

**Structure Based Drug Design Against
c-Rel Transcription Factor**

by

Mümin Öztürk

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

Master of Science

in

Chemical and Biological Engineering

Koc University

February 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

Mümin Öztürk

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:

Prof. Burak Erman

Prof. Özlem Keskin

Prof. Attila Gürsoy

Date:

ABSTRACT

NF- κ B signaling cascade plays a major role in many physiological processes such as immune and inflammatory responses, developmental processes, cellular growth, and apoptosis. The dysregulation of NF- κ B pathway and other signaling pathways that control its activity is a common feature of many diseases including cancer development and progression, cardiovascular and inflammatory diseases. c-Rel is one of the member of NF- κ B transcription factor family. It is shown that c-Rel proto-oncogene is the only NF- κ B family member that can malignantly transform cells in culture.

There have been various ongoing studies focused on blocking NF- κ B signaling by preventing binding of transcription factor to its target sites on DNA. In this study, we rather focus on termination of NF- κ B signaling. We propose a novel way of inhibiting NF- κ B action by removing it from DNA. c-Rel is chosen for this study because it is the most specialized member of NF- κ B family and termination of c-Rel response will have the least undesired side effects.

We first predict the structure of human c-Rel protein with a template based homology modeling server. The critical residues that will induce unbinding of c-Rel from DNA are determined by Gaussian Network Model. These residues are considered as drug interaction site and virtual screening studies is performed against c-Rel. Molecular modifications are applied to best hit compounds to increase the binding affinities. Potential drug molecules are then docked to other NF- κ B members to check specificity to c-Rel.

ÖZET

NF- κ B sinyal iletim yolađı bađışıklık ve enflamasyon tepkileri, embriyonik gelişim süreçleri, hücresel gelişim ve apoptoz gibi birçok fizyolojik süreçte önemli rol oynar. NF- κ B yolađının ve NF- κ B aktivitesini kontrol eden diđer yolakların fonksiyonlarının bozulması kanser oluşumunda ve yayılmasında, kardiyovasküler ve bađışıklık sistemi hastalıklarında görülen ortak bir özelliktir. c-Rel, NF- κ B transkripsiyon faktörü ailesinin bir üyesidir ve c-Rel proto-onkogeninin hücre kültürlerinin kötücül şekilde transformasyonunu sađlayan tek NF- κ B ailesi üyesi olduđu gösterilmiştir.

Transkripsiyon faktörün DNA üzerindeki hedef bölgeleri bađlanmasını engelleme amaçlı NF- κ B sinyal yolađının bloke edilmesi üzerine çok çeşitli çalışmalar yapılmaktadır. Bu çalışmada bu açıdan ziyade, NF- κ B sinyalinin sonlanması üzerine yoğunlaşıyoruz. DNA dan ayırarak NF- κ B etkisinin engellemesi üzerine yeni bir yol öneriyoruz. Bu çalışma için c-Rel seçildi, çünkü c-Rel NF- κ B ailesi içinde en özelleşmiş üye ve c-Rel etkisinin sonlandırılması aile içerisinde en az istenmeyen yan etkilere yol açacaktır.

İlk olarak, insan c-Rel proteinin yapısı homoloji modelleme sunucusu ile tahmin edildi. Gaussian Network Model kullanılarak c-Rel in DNA dan kopmasına yol açacak kritik amino asitler belirlendi. Bu amino asitler ilacın etkileşim yeri olarak kabul edildi ve bu kısım kullanılarak c-Rel e karşı ilaç tarama gerçekleştirildi. En iyi skor veren moleküllerin bađlanma afinitelerini daha da arttırmak için moleküler modifikasyonlar uygulandı. Belirlenen potansiyel ilaç molekülleri c-Rel e karşı özgülüğü test etmek için daha sonra diđer NF- κ B proteinlerine yanıştırıldı.

ACKNOWLEDGEMENTS

I would like to thank **Prof. Burak Erman** (my advisor) for giving me the chance of participating in an intriguing study. His guidance and support throughout thesis studies helped me overcome my indolence. It is an honor for me to graduate as one of his students. I would like to thank also **Prof. Özlem Keskin, Prof. Attila Gürsoy** and **Assistant Prof. Halil Kavaklı** for their patience and guidance. I shall extend my grateful thanks to all my friends in Graduate School of Sciences and Engineering. It has been such a great pleasure for me to spend more than two years at Koç University.

Last but not least, I would like to express my gratitude to my family for their endless love and support.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZET	iv
ACKNOWLEDGEMENTS	v
LIST of TABLES	viii
LIST of FIGURES.....	ix
INTRODUCTION.....	1
LITERATURE REVIEW	3
2.1 Introduction to NF- κ B	3
2.2 Canonical and Non-canonical NF- κ B Signaling.....	8
2.3 Termination of NF- κ B Response	13
2.4 Inhibitors of NF- κ B	14
COMPUTATIONAL TOOLS	18
3.1 Homology Modeling and Phyre Server	18
3.2 Molecular Dynamics	21
3.3 Gaussian Network Model	25
3.4. Molecular Docking	29
3.5 Autodock.....	32
3.6 Autodock Vina	37
RESULTS and DISCUSSIONS.....	41
4.1 c-Rel Transcription Factor:.....	41
4.2 c-Rel Crystal Structure	44
4.3 Homology modeling of human c-Rel.....	48
4.4 Energy Minimization of Modeled Structure	51
4.5 Drug Interaction Site Finding	54
4.6 Docking Studies	56
4.6.1 Zinc Fragment-Like Set Docking.....	56
4.6.2 Zinc Natural Products Docking.....	73

4.6.3 Amino Acid Derived Drug Discovery.....	81
CONCLUSIONS.....	87
APPENDIX A.....	89
Configuration File of NAMD Minimization.....	89
BIBLIOGRAPHY.....	91
VITA.....	102

LIST of TABLES

2.1	A list of genes induced by NF- κ B signaling system.....	4
2.2	Role of NF- κ B in different human ailments.....	5
4.1	Top ten docking results of fragment like ligands.....	59
4.2	Binding energies of top ten compounds over Rel family members.....	65
4.3	Best ten natural compounds bind to c-Rel.....	73
4.4	Binding energies of top ten natural compounds over Rel family.....	76
4.5	Free energy of binding values of best five amino acids.....	81

LIST of FIGURES

2.1	The domains of each NF- κ B family member.....	6
2.2	Canonical (classical) pathway and non-canonical (alternative) pathway	9
2.3	Structural domains for I κ B and IKK protein families.....	10
2.4	The inhibition steps of various inhibitors of NF- κ B signaling cascade.....	15
3.1	Steps in template based protein structure modeling.....	19
3.2	Lennard-Jones potential.....	23
3.3	Internal coordinates for bonded interactions.....	24
3.4	GNM description as a minimalist Elastic Network Model.....	26
3.5	Analysis of binding process in Autodock.....	36
4.1	c-Rel homodimer bound to DNA.....	45
4.2	Overall structure c-Rel-DNA complex.....	46
4.3	Pairwise alignment of chicken and human RHD of c-Rel.....	47
4.4	Schematic representations of modeled structures.....	48
4.5	Superimposed images of modeled polypeptides vs. chicken c-Rel RHD.....	50
4.6	RMSD graph of entire protein over the course of minimization.....	52
4.7	Superimposition of minimized modeled human c-Rel and chicken c-Rel.....	53
4.8	Energetically active residues on c-Rel.....	55
4.9	Surface representation of c-Rel bound to DNA.....	56
4.10	Grid box of receptor used during docking studies.....	58
4.11	Multiple sequence alignment of NF- κ B members.....	62
4.12	Grid boxes of p50 and p52.....	63
4.13	Grid boxes of RelA and RelB.....	64
4.14	Similar compounds of best docked ZINC00121681.....	66
4.15	Compound I and its docked position.....	67
4.16	Compound II and its docked pose.....	68

4.17	Superimposition of Vina and Autodock docked poses of compound II	70
4.18	Compound II and hydrogen bonding residues.....	71
4.19	CH--- Π interaction between Ala118 and compound II	72
4.20	C-H---O interaction c-Rel residues and compound II	72
4.21	Molecular structure of ZINC12659436 and testosterone.....	77
4.22	Docked pose of compound III	78
4.23	Superimposition of Vina and Autodock docked poses of compound III	79
4.24	Compound III and hydrogen bonding residues.....	80
4.25	C-H---O and C-H--- Π interactions c-Rel residues and compound III	80
4.26	2D structure of modified D-PHE and its docked figure.....	82
4.27	2D structure of compound V and its docked pose.....	83
4.28	Compound V and hydrogen bonding residues.....	84
4.29	Cation- Π interactions between compound V and positively charged residues.....	85
4.30	C-H---O and C-H--- Π interactions c-Rel residues and compound V	86
4.31	Superimposition of Autodock and Vina docked poses.....	86

Chapter 1

INTRODUCTION

NF- κ B family proteins are pleiotropic transcription factors that are structurally related and evolutionarily highly conserved. These transcription factors are so versatile that they exist practically in all cell types; however the most important and conserved role of NF- κ B is in immune system covering innate and adaptive immune responses, lymphogenesis and hematopoietic differentiation. The reach of NF- κ B is not confined in immune system; it extends to cell survival, differentiation and proliferation. The intriguing aspect of the pathway is how such a limited set of signaling NF- κ B proteins ultimately regulates transcriptional responses in diverse key cellular responses. As a consequence of these broad activities, it is not surprising that dysfunction of NF- κ B results in diverse pathologies. A thorough research is carrying on designing efficient therapeutical approaches in order to control and regulate NF- κ B activity for treatment of complex diseases. Since NF- κ B signaling is involved in variety of cellular processes, generic inhibition of NF- κ B likely leads to undesired side effects.

c-Rel is one of the five members of NF- κ B family and exists as homodimer or heterodimer with other members. While the various NF- κ B subunits have overlapping functions, c-Rel has some distinct roles in B-cell development, growth and survival. The specialized role of c-Rel and not involvement in many functions of NF- κ B master transcription factor make it good candidate for a drug target. In this thesis study, we approach c-Rel inhibition by inducing removal of bound transcription factor from DNA which will result in signaling termination. There is little known about the termination of NF- κ B signaling, so we propose a pocket that will induce a conformational change around

DNA binding residues when the drug molecule interacts with residues around the pocket. We screen fragment-like compounds, natural product and derivatives and D- and L-amino acids to find the best hits. These hits are later modified by combinatorial chemistry to improve the binding affinities.

Chapter 2 of thesis covers the general information about NF- κ B pathway; its elements, targets and signaling cascade. The limited known ways of termination of the response and inhibitors of the pathways are also included in this chapter.

Chapter 3 gives basic information about the subjects introduced in this work including homology modeling, molecular dynamics and molecular docking. The tools that are used, Phyre, Gaussian Network Model, Autodock, Autodock Vina, are described in detail in this chapter.

Chapter 4 starts with detailed information on c-Rel and its structure. Moreover, it includes the results of each step in this work and discussion of them. This thesis ends with a final chapter which includes conclusion of the study and future directions.

Chapter 2

LITERATURE REVIEW

2.1 Introduction to NF- κ B

Nuclear factor- κ B (NF- κ B) is a master transcription factor that plays important roles in inflammation, immunity, cell proliferation, cell development and cell survival. In 1986, David Baltimore and his colleagues discovered the first component of NF- κ B signaling system in enhancer regions of immunoglobulin κ light chains in activated B cells as a transcription-enhancing DNA binding complex [1]. The transcription factor has been one of the major research focuses in cell signaling since its discovery because of its diverse effects in cellular activities and broad spectrum of modulating stimuli. To emphasize the hotness of the topic, a PUBMED search results in about 32000 articles and more than 3000 reviews. After many years of intense research, it has been shown that the signaling system can be activated in almost all cell types and there is an extensive gene pool that has the specific NF- κ B binding site on the promoter or enhancer regions. So far, more than 400 genes have been found as NF- κ B target genes including cytokines, immune receptors, cell adhesion molecules, acute phase response proteins, regulators of apoptosis, growth factors, stress response genes and cell cycle regulators (Table 2.1). The pleiotropic transcription factor is linked to a large number of diseases ranging from cancer, diabetes, allergy, cardiovascular diseases, immune system disorders and ageing (Table 2.2).

NF- κ B is actually a collective term for a family of transcription factors. There are five members of the NF- κ B family in mammals: RelA (p65), RelB, c-Rel, p50 (NF- κ B1 and its precursor form is p105) and p52 (NF- κ B2 and its precursor form is p100) encoded by *rela*, *relb*, *rel*, *nfkb1* and *nfkb2* genes, respectively. These subunits of NF- κ B can homo-

or heterodimerize and form NF- κ B dimers; however not all possible dimeric combinations have been demonstrated. Theoretically, the members of NF- κ B family can form up to 15 unique dimers; but 12 of them are discovered in vivo. RelB:RelB, RelB:c-Rel and RelA:RelB are not known to exist in vivo up to now [3]. The NF- κ B transcription factor function only in dimeric form and the most abundant form is p50/p65 dimer. This heterodimer is constitutively expressed in almost all tissues and commonly referred to as NF- κ B protein. The prototypical NF- κ B is the best studied member of the family.

Table 2.1: The list of genes induced by NF- κ B signaling system

<i>Acute phase response proteins</i>	IL-1 α	<i>Growth factors</i>	<i>Viruses</i>	<i>Miscellaneous</i>
Angiotensinogen	IL-1 β	Bone morphogenic protein-2	Adenovirus (E3 region)	α -1 acid glycoprotein
β -defensin-2	IL-1 receptor antagonist	Granulocyte colony stimulating factor	Avian leukosis virus	Apolipoprotein C III
C4b binding protein	IL-2	Granulocyte macrophage colony stimulating factor	Bovine leukemia virus	AMH ^a
Complement factor B	IL-6	Erythropoietin, macrophage colony stimulating factor (M-CSF)	Cytomegalovirus	Cyclin D1
Complement factor C4	IL-8	Neurokinin-1 receptor	Epstein-Barr virus (Wp promoter)	Factor VIII
C-reactive protein	IL-9	Hepatocyte growth factor	Hepatitis B virus (pregenomic promoter)	Gadd45 β
Lipopolysaccharide binding protein	IL-10	Platelet-derived growth factor B chain	HIV-1	Galectin 3
Pentraxin PTX3	IL-11	Proenkephalin	HSV ^a	Epsilon-globin
SAA1 and SAA2 ^a	IL-12 (p40)	Vascular endothelial growth factor	JC virus	HMG-14 ^a
Tissue factor-1	IL-13		Human papillomavirus type 16	K3 keratin
Urokinase-type plasminogen activator	IL-15		SIV ^a	Laminin B2 chain
	β -Interferon		SV-40 ^a	Mts1
	IP-10 ^a	<i>Immunoreceptors</i>		MUC-2
	KC ^a	B7.1		Perforin
<i>Cell adhesion molecules</i>		BRL-1 ^a		Pregnancy-specific glycoprotein rnC ^a
E-selectin	ENA-78 (CXCL5)	CCR5		Prostate-specific antigen
Endoglin	GCP-2 (CXCL6)	CCR7	<i>Transcription/growth control factors</i>	S100A6 (calcyclin)
Fibronectin	Lymphotoxin α	CD137	A20	Syndecan-4
ICAM-1 ^a	Lymphotoxin β	CD154	Androgen receptor	Vimentin
MadCAM-1 ^a	MCP-1/JE ^a	CD40	c-myc	Wilms' tumor suppressor gene
P-selectin, tenascin-C	MIP-1 α β ^a	CD40 ligand	IRF-1 ^a	
VCAM-1 ^a	MIP-3 α	CD48	IRF-2	<i>Regulators of apoptosis</i>
DC-SIGN ^a	mob-1	CD83	IRF-4	TRAF-1 ^a
			IRF-7	TRAF-2 ^a
<i>Cell surface receptors</i>	Neutrophil activating peptide-78	Fc epsilon receptor II	Rel/NF- κ B proteins (p52/p100, p50/p105, c-Rel, and RelB)	IEX-11 ^a
RAGE — receptor for advanced glycation end products	RANTES ^a	IL-2 receptor α -chain	I κ B proteins (I κ B α , I κ B β , Bcl-3, JunB, Stat5a, WT1, p53, Ras)	Fas-ligand
Platelet activator receptor-1	TCA3 ^a	Immunoglobulin C γ 1		CD95 (Fas)
Neuropeptide Y Y1-receptor	TNF α	IgG γ 4		c-FLIP
Mu-opioid receptor	TNF β	Immunoglobulin epsilon heavy chain		Nr13
Mdr1 ^a	TRAIL ^a	Immunoglobulin k light chain		Caspase-11
Lox-1 ^a	THF3 ^a	Invariant chain II		Bcl-2
Gal1 receptor		MHC class I (H-2Kb)		Bcl-xL
CD69	<i>Enzymes</i>	MHC Class I HLA-B7		Bfl1/A1
Bradykinin B1-receptor	Liver alcohol dehydrogenase	β_2 -microglobulin		Survivin
	Collagenase 1	Nod2		
	Glutathione S-transferase	Polymeric Ig receptor		<i>Stress-response genes</i>
<i>Cytokines/chemokines</i>	Hyaluronan synthase	T-cell receptor β chain		Angiotensin II
CCL5 ^a	H+ -K+ ATPase α 2	T-cell receptor/CD3 γ		Cytochrome p450 gene
CCL15/leukotactin	Lysozyme	p80 TNF-receptor		COX-2 ^a
CCL22	Matrix metalloproteinase-9	Complement B		Ferritin H chain
CCL28	GD3-synthase	Complement component 3		12-Lipoxygenase
CINC-1 ^a	Gelatinase B	Complement receptor 2		iNOS ^a
CXCL 11 ^a	PIM-1	Proteasome subunit LMP2		Mn SOD ^a
Eotaxin	PKC δ	Peptide transporter TAP1		NQO1 ^a
Fractalkine	Phospholipase C δ 1	Tapasin		Phospholipase A2
Gro- α -g	Serpin 2A			
Gro-1	Transglutaminase			
ICOS ^a	TIRT ^a			
IFN-g				

For more information please visit <http://people.bu.edu/gilmore/nf-kb/target/index.html>

All subunits share an amino-terminal 300 amino acid long conserved domain, Rel homology domain (RHD), which is composed of two distinctly folded domains linked by a short polypeptide. The domain is responsible for nuclear localization, DNA binding, dimerization of subunits and also inhibition by I κ B (Inhibitor of κ B) proteins [4]. Crystal structures of NF- κ B dimers deposited to Protein Data Bank, c-Rel-c-Rel homodimer (PDB

Cancers	Cardiovascular diseases	Inflammatory diseases	Other diseases
Acute lymphoblastic leukemia	Atherosclerosis	Alzheimer's disease	Aging
Anaplastic large-cell lymphoma	Cardiac hypertrophy	Arthritis	AIDS
Breast	Heart failure	Asthma	Diabetes mellitus
Burkitt lymphoma	Hypercholesterolemia	Bone resorption	Ischemia
Cervical		Chronic obstructive pulmonary disease	Muscular dystrophy
Colorectal		Crohn's disease	Renal diseases
Diffuse large B-cell lymphoma		Multiple sclerosis	Sepsis
Fibrosarcoma		Ocular allergy	Skin diseases
Head and neck		Parkinson's disease	
Hodgkin's lymphoma			
Mammary carcinoma			
Mantle cell lymphoma			
Melanoma			
Multiple myeloma			
Lung			
Ovarian			
Pancreatic			
Prostate			
Squamous-cell carcinoma			
Thyroid			
Vulva			

Table 2.2: Role of NF- κ B in different human ailments [2]

code: 1GJI), p50-p50 homodimer (1NFK), p50-RelA heterodimer (1VKX), p50-RelB heterodimer (2V2T), p65-p65 homodimer (1RAM), relB-relB homodimer (1ZK9), reveals that amino-terminal part of RHD domain mediates binding to specific κ B sequences, while carboxy-terminal plays a main role for dimerization and interaction with I κ B proteins. NF- κ B specifically recognizes κ B DNA sequences; with a consensus sequence of 5'-GGGRNYYCC-3' (R is an unspecified purine -A or G-; Y is an unspecified pyrimidine -

C or T-; and N is any nucleotide). The family is also called the Rel protein family because of this conserved domain. Rel proteins can be divided according to their transactivation potentials. Unlike p50 and p52; RelA, RelB and c-Rel have transactivation domains (TAD), indicating that not all combinations of the Rel family are transcriptionally active (Figure 2.1). p50 and p52 homodimers repress κ B dependent genes by preventing those sites from binding of transcriptionally active dimers or recruiting histone deacetylases (HDAC) to promoter sites. It is shown that in unstimulated cells p50 homodimers are bound to κ B site containing promoters associated with HDAC-1 which mediates gene repression. Upon stimulation of cells by NF- κ B agonists, p65-p50 heterodimers enter the nucleus and replace p50 homodimers which have weaker affinity to κ B sites compared to p65/p50 heterodimers. Phosphorylated p65s interact with histone acetyltransferase (HAT) containing CBP/p300 coactivators. This mechanism not only explains the repression of NF- κ B target genes, also implicates acetylation/deacetylation gene regulatory processes in NF- κ B signaling [5].

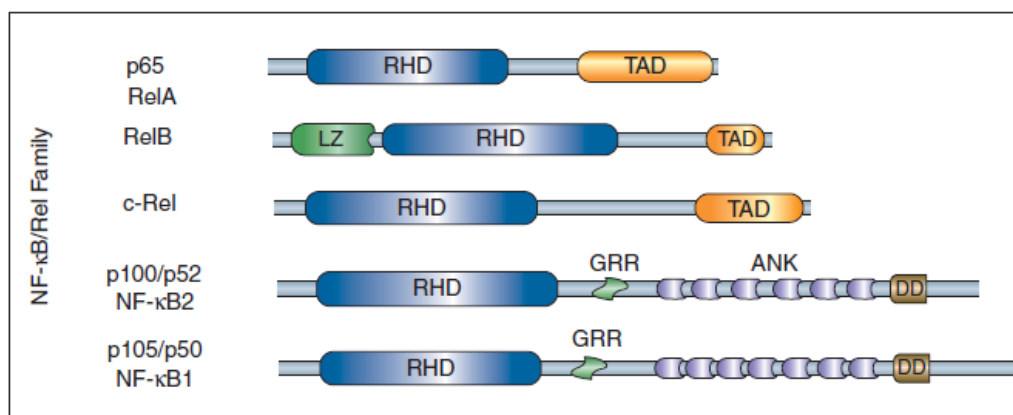


Figure 2.1: The domains of each NF- κ B family member are presented [3].

RelB has a unique feature that N terminal is also required for full transactivation of the transcription factor. Deletion studies showed that C terminal and N terminal of RelB

can show weak transcription activity independently, but the presence of both domains seriously enhances transcription [6]. Leucine zipper (LZ) domain presents this transactivation potential to N terminal. p100 (NF- κ B2) and p105 (NF- κ B1) are post-transcriptionally degraded from their C terminals to introduce p52 and p50; respectively. The cleavage site resides between glycine rich region (GRR) and ankyrin (ANK) repeats. GRR is a 23 amino acid length region (15 among of 23 are glycine) and functions as a degradation signal. Processing of p105 and p100 are mediated by proteasome in a unique manner. The substrates are subjected to incomplete degradation; N terminal portions escape degradation and eventually function as p50 and p52 [7]. Degraded segment includes several ankyrin repeats which are characteristic domains for I κ B proteins; that is why p100 and p105 are also considered as a member of I κ B protein family. p100 and p105 also contain the Death Domain (DD) in their outermost C terminal. DD is a protein-protein interaction motif and named after its presence in apoptotic proteins. Degradation of p105 to p50 occurs constitutively and DD is shown to be absolutely required for phosphorylation at Ser927 by IKK following recruitment of ubiquitin ligase complex for ubiquitinylation and eventually partial degradation by proteasome complex [8]. However; processing of p100 is tightly regulated, its phosphorylation and degradation is akin to I κ B. Interestingly, p100 is also found to be interacting with death receptors like TNF receptor 1 (TNFR1) and FAS via DD to induce apoptosis; on the contrary, NF- κ B is admitted as pro-survival factor [9].

NF- κ B family is functionally and structurally conserved from *Drosophila melanogaster* to *Homo sapiens*; with Relish, Dorsal and Dif proteins are taking role in fruit flies instead of Rel proteins in mammals. First identified protein in *Drosophila* Rel family was Dorsal which takes role in dorsal/ventral (back/abdomen) morphogenesis of fruit fly embryos. Dorsal-ventral polarity is accomplished by Dorsal transcription factor by turning gene expression on in one side and off in other side of the embryo. The activation signal to Dorsal passes through a transmembrane protein named Toll to a kinase named Pelle. In

unstimulated cells, Dorsal retained in cytosol by an inhibitor protein Cactus, homologous of I κ B, and activation of Dorsal requires release from Cactus. Phosphorylation of Cactus protein by Pelle kinase results in degradation of Cactus. The cytoplasmic domain of human IL-1 receptor shows high resemblance to Toll receptor and Pelle kinase is homologous to IRAK (IL-1 receptor associated kinase). The whole system is mostly similar to IL-1 activated NF- κ B modulation in mammalian cells. Dif shows more striking resemblance to mammalian system because it activates fly immune response genes, besides the same activating mechanism of Dorsal. The other member of *Drosophila* Rel family Relish protein is known to be the homologous of mammalian p105 and p100 proteins [10].

2.2 Canonical and Non-canonical NF- κ B Signaling

Since NF- κ B has an ability to influence expression levels of ever-increasing array of genetic targets, its activity is tightly regulated at multiple levels. The primary mechanism for regulating NF- κ B is through inhibitory I κ B proteins, and the kinase that phosphorylates I κ Bs, called the I κ B kinase (IKK) complex. A number of post-translational modifications also modulate the activity of I κ B and IKK proteins as well as NF- κ B molecules [11].

There are two major cascades have been demonstrated to activate NF- κ B transcription factor. In canonical pathway, activated IKK β phosphorylates canonical I κ Bs, I κ B α , I κ B β and I κ B ϵ . Diverse stimuli including tumor necrosis factor-alpha (TNF α), bacterial lipopolysaccharides, CD40 ligand, and several more cytokines recruit canonical signaling. Non-canonical NF- κ B activating stimuli are more restricted such as lymphotoxin β , B-cell activating factor (BAFF) and CD154. In non-canonical signaling, NF- κ B inducing kinase (NIK) activates IKK α which utilizes the precursor I κ B, p100. p100 is subsequently degraded to p52 that allows nuclear translocation of RelB-p52 dimers and binding to target gene promoters (Figure 2.2).

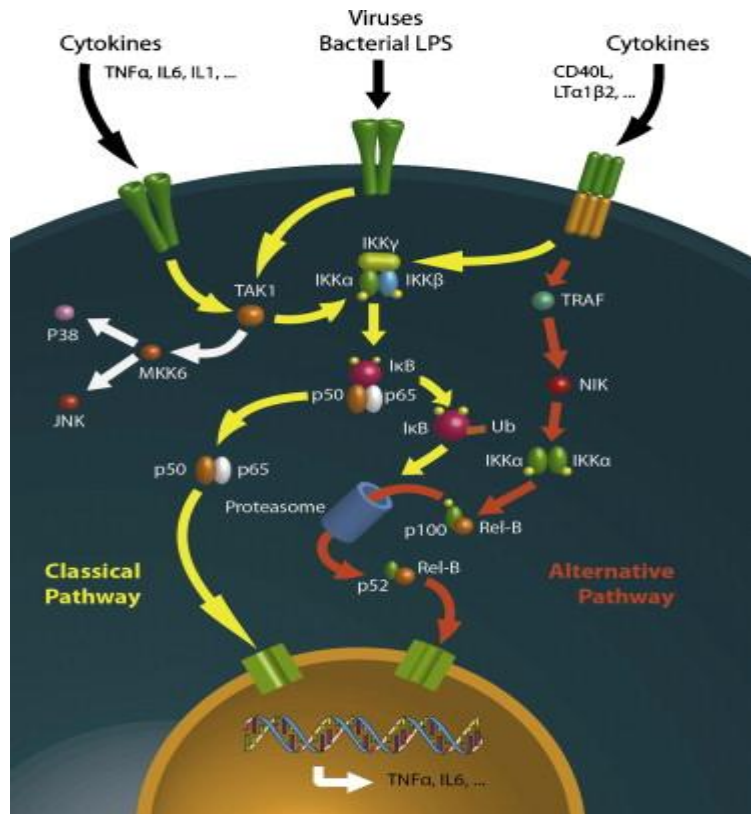


Figure 2.2: Canonical (classical) pathway is shown in yellow, non-canonical (alternative) is shown in red [12].

In the first years of NF- κ B research it was extensively believed that NF- κ B was expressed only in mature B cells and plasma cells. Gel-shift assays using κ B binding site in other cell types did not result in NF- κ B protein. It was later found that binding of NF- κ B was prevented by an inhibitor protein [13]. The interaction of the NF- κ B dimer with the I κ B protein generally sequesters the complex in the cytoplasm by masking the NLS (nuclear localization signal) sequence that resides on the N-terminal of NF- κ B subunits and blocks the dimer's ability to bind to DNA. Activation of the NF- κ B pathway by various inducers activates the IKK complex, which then phosphorylates I κ B.

Phosphorylated I κ B is then a preferred substrate for ubiquitination and degradation by the proteasome, which demasks the NLS sequence, frees the NF- κ B complex to enter the nucleus and eventually results in binding to target gene promoters or enhancers [14]. NF- κ B dimer is kept latent in cytosol by I κ Bs in most unstimulated cells; however it is discovered as constitutively active in mature B cells, macrophages, neurons and as well as numerous types range of cancer cells.

There are three types of I κ B protein; first group is called canonical or prototypical I κ Bs including I κ B α , I κ B β and I κ B ϵ , second group is called the precursor I κ Bs including p100 and p105, and the last group is called atypical I κ Bs including I κ B ζ and Bcl-3 which regulate the activity of NF- κ B dimers in nucleus. The I κ B family is characterized by five to

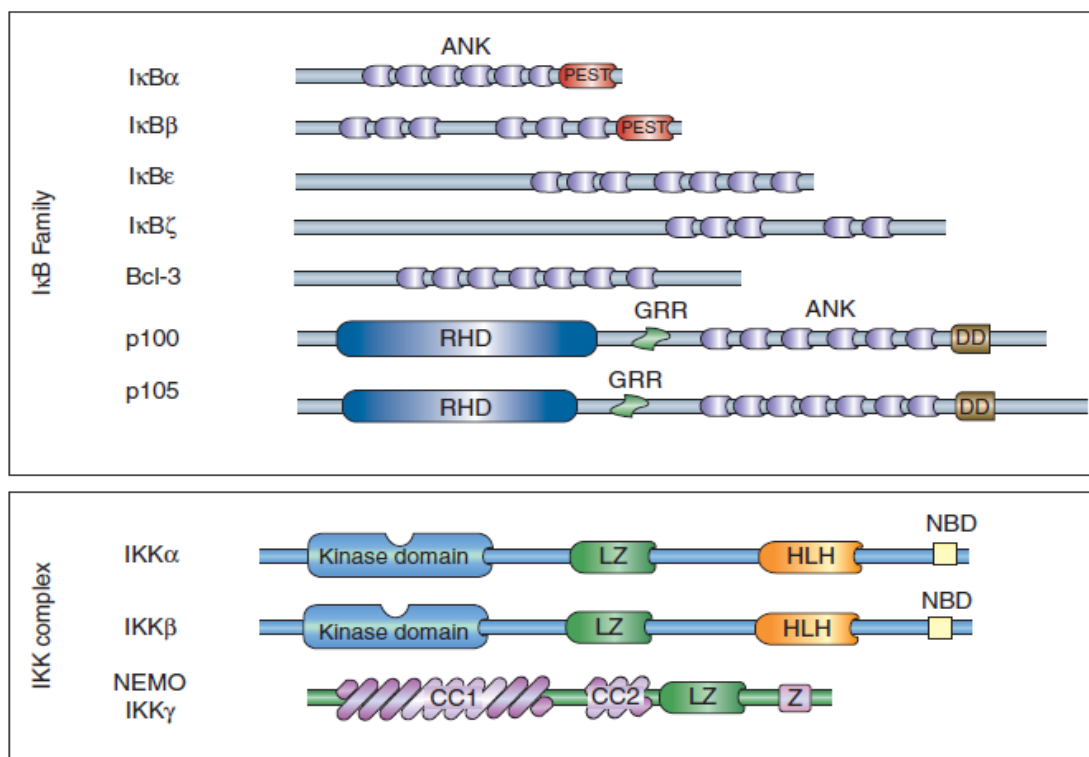


Figure 2.3: Structural domains are represented for I κ B and IKK protein families [3].

seven ankyrin repeat motifs (Figure 2.3). Ankyrin repeats are 33-residue length and one of the most common protein-protein interaction motifs. They are involved in the interaction with RHD of NF- κ B members. I κ B α and I κ B β also contain PEST region in their C terminal. PEST region is rich in amino acids proline (P), glutamic acid (E), serine (S) and threonine (T) and commonly found in rapidly degraded proteins. It has been shown that there are consensus casein kinase II phosphorylation sites in PEST region and it is required for degradation of free I κ B α [15]. Individual I κ B proteins show preference to a particular subset of NF- κ B dimers. Classical I κ B proteins are primarily believed to inhibit dimers that possess at least one p65 or c-Rel subunit. While nonclassical I κ Bs associate with all NF- κ B subunits; RelB exclusively binds p100. Nuclear or atypical I κ Bs prefer p50 and p52 homodimers [14, 16].

Prototypical I κ Bs have N-terminal signal responsive regions that contain conserved serine residues. In canonical signaling, IKK β subunit of IKK complex phosphorylates serine 32 and serine 36 residues on I κ B α . Phosphorylated I κ B α later undergoes lysine 48 linked-polyubiquitinylation by SCF (Skp1-Culin-F-box) complex containing β -TRCP (β -transducin repeat containing protein), a multi protein E3 ubiquitin ligase complex. Ubiquitin is a small protein, 76 amino acid length, is covalently attached to target protein in an ATP dependent fashion. Ubiquitinylation is one of the post-transcriptional modifications and plays key role in protein degradation, protein localization and cell-cycle regulation. The ubiquitinylation process is performed by sequential enzymatic reactions of three proteins. E1 ubiquitin-activating enzyme hydrolyses ATP, E2 ubiquitin-conjugating enzyme catalyzes ubiquitin attachment to the substrate protein and E3 ubiquitin ligase is known to have role in recognizing substrate and mediates the conjugation of ubiquitin in a concerted manner with E2. After first ubiquitin is linked to substrate, further ubiquitins can be attached to the first added moiety generating polyubiquitin chains. There are seven lysine residues in ubiquitin protein that can act in ubiquitinylation; however K48 linked

polyubiquitylation is the most prominent one. SCF^{β-TRCP} recognizes DS^PGXXS^P motif on phosphorylated N-terminal of IκBα and catalyzes degradative polyubiquitylation which results in degradation by 26S proteasome complex. NLS is exposed in NF-κB dimers subsequently and nuclear translocation is induced by karyopherin proteins [17].

The nuclear or atypical IκBs are the least studied ones and their mechanism of action are not fully understood. Bcl-3 has a TAD and interacts with p50 and p52 containing homo- or heterodimers. First studies done on Bcl-3 showed that it interacts and removes transcriptionally inactive or repressive p50 and p52 homodimers from κB sites. Nonetheless, further studies suggested that it can also provide transcriptional activation function to p50 and p52 homodimers by interacting with them. These different actions of Bcl-3 on promoting transcription have been shown to be dependent on different phosphorylation states but upstream elements remain unknown in this cascade [18]. Interestingly, it has been also proved that Bcl-3 can inhibit NF-κB activation by stabilizing repressing dimers at κB sites, thus preventing transcriptionally active NF-κB dimers' access and ensuing transcription [19]. IκBζ is also shown to form stable ternary complexes with p50 homodimers like Bcl-3. The other common feature of nuclear IκBs is having a NF-κB driven expression pattern.

Most of the diverse cell stimuli that activate NF-κB firstly converge on IKK complex. The complex consists of two catalytic subunits, IKKα and IKKβ, and one regulatory scaffold subunit, NEMO (NF-κB essential modifier) also called as IKKγ. Both IKK enzymes have N terminal serine/threonine kinase domain, a leucine zipper (LZ) domain and a C terminal helix-loop-helix (HLH) domain (Figure 2.3). Dimerization of catalytic subunits is required for kinase activity and LZ domain is responsible for dimerization. HLH is reported essential for optimal kinase activity and intramolecular interaction with kinase domain was suggested to regulate post-inductive downregulation of

kinase activity. The enzymes also contain carboxy terminal NEMO binding domain (NBD) that is critical for their interaction with regulatory IKK γ . NEMO is a highly conserved protein and consists of coiled coil regions (CC), leucine zipper and C terminal zinc finger (Z) motif [20].

2.3 Termination of NF- κ B Response

There exists limited information in literature about termination of NF- κ B transcription. Appropriate temporal activity is required to prevent deleterious overinduction effects and maintain cell's response to succeeding stimuli. Negative feedback regulation is generally accepted mechanism but how NF- κ B is removed from DNA remains unknown. In this mode of regulation, the expression of canonical I κ Bs is induced after NF- κ B enters the nucleus. Newly generated I κ Bs can enter the nucleus and pull the active dimer away from the κ B sequences back to cytoplasm. I κ B α can provide fast turn-off mechanism; whereas I κ B β and I κ B ϵ have slower kinetics [21].

Second downregulatory mechanism involves proteasomal degradation of bound p65 subunit. PDLIM2 is a nuclear ubiquitin ligase and has been shown to get involved in degradative ubiquitinylation of p65; however whether the degradation process occurs in cytosol or nucleus is unknown. PDLIM2 also targets ubiquitinated p65 dimers to PML nuclear bodies that the nuclear degradation is suggested to happen in these nuclear compartments [22]. Another ubiquitin ligase complex has been also identified in p65 ubiquitinylation, ECS complex (Elongins B and C, Cul2 and SOCS1). COMMD1 (copper metabolism Murr1 domain 1) plays a nonenzymatic role in this mechanism and promotes the association between p65 and SOCS1 [23]. Moreover, IKK α has been demonstrated to phosphorylate C- terminals of p65 and c-Rel to promote rapid turnover and clearance from κ B sites in macrophages. It has been suggested that IKK β mediated canonical signaling

involves I κ B-based negative feedback but this mechanism may not be sufficient or fast enough to ensure transient nature NF- κ B response; so IKK α -dependent termination mode is evolved [24]. There are probably more specific E3 ubiquitin ligases that target other NF- κ B subunits or other mechanisms to provide negative control on NF- κ B signaling that need to be identified.

2.4 Inhibitors of NF- κ B

Hyperactivation of NF- κ B is frequently seen in human diseases. NF- κ B is constitutively activated in breast cancer, esophageal cancer, hepatocellular carcinoma, Hodgkin lymphoma, lung cancer, ovarian cancer, prostate cancer, autoimmune diseases such as arthritis, multiple sclerosis, various skin diseases and neurodegenerative disorders such as Alzheimer's disease. Given that NF- κ B has a central role in many ailments; immense research has been conducted to develop strategies that block NF- κ B signaling. More than 750 compounds including natural products, proteins, peptides, natural compounds, antisense RNAs and synthetic compounds have been reported to inhibit NF- κ B activation [25]. Approach to inhibition of NF- κ B signal transduction pathway can be classified into three steps (Figure 2.4):

- a. Blockage of signaling upstream of IKK
- b. Blockage at IKK or I κ B degradation state
- c. Prevention of nuclear activity of NF- κ B

Cytokines are polypeptide regulators used extensively in cellular communication in the immune system. Many cytokines signal via NF- κ B cascade; e.g. TNF- α , TNF- β , IL-1. Anti-cytokine therapy has been shown to suppress NF- κ B activity in patients with rheumatoid arthritis (RA) and Crohn's disease (CD), i.e. inflammatory bowel disease. Infliximab (trade name Remicade) is a monoclonal antibody against TNF- α , blocks

interaction with TNF- α receptor, is approved for treatment of RA and CD. Etanercept (Enbrel) is an immunoadhesin, an engineered fusion protein comprised of a binding domain and an immunoglobulin constant domain, consists of TNF receptor II fused with constant domain of IgG1 heavy chain. It works as a decoy receptor that binds to TNF- α and competes with cell surface TNF- α receptors. Anakinra (Kineret) is an IL-1 receptor antagonist that blocks the IL-1 receptor activity by competitive inhibition and approved by the Food and Drug Administration (FDA) for the treatment of RA [26].

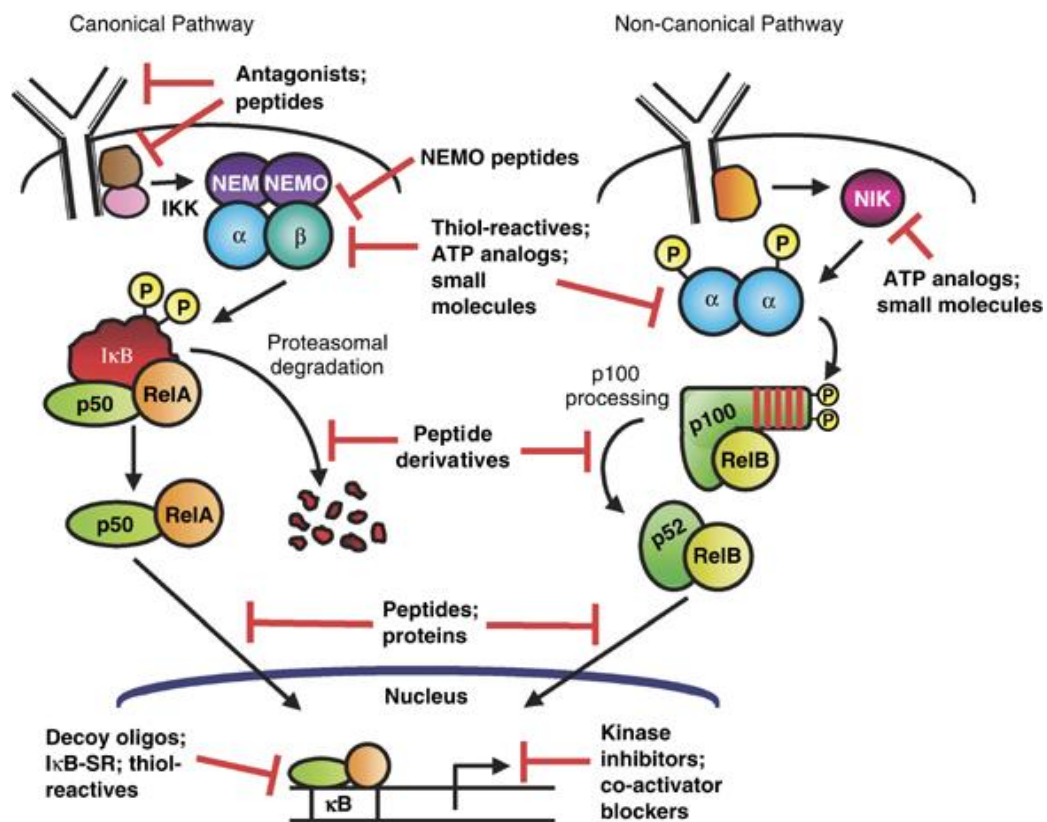


Figure 2.4: The inhibition steps of various inhibitors of NF- κ B signaling cascade [25].

Upstream elements of IKK can also be the target of NF- κ B signal transduction, but they contribute to various signaling pathways and inhibition may result in non-specific

blocking. Phosphoinositide 3-kinase (PI3K) and Akt are upstream elements of IKK α that wortmannin and LLY294002 for PI3K and potent Akt inhibitors may dampen NF- κ B signaling [27]. IKK has been a main target in suppressing NF- κ B activation because of its convergent role on diverse stimuli and latest successes on other kinase inhibitors in drug discovery. ATP binding site competitors such as aspirin, sodium salicylate, SPC839 and SC-514 effect on ATP binding pocket of IKK β ; by that means have a diminishing effect on phosphorylation of I κ Bs [28]. BMS-345541 is a highly selective inhibitor binds to non-equivalent allosteric sites on IKK α and IKK β but IKK β shows 10-fold greater inhibitory effect [29]. The association between NEMO and IKK α/β has also been targeted and researchers come up with a decoy peptide that is cell-permeable and spanning NBD of IKK β . It is shown that NBD peptide effectively blocks IKK β -NEMO interaction thereby TNF- α induced NF- κ B activation [30].

Maintaining a high level of I κ B concentration in the cytoplasm can also be the goal to inhibit NF- κ B activation. One approach is blocking the ubiquitinylation or promoting deubiquitinylation of I κ Bs. Bacterial protein YopJ is shown to be involved in reversing ubiquitinylation of phosphorylated I κ B α [31]. Ro106-9920 was reported to inhibit E3 ubiquitin ligase activity and as a consequence block NF- κ B activity [32]. There are various proteasome inhibitors that block the 26S proteasome which degrades ubiquitin-tagged I κ Bs. Proteasome inhibitors have drawn attention because of their potential role in antitumor activity. Lactacystine, synthetic peptide aldehydes, e.g. PS-341, MG-132 are well-known peptide inhibitors. Bortezomib (PS-341) is a FDA approved proteasome inhibitor administered against multiple myeloma. It has been showed that PS-341 causes perturbation in NF- κ B signaling in colorectal cancer cells. MG-132 also caused malfunction in NF- κ B signal activation in Hodgkin's lymphoma cells [33].

Direct inhibition of NF- κ B could be achieved by concentrating on blocking either nuclear translocation, DNA binding or transactivation of the dimers. SN50 peptide is designed to contain NLS of p50 and a cell permeable motif. Delivery of engineered peptide to cells inhibits nuclear import of NF- κ B and it is shown to have therapeutic effects on acute pancreatitis [34]. DHMEQ is also reported to block nuclear translocation of freed NF- κ B dimers in canonical and non-canonical signaling and has anti-inflammatory and antitumor activity in various carcinomas without any toxicity [35]. Sesquiterpene lactones (SL) are a diverse group of natural products that are isolated from medicinal plants. It has been shown that many SLs can work as a sensitization agent to overcome chemoresistance of tumor cells, anti-metastasis agent, and anti-inflammatory agent by directly inhibiting NF- κ B pathway at multiple steps. One of the acting mechanisms is found to be inhibiting DNA-binding feature of p65. Parthenolide is thought to modify p65 at Cys38 residue in the DNA-binding loop via alkylation [36]. Usage of decoy oligodeoxynucleotides (ODN) spanning κ B sites can result in binding of corresponding NF- κ B dimers, thereby preventing nuclear trafficking of those dimers and blocking specific NF- κ B-target gene binding [37]. The final step in NF- κ B cascade that could be blocked is transactivation potential of Rel family members. D609, phosphatidylcholine-specific phospholipase C inhibitor, RO31-8220, protein kinase C inhibitor, SB203580, p38 mitogen-activated protein kinase inhibitor, are shown to inhibit transactivation ability, but the mechanisms are unknown [38].

Chapter 3

COMPUTATIONAL TOOLS

3.1 Homology Modeling and Phyre Server

There are nearly 12 million non-redundant protein sequences included in January 2011 update of The Reference Sequence database (<http://www.ncbi.nlm.nih.gov/RefSeq/>); however about 68000 protein structures are deposited to Protein Data Bank, at present. The significant discrepancy between number of structures and sequences draws attention to tertiary structure prediction from amino acid sequence in computational biology area. A biennial, international blind structure prediction experiments Critical Assessment of Structure Prediction (CASP), first launched in 1994, has attracted even more researchers in structure prediction from amino acid sequence [39]. The prediction methods can be grouped into three categories, namely, *ab initio* or *de novo* modeling, threading or fold recognition and homology modeling. *Ab initio* modeling starts from scratch and based on the assumption that the folded –native- state is at the global free energy minimum. The method simulates folding process using a Monte Carlo optimization based on physicochemical or knowledge-based statistical potentials established upon contact potentials, solvation potentials, Van der Waals interactions, attractive forces and hydrogen bond potentials. Unlike template based modeling, homology modeling and threading, *ab initio* modeling is not limited to known structure and helps to understand how proteