILE-ILE-LYS-PRO	WPIIIKP	-8.8	-9.83	-6.93
25	HIS-PRO-CYS-CYS-PRO-PRO-PRO	HPCCPPP	-8.8	-7.36	-6.51
26	HIS-PRO-CYS-CYS-HIS-PRO-PRO	HPCCHPP	-8.0	-6.70	-4.14
27	ALA-LEU-SER-SER-LEU-SER-SER	ALSSLSS	-7.04	-7.97	-1.69
28	HIS-PRO-PRO-PRO-PRO-PRO-PRO	HPPPPPP	-8.3	-8.08	-7.24
29	TYR-LEU-LEU-LEU-LEU-LEU-LYS	YLLLLLK	-6.9	-6.72	-7.74

30	PRO-GLU-GLU-PRO-PRO-PRO-PRO	PEEPPPP	-9.1	-9.31	-4.54

Ligand numbers 1, 3, 11, 14, 15, 18, 19, 23 and 24 eliminated due to containing tryptophan. Recursive elements mostly occur in 6, 9, 10, 16, 21, 28, 29 and 30. These ligands were also discarded. From the remaining ligands 5, 22 and 27 gave uncorrelated results with GOLD. Due to the time-consuming process of Molecular Dynamics Simulations 4 potent inhibitors were decided to select. From the rest of ligands, the ones with higher binding free energy than -7 kcal/mol were also ignored. Remaining numbers were 2, 13, 20, 25 and 26. Since 25 and 26 are so close both in sequence and binding energy, the one with the lowest docking energy was chosen (number 25).

4.3. STEERED MOLECULAR DYNAMICS RESULTS

Remaining ligands from the Viterbi Algorithm elimination and their binding free energies can be seen in Table 21.

Table 21 Binding Free Energies

LIGANDS	AutoDock (kcal/mol)
EATVIII	-10.12
DYCYTVV	-8.4
PELPPIP	-7.93
HPCCPPP	-8.8

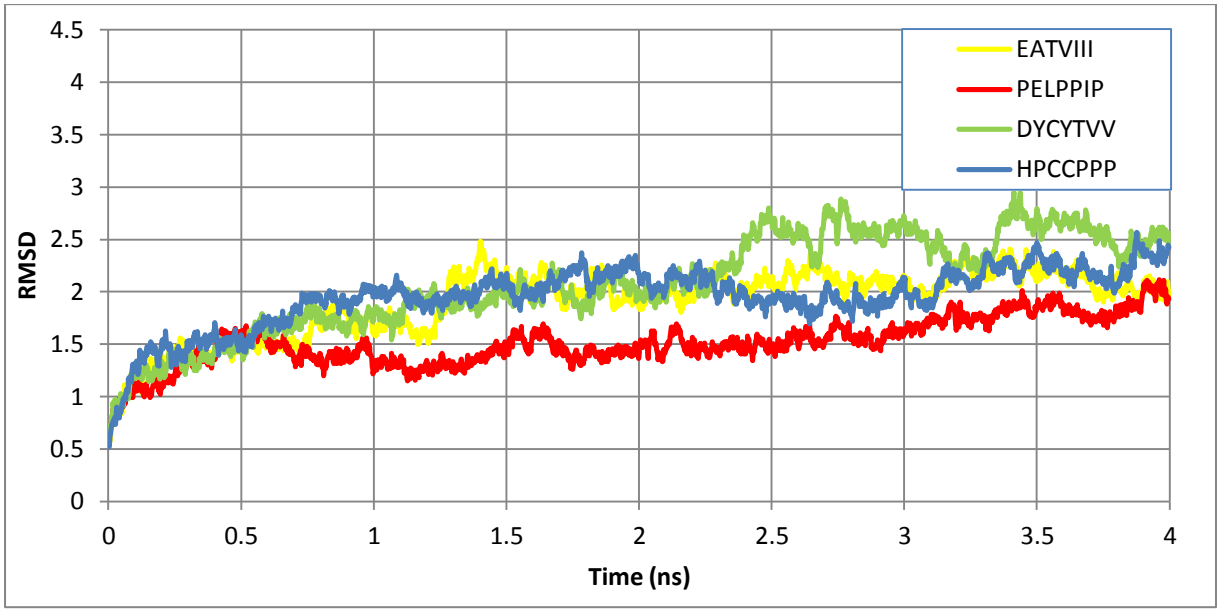

Figure 35 RMSD Graphs for four ligands.

To choose most proper inhibitor to IL-1β, this time elimination with Steered Molecular Dynamics (SMD) were done. For each ligand three different steered molecular dynamics simulations were carried on from different snapshots of NVT simulation. After RMSD graphs converged for each ligand, snapshots form 3 ns, 3.5 ns and 4 ns of total 4 ns simulations were taken and SMD were applied.

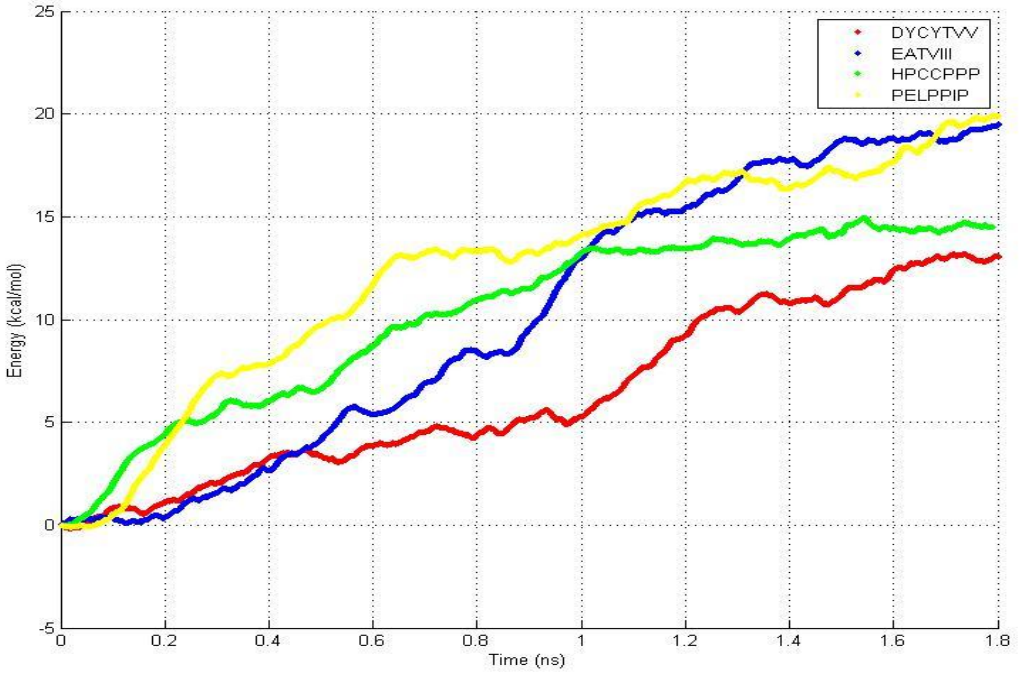

Figure 36 Unbinding Energies of remaining four ligands.

Figure 36 shows the calculated unbinding free energies for chosen ligands. All energy graphs converge after its unbinding occurs. Here the ligand with the sequence D-Y-C-Y-T-V-V converges about 1.2 ns of the total 1.8 ns simulation. The increase after convergence comes from the friction forces encountered during unbinding path. This ligand gives approximately 10 kcal/mol unbinding energy and eliminated for later simulations due to its lower trend within others. As expected from its Autodock results, the ligand with the sequence E-A-T-V-I-I-I gave a higher trend in SMD simulations and saved for later simulations. It can be said that its convergence and unbinding started at 1.5 ns of the simulation and its unbinding energy is 17 kcal/mol. Although the ligand P-E-L-P-P-I-P gave highest energy from Autodock, it showed more consistent results with E-A-T-V-I-I-I. Its unbinding energy is similar to E-A-T-V-I-I-I while H-P-C-C-P-P-P showed lower trend in unbinding. Thus, these two ligands (E-A-T-V-I-I-I and P-E-L-P-I-P-P) were chosen for later simulations. All energies were calculated using Jarzynski equality.

4.3.1. STEERED MD RESULTS FOR E-A-T-V-I-I-I AND P-E-L-P-P-I-P

Ten SMD simulations from different snapshots of NVT simulation were carried out in order to understand most proper candidate. After ten simulations, trend of both energy and force graphs became more accurate. Snapshots were taken from 3.00, 3.125, 3.25, 3.375, 3.5, 3.625,

3.75, 3.875, 3.9375 and 4.00 ns of total 4 ns simulation. It is especially considered to take snapshots from the most converged part of RMSD graphs.

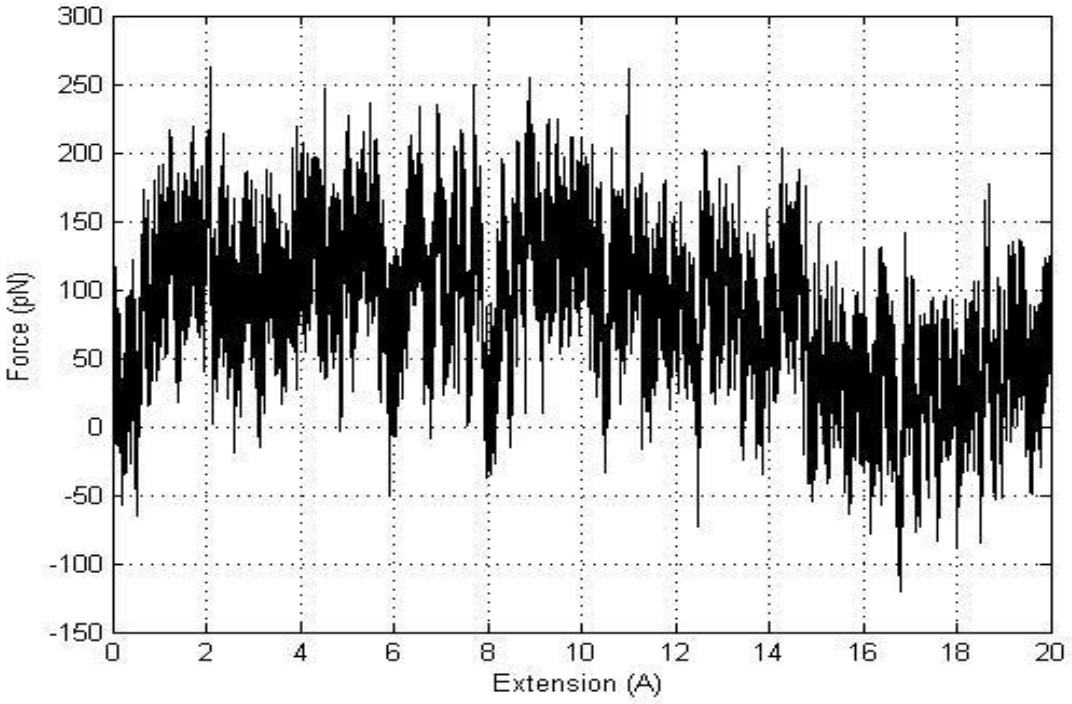

Figure 37 Force distributions through extension for E-A-T-V-I-I-I.

Figure 37 shows the distribution through extension graph of candidate E-A-T-V-I-I-I. An increasing trend between 0 and 14 Angstrom of total 20 Angstrom or 2 ns simulation indicates that the most of the interaction loss occurs within this region. The highest force peak was recorded as 270pN while the lowest was -60pN. Force values can fluctuate between positive-negative values due to the thermal fluctuations in protein.

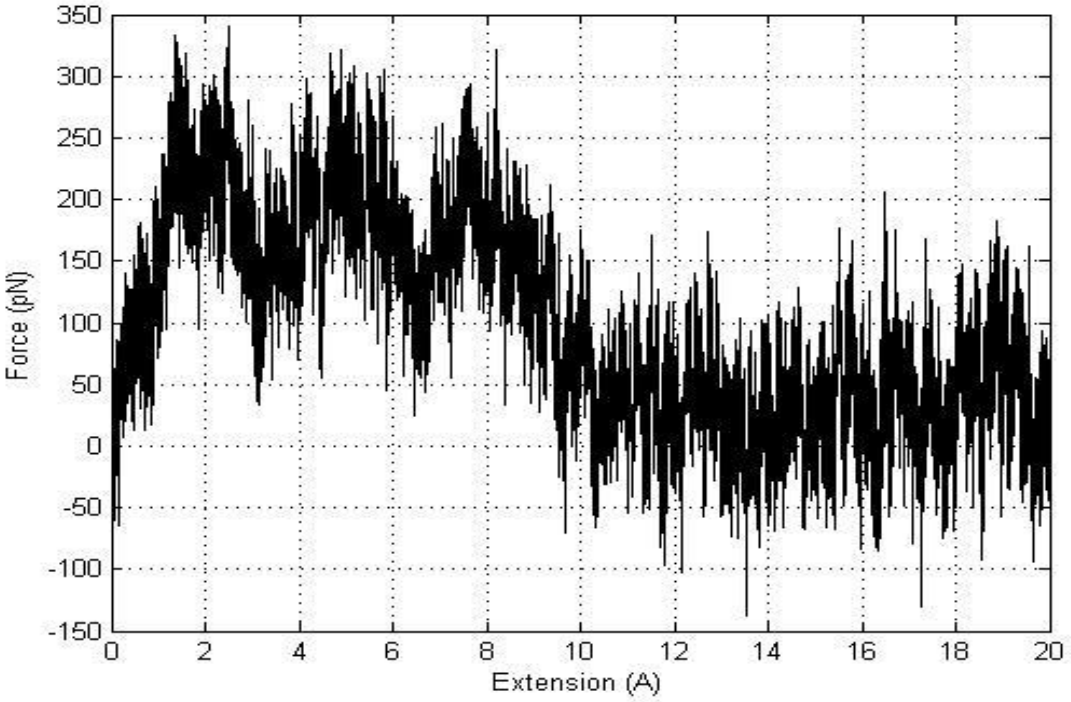

Figure 38 Force distributions through extension graph for P-E-L-P-P-I-P.

Figure 38 shows the Force distribution through extension graph for the second candidate P-E-L-P-P-I-P. This candidate gave higher peaks than E-A-T-V-I-I-I within a short range. The highest peak was recorded 330pN within 10 Angstrom or 1 ns simulation.

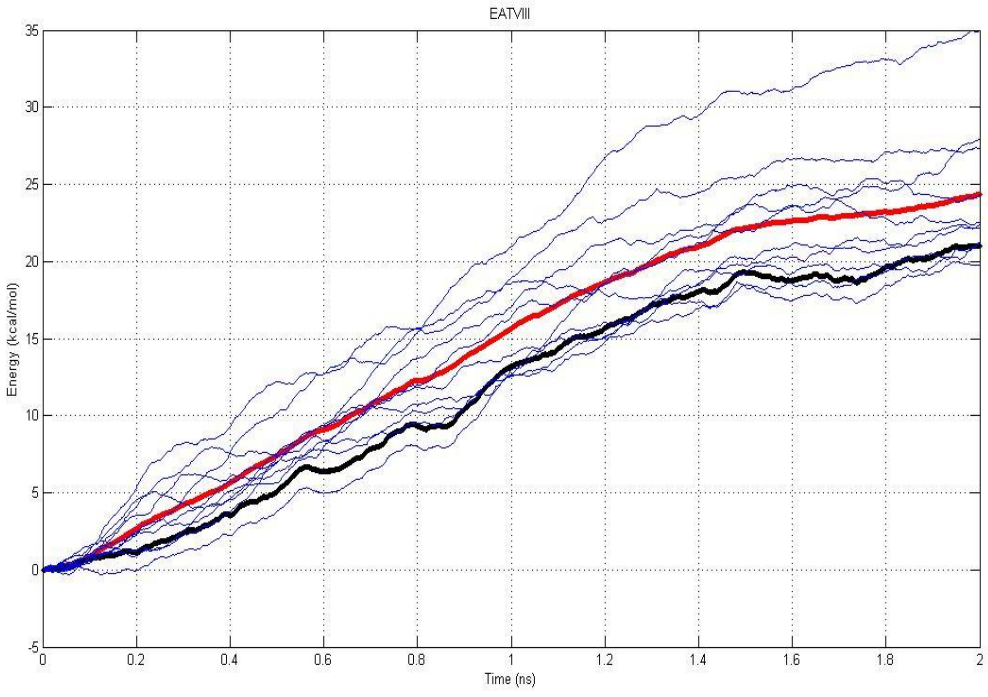

Figure 39 Unbinding energy change graph for E-A-T-V-I-I-I.

As can be seen from Figure 39, the unbinding energy of E-A-T-V-I-I-I is similar to what we observed from Figure 36, 18kcal/mol. Its convergence starts from 1.4ns simulation meaning no interactions were recorded after this point. The blue lines indicate 10 SMD simulations, red line indicates the average of simulations and black line indicates the profile calculated from Jarzynski Equation.

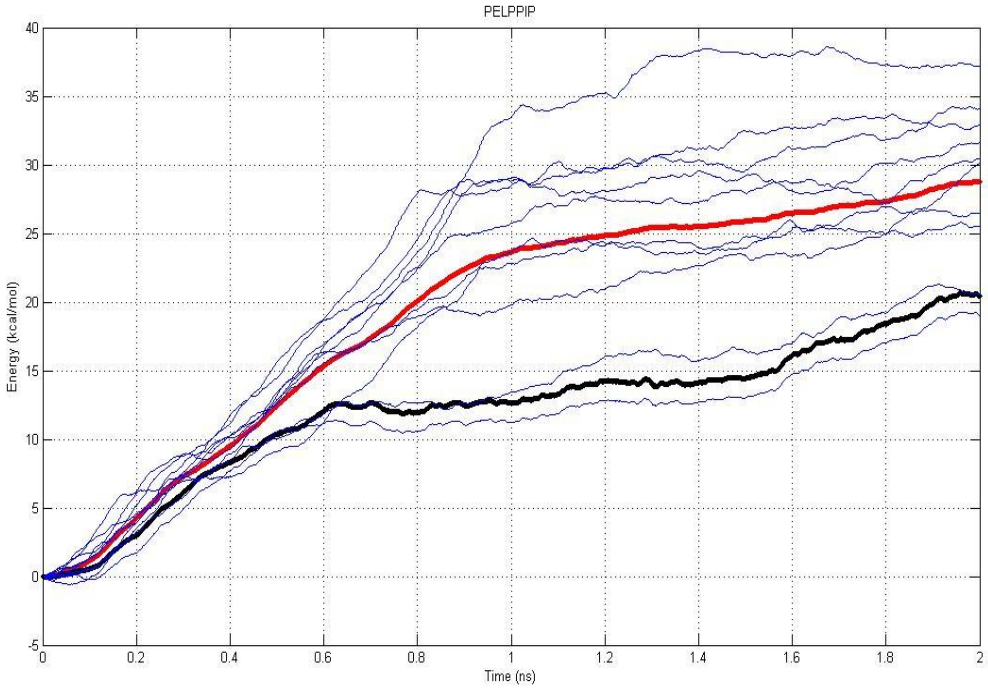

Figure 40 Unbinding energy change graph for P-E-L-P-P-I-P.

Unbinding profile of P-E-L-P-P-I-P changed dramatically in comparison with Figure 36. Its energy decreases to the 12 kcal/mol while the unbinding finishes at 0.6 ns of 2 ns simulation. Again, the blue lines indicate all 10 SMD simulations while red gives the average and the black one energy from the Jarzynski Equality.

By referencing the force and energy graphs, candidate P-E-L-P-P-I-P requires less energy, unbinds easier in comparison with E-A-T-V-I-I-I. But the unbinding of P-E-L-P-P-I-P gave higher peaks in force graphs while E-A-T-V-I-I-I force profile trends in more smooth way. This means E-A-T-V-I-I-I interactions are stronger than P-E-L-P-P-I-P and unbinding occurs in a slower way.

To observe unbinding process in detail, behaviour of each atom during the path was plotted. Atoms which are further to the protein than 3.5 Angstrom were not considered due to the Van der Waals interaction radius standards [76].

Figure 41 and Figure 42 show the cumulative probability distribution of each atom through extension.

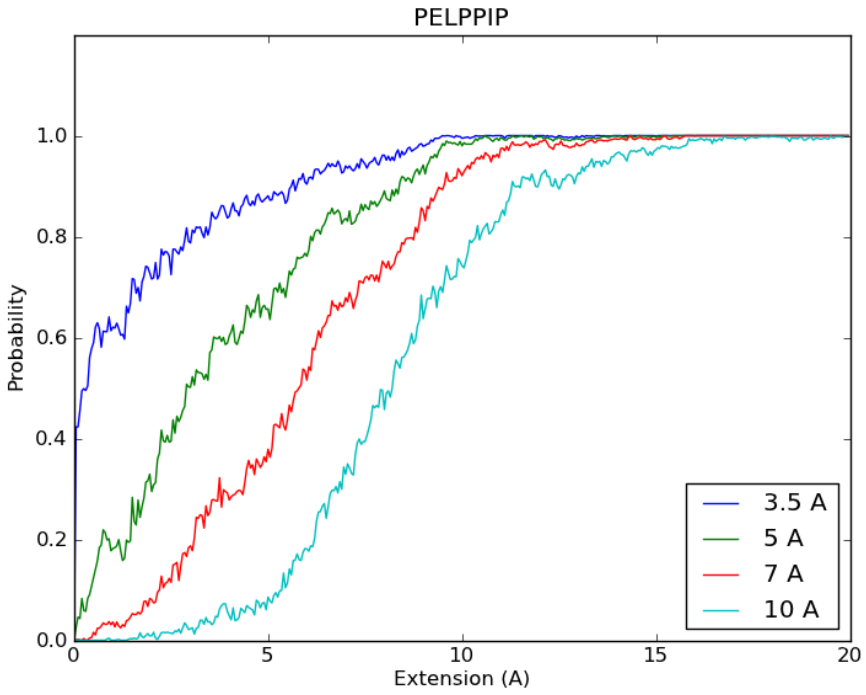

Figure 41 Probability distribution of P-E-L-P-P-I-P through unbinding path with changing protein-ligand intervals.

Figure 41 shows the probability distribution of candidate P-E-L-P-P-I-P. We considered four different distances as a start (D). These distances indicate the intervals between protein and ligand in Angstroms at the start of the simulation. As the distance between protein and ligand increases the graph shifts to the right.

Due to the van der Waals radius limitations we considered the atoms between two chains which are closer than 3.5 Angstrom. If we were to measure the probability of atoms that pass 5 Angstrom through simulation, the graph trend to increase until each atom passes 5 Angstrom distance. For D=5 Angstrom, we can conclude that all atoms passed the defined distance about 10 Angstrom extension. The reason of coupling is the flexibility and conformational changes of protein during unbinding. In SMD simulations, only one atom was defined as pulling atom. Thus not each part of the ligand moves in the same fashion. If we

were to measure the atoms passing 7 Angstrom distance, the stability point shifts to 11 Angstrom. This situation is also similar for D=10 Angstrom. The unbinding point shifts to 15 Angstrom. If we look back Figure 40, we notice that SMD simulations gave the unbinding point about 6 Angstrom.

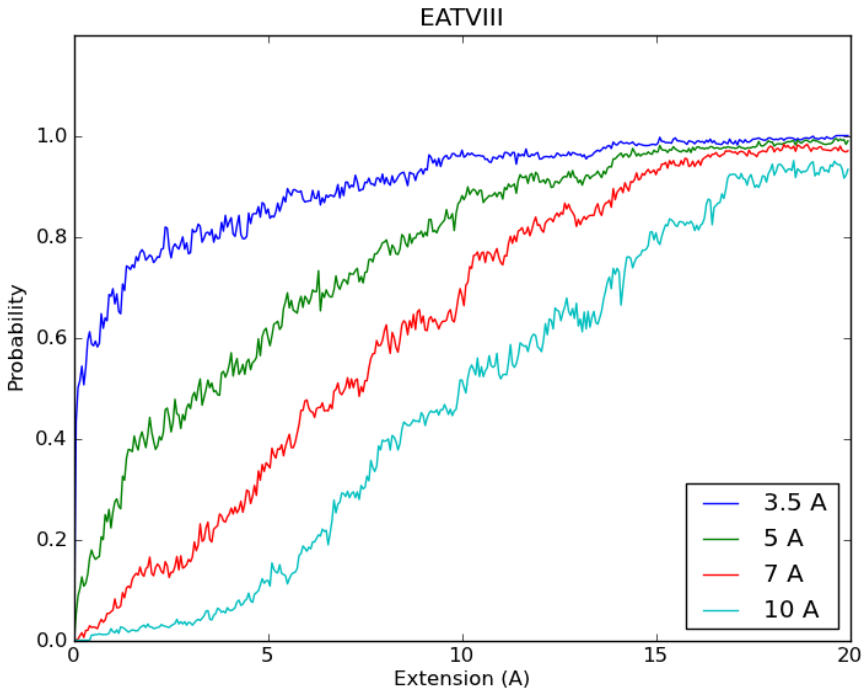

Figure 42 Probability distribution of E-A-T-V-I-I-I through unbinding path with changing protein-ligand intervals.

Figure 42 shows the probability distribution of the second candidate. As we concluded, this ligand requires less force but larger time and distance to unbind due to the force & energy graphs from SMD simulations. Again, distance changes made the plots to shift as expected. Here the unbinding points also correlate with the one obtained from Figure 39. All unbinding points except while D=10, tends to converge about 14 Angstrom of total 20 Angstrom extension.

Thus, the detailed probability distribution of each atom through extension graphs helped us which candidate is more proper for inhibition. With the help of Jarzynski equality, the candidate with the sequence E-A-T-V-I-I-I said to be more convenient due to its difficulty in unbinding with greater requirement of the unbinding energy.

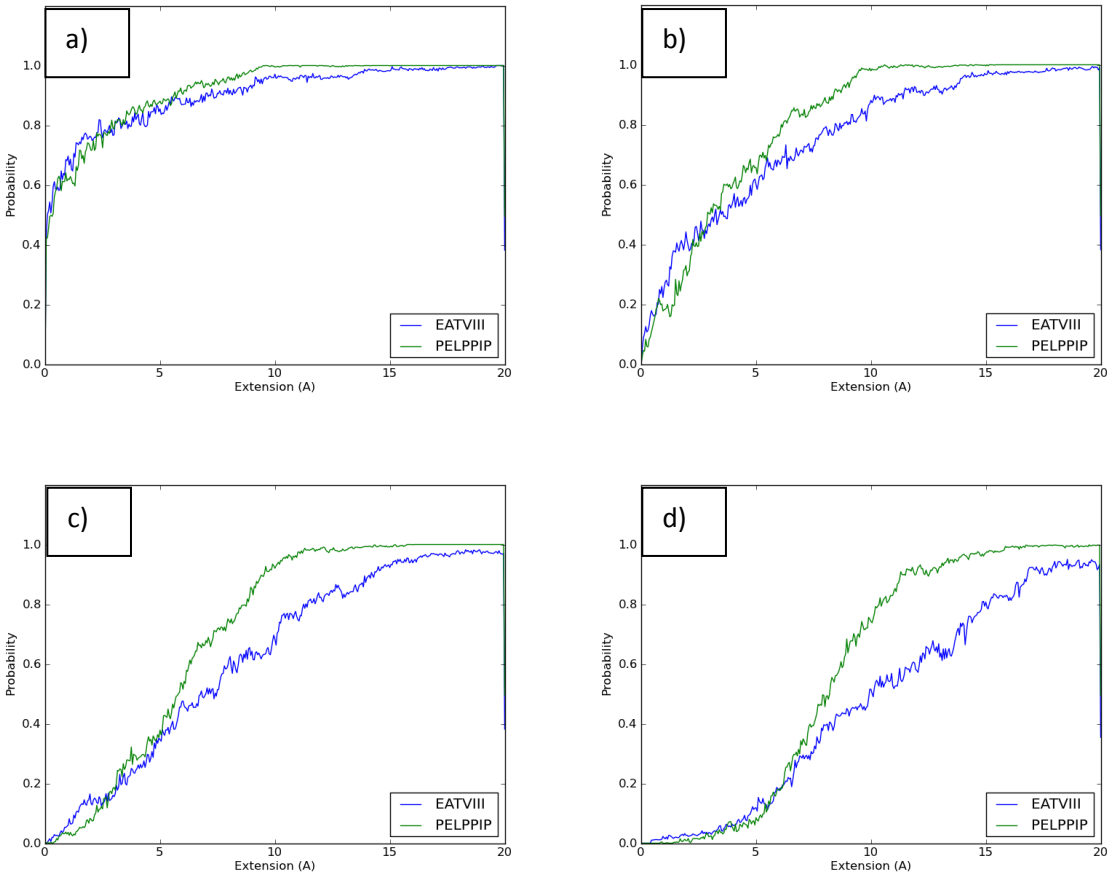

Figure 43 Cumulative differences in two candidates.

a) D=3.5Å b) D=5Å c) D=7Å d) D=10Å

Figure 43 shows the differences of two candidates in cumulative probability distribution. The differences in two ligands increase while reference interval between ligand and protein grows. Starting from the Van der Waals interaction limit to 10Å, the passage of each atom from the interval gets harder. Thus the probability of the atom that passes to limit grows slowly. Candidate E-A-T-V-I-I-I shows a lower trend and greater time in unbinding.

Curve fitting to the ligands with different distances was applied in order to obtain numerical differences in two ligands. They both resemble to the function $\int x^4 e^{Dx}$ where D is the distance change.

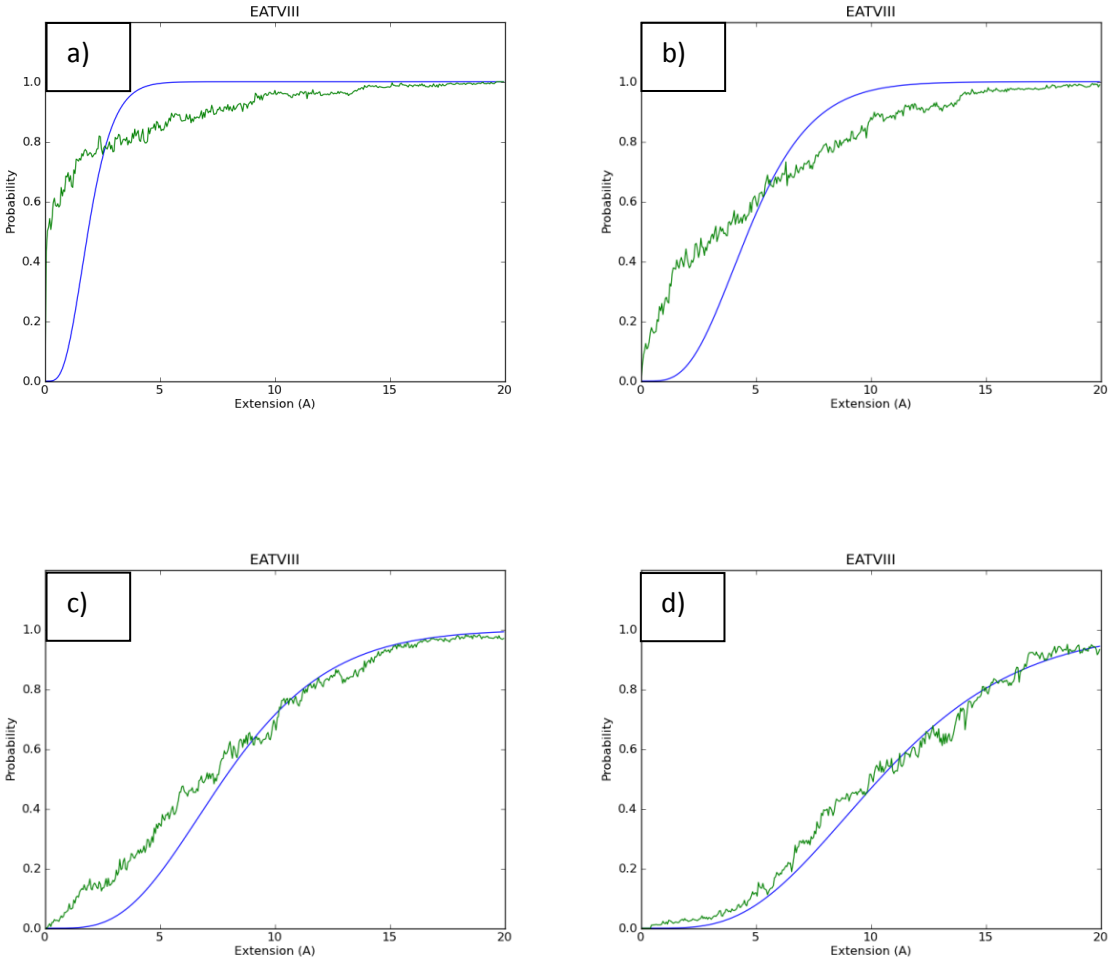

Figure 44 Curve fitting for E-A-T-V-I-I-I.

a) D=3.5Å b) D=5Å c) D=7Å d) D=10Å

Figure 44 shows the curve fitting graphs with the changes in intervals for ligand E-A-T-V-I-I-I. As the distance increases our data fits curve function better.

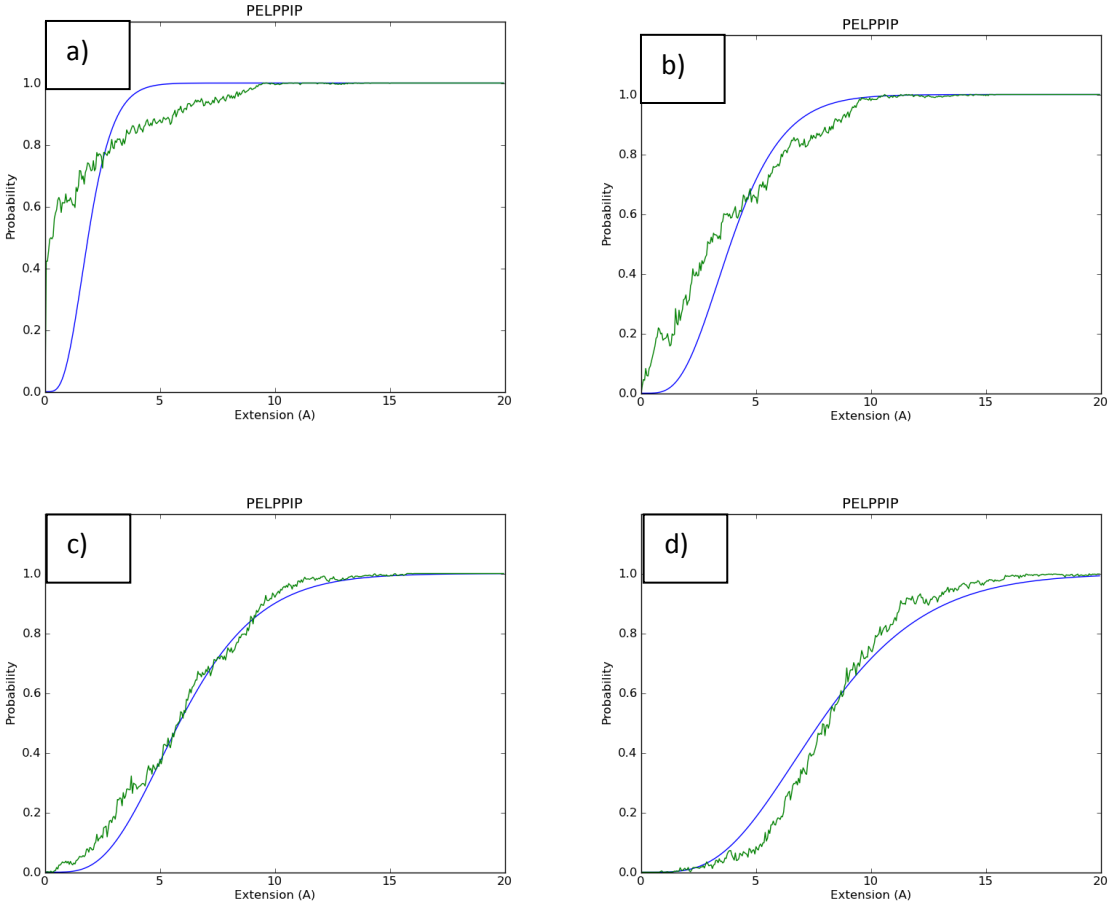

Figure 45 Curve fitting for P-E-L-P-P-I-P.

a) D=3.5Å b) D=5Å c) D=7Å d) D=10Å

Figure 45 gives the curve fitting graphs with the changes in intervals for ligand P-E-L-P-P-I-P. The difference between two ligands increased as the distance increased. In numerical expression, the constants in function that determine the difference between two ligands diverge.

CHAPTER 5

5. CONCLUSION

In this study two computational methods were applied in order to design a peptide inhibitor to IL-1β to control over secretion to the extracellular region.

Firstly a tripeptide candidate T-S-W (THR-SER-TRP) was obtained as a potent inhibitor from Genetic Algorithm. This algorithm mimics the biological evolution. Each candidate exposed to mutation by replacing the amino acids in each population. By ranking their binding free energies results evolved and started to converge to the same sequence in each population. At the end of 13^{th} population, sequence T-S-W appeared and continued till the 20^{th} population as the most potent one. After obtaining the converged sequence, by fixing the sequence we increased the amino acid number to increase specificity. This is done by docking amino acids to the right and left hand sides of the fixed protein. Also 400 dipeptide combinations were docked to the selected tripeptide and showed better results. W-T-S-W-R, C-T-S-W-S and W-T-S-W-S were showed lower binding energies as potent inhibitors.

Secondly, to increase the specificity potent heptapeptides were designed via the Viterbi Algorithm which controls the peptides compatibility by calculating both its binding free energy and ψ-φ angles. This algorithm generates peptide inhibitors by docking its residues pair by pair to a selected path, called the Viterbi path, along the protein surface by checking its secondary structure conformation. The Hidden Markov Models states that the outcome of any process affects the next process in a chainlike manner. By saying Hidden, one considers the states which are not directly visible to observer but visible according to output.

Total of thirty candidates were obtained from Viterbi Algorithm and firstly eliminated according to their binding free energies and non-recursive sequences. Recursive amino acid sequences are rare in nature thus two candidates which give higher unbinding free energies via GOLD and Autodock Programs and show non-recursive sequences were selected as the potent candidates. All of the potent inhibitors gathered in one cavity on the surface which includes most of the hot points including ILE 56 from the Boraschi Loop. The Boraschi Loop, also known as β bulge loop, plays an important role in binding to the receptor. This gathering shows that our algorithm finds the most probable path correctly and gives consistent results with the literature.

These four remaining candidates with sequences E-A-T-V-I-I-I (GLU-ALA-THR-VAL-ILE-ILE-ILE), D-Y-C-Y-T-V-V (ASP-TYR-CYS-TYR-THR-VAL-VAL), H-P-C-C-P-P-P (HIS-PRO-CYS-CYS-PRO-PRO-PRO) and P-E-L-P-P-I-P (PRO-GLU-LEU-PRO-PRO-ILE-PRO) were then eliminated by a second method, Molecular Dynamics. The four peptide complexes were minimized and equilibrated. By an external force peptides were pulled through chosen direction, and the free energy during this unbinding process was calculated with the Jarzynski Equality. This equality states that the Helmholtz free-energy difference between two equilibrium configurations of a system can be obtained from an ensemble of non-equilibrium measurements of the work performed in switching an external parameter of the system. Steered molecular dynamics applied to each candidate three times for a third elimination. From them two of them gave higher unbinding energies and further calculations were continued with the candidates E-A-T-V-I-I-I and P-E-L-P-P-I-P. For each ligand ten different steered molecular dynamics simulations were carried on from different snapshots of simulations.

As a result, candidate E-A-T-V-I-I-I showed greater unbinding energy and required more time through extension which means this ligand is a more suitable candidate than the candidate P-E-L-P-P-I-P. It fits the cavity in a way that unbinding requires more energy which makes the unbinding process difficult. To understand the changes in unbinding process, movement of each atom was followed and a distribution through unbinding was obtained. These distributions helped us to understand the differences of the ligands in graphical representations.

Also our inhibitor candidates showed high binding free energies in the similar structure IL-1α which also binds to the same receptor as IL-1β. This situation resembles anakinra treatment which has been used over 10 years in the treatment of interleukin family related diseases.

Among two algorithms, the Viterbi Algorithm results are considered to be more accurate due to following reasons: 1) The Viterbi Algorithm controls both binding free energies to the protein and its structural conformation coherence 2) It gives seven peptide long sequence inhibitors which increases the specificity and prevents binding to untargeted surfaces. 3) The results were also controlled via Molecular Dynamics Simulations. On the other hand, the algorithm choice depends on the length of the peptide and for small peptide inhibitors the Genetic Algorithm is more applicable.

As a result, a potent inhibitor for IL-1β and IL-1α was found and two algorithms in peptide design were applied successfully. A seven amino acid long peptide sequence E-A-T-V-I-I-I was found as the most appropriate one. Future work of this study might be the experimental control in binding since all results given here verified computationally.

6. BIBLIOGRAPHY

1. Sato, A.K., et al., *Therapeutic peptides: technological advances driving peptides into development.* Curr Opin Biotechnol, 2006. **17**(6): p. 638-42.
2. Dinarello, C.A., *Immunological and inflammatory functions of the interleukin-1 family.* Annu Rev Immunol, 2009. **27**: p. 519-50.
3. Unal, E.B., Gursoy, A., Erman, B., *Inhibitor Peptide Design For NF-ΚB: Markov Model & Genetic Algortihm*, in *HIBIT*. 2010: Antalya,TURKEY.
4. Kamphausen, S., et al., *Genetic algorithm for the design of molecules with desired properties.* J Comput Aided Mol Des, 2002. **16**(8-9): p. 551-67.
5. Holland, J.M., *Natural history and staging of renal cell carcinoma.* CA Cancer J Clin, 1975. **25**(3): p. 121-33.
6. Erman, B., *The gaussian network model: precise prediction of residue fluctuations and application to binding problems.* Biophys J, 2006. **91**(10): p. 3589-99.
7. Haliloglu, T., E. Seyrek, and B. Erman, *Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model.* Phys Rev Lett, 2008. **100**(22): p. 228102.
8. Vigers, G.P., et al., *Crystal structure of the type-I interleukin-1 receptor complexed with interleukin-1beta.* Nature, 1997. **386**(6621): p. 190-4.
9. Tuncbag, N., A. Gursoy, and O. Keskin, *Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy.* Bioinformatics, 2009. **25**(12): p. 1513-20.
10. Vigers, G.P., et al., *X-ray crystal structure of a small antagonist peptide bound to interleukin-1 receptor type 1.* J Biol Chem, 2000. **275**(47): p. 36927-33.
11. Heidary, D.K., et al., *Long-range coupling between separate docking sites in interleukin-1beta.* J Mol Biol, 2005. **353**(5): p. 1187-98.
12. Quiniou, C., et al., *Development of a novel noncompetitive antagonist of IL-1 receptor.* J Immunol, 2008. **180**(10): p. 6977-87.
13. Wang, D., et al., *Structural insights into the assembly and activation of IL-1beta with its receptors.* Nat Immunol, 2010. **11**(10): p. 905-11.
14. Schreuder, H., et al., *A new cytokine-receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist.* Nature, 1997. **386**(6621): p. 194-200.
15. Charles M. Grinstead, J.L.S., *Introduction to Probability*. 1997: American Mathematical Society.
16. Unal, E.B., A. Gursoy, and B. Erman, *VitAL: Viterbi algorithm for de novo peptide design.* PLoS One, 2010. **5**(6): p. e10926.
17. Krogh, A., et al., *Hidden Markov models in computational biology. Applications to protein modeling.* J Mol Biol, 1994. **235**(5): p. 1501-31.
18. Sramek R., B.B., Vinar T.,, *On-line Viterbi Algorithm for Analysis of Long Biological Sequences*, in *Algorithms in Bioinformatics, 7th International Workshop, WABI 2007*. 2007, Springer: Philadelphia, PA, USA.
19. Mirabeau, O., et al., *Identification of novel peptide hormones in the human proteome by hidden Markov model screening.* Genome Res, 2007. **17**(3): p. 320-7.
20. Karplus, K., C. Barrett, and R. Hughey, *Hidden Markov models for detecting remote protein homologies.* Bioinformatics, 1998. **14**(10): p. 846-56.
21. Noguchi, H., et al., *Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules.* J Biosci Bioeng, 2002. **94**(3): p. 264-70.

22. Alonso, H., A.A. Bliznyuk, and J.E. Gready, *Combining docking and molecular dynamic simulations in drug design.* Med Res Rev, 2006. **26**(5): p. 531-68.
23. C.Jarzynski, *A nonequilibrium equality for free energy differences.* physical review letters, 1997. **78**: p. 2690-2693.
24. Boraschi, D., A. Tagliabue, and A.D. Miller, *The immunostimulatory effect of IL-1beta in vivo is blocked by antisense peptides complementary to the loop sequence 163-171.* FEBS Lett, 2009. **583**(4): p. 792-6.
25. Navikas, V. and H. Link, *Review: cytokines and the pathogenesis of multiple sclerosis.* J Neurosci Res, 1996. **45**(4): p. 322-33.
26. Sims, J.E. and D.E. Smith, *The IL-1 family: regulators of immunity.* Nat Rev Immunol, 2010. **10**(2): p. 89-102.
27. Auron, P.E., *The interleukin 1 receptor: ligand interactions and signal transduction.* Cytokine Growth Factor Rev, 1998. **9**(3-4): p. 221-37.
28. Veerapandian, B., *Structure and function of interleukin-1, based on crystallographic and modeling studies.* Biophys J, 1992. **62**(1): p. 112-5.
29. Priestle, J.P., H.P. Schar, and M.G. Grutter, *Crystal structure of the cytokine interleukin-1 beta.* EMBO J, 1988. **7**(2): p. 339-43.
30. *String Database*.
31. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nat Protoc, 2007. **2**(10): p. 2366-82.
32. Bin, W., et al., *IL-1beta enhances beta2-adrenergic receptor expression in human airway epithelial cells by activating PKC.* Am J Physiol Lung Cell Mol Physiol, 2001. **280**(4): p. L675-9.
33. Irmler, M., et al., *Granzyme A is an interleukin 1 beta-converting enzyme.* J Exp Med, 1995. **181**(5): p. 1917-22.
34. Coeshott, C., et al., *Converting enzyme-independent release of tumor necrosis factor alpha and IL-1beta from a stimulated human monocytic cell line in the presence of activated neutrophils or purified proteinase 3.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6261-6.
35. Borth, W. and T.A. Luger, *Identification of alpha 2-macroglobulin as a cytokine binding plasma protein. Binding of interleukin-1 beta to "F" alpha 2-macroglobulin.* J Biol Chem, 1989. **264**(10): p. 5818-25.
36. Ito, A., et al., *Degradation of interleukin 1beta by matrix metalloproteinases.* J Biol Chem, 1996. **271**(25): p. 14657-60.
37. Schonbeck, U., F. Mach, and P. Libby, *Generation of biologically active IL-1 beta by matrix metalloproteinases: a novel caspase-1-independent pathway of IL-1 beta processing.* J Immunol, 1998. **161**(7): p. 3340-6.
38. Kapur, V., et al., *Cleavage of interleukin 1 beta (IL-1 beta) precursor to produce active IL-1 beta by a conserved extracellular cysteine protease from Streptococcus pyogenes.* Proc Natl Acad Sci U S A, 1993. **90**(16): p. 7676-80.
39. Masters, S.L., et al., *Regulation of interleukin-1beta by interferon-gamma is species specific, limited by suppressor of cytokine signalling 1 and influences interleukin-17 production.* EMBO Rep, 2010. **11**(8): p. 640-6.
40. Adessi, C. and C. Soto, *Converting a peptide into a drug: strategies to improve stability and bioavailability.* Curr Med Chem, 2002. **9**(9): p. 963-78.
41. Hea, J.R., et al., *Mechanistic investigation into complementary (antisense) peptide mini-receptor inhibitors of cytokine interleukin-1.* Chembiochem, 2002. **3**(1): p. 76-85.
42. Heal, J.R., et al., *A search within the IL-1 type I receptor reveals a peptide with hydropathic complementarity to the IL-1beta trigger loop which binds to IL-1 and inhibits in vitro responses.* Mol Immunol, 1999. **36**(17): p. 1141-8.
43. G.M, M., *Automated docking using a Lamarckian genetic algorithm and an empricial binding free energy function.* Computational Chemistry, 1998. **19**: p. 1639.
44. Jones, G., P. Willett, and R.C. Glen, *Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation.* J Mol Biol, 1995. **245**(1): p. 43-53.

45. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein-ligand docking: current status and future challenges.* Proteins, 2006. **65**(1): p. 15-26.
46. Huey, R., et al., *A semiempirical free energy force field with charge-based desolvation.* J Comput Chem, 2007. **28**(6): p. 1145-52.
47. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD.* Proteins, 2003. **52**(4): p. 609-23.
48. *Hyperchem(TM) Professional 7.51*, Hypercube Inc: 1115 NW 4th Street, Gainesville, Florida 32601, USA.
49. in *DS Modelling*, Accelerys Inc.
50. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
51. Haliloglu, T. and B. Erman, *Analysis of correlations between energy and residue fluctuations in native proteins and determination of specific sites for binding.* Phys Rev Lett, 2009. **102**(8): p. 088103.
52. Budin, N., et al., *An evolutionary approach for structure-based design of natural and non-natural peptidic ligands.* Comb Chem High Throughput Screen, 2001. **4**(8): p. 661-73.
53. Lunter, G., *HMMoC--a compiler for hidden Markov models.* Bioinformatics, 2007. **23**(18): p. 2485-7.
54. Burge, C. and S. Karlin, *Prediction of complete gene structures in human genomic DNA.* J Mol Biol, 1997. **268**(1): p. 78-94.
55. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.
56. RABINER, L.R., *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition* PROCEEDINGS OF THE IEEE, 1989. **77**(2).
57. G. DAVID FORNEY, J., *The Viterbi Algorithm.* 1972.
58. ANDREW J. VITERBI, *Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm* 1966.
59. Brona Brejov´a, D.G.B.a.T.V., *The Most Probable Labeling Problem in HMMs and Its Application to Bioinformatics.* 2004: p. 426–437.
60. Fischer, B., et al., *NovoHMM: a hidden Markov model for de novo peptide sequencing.* Anal Chem, 2005. **77**(22): p. 7265-73.
61. PJ, F., *Statistical Mechanics of Chain Molecules.* 1969, New York: Wiley.
62. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* J Mol Biol, 1963. **7**: p. 95-9.
63. Keskin, O., et al., *Relationships between amino acid sequence and backbone torsion angle preferences.* Proteins, 2004. **55**(4): p. 992-8.
64. Unal, E.B., A. Gursoy, and B. Erman, *Conformational energies and entropies of peptides, and the peptide-protein binding problem.* Phys Biol, 2009. **6**(3): p. 036014.
65. Ewens WJ, G.G., *Statistical methods in bioinformatics: an introduction.* 2001, New York: Springer.
66. Patra, S.M., T. Bastug, and S. Kuyucak, *Binding of organic cations to gramicidin A channel studied with AutoDock and molecular dynamics simulations.* J Phys Chem B, 2007. **111**(38): p. 11303-11.
67. Bastug, T., et al., *Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling.* J Chem Phys, 2008. **128**(15): p. 155104.
68. Mai, B.K., M.H. Viet, and M.S. Li, *Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach.* J Chem Inf Model, 2010. **50**(12): p. 2236-47.
69. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-615.
70. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

71. D.Brown , J.H.R.C., *A comparison of constant energy,constant temperature and constant pressure ensembles in molecular dynamics simulation of atomic liquids.* molecular physics, 1984. **51**(5): p. 1243-1252.
72. MacKerell, A.D., Jr., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids.* Biopolymers, 2000. **56**(4): p. 257-65.
73. Zarate, H.H.a.J.M.O.d., *Jarzynski's equality illustrated by simple examples.* European Journal Of Physics, 2010. **31**: p. 1097-1106.
74. Shi, Y., *Mechanisms of caspase activation and inhibition during apoptosis.* Mol Cell, 2002. **9**(3): p. 459-70.
75. Changsik Kim, J.C., Seong Joon Lee, William J. Welsh and Sukjoon Yoon, *NetCSSP: web application for predicting chameleon sequences and amyloid fibril formation*. 2009, Nucleic Acids Research.
76. A.Bondi, *van der Waals Volumes and Radii.* Physical Chemistry, 1964. **68**(3).

VITA

Ece Bulut was born in Çanakkale, Turkey in 1987. She received her Bachelor of Science Degree in Chemical Engineering from Hacettepe University. In September 2009 she attended Chemical and Biological Engineering M.S. program at Koç University. From 2009 to 2011, she worked as a research and teaching assistant at the same institution.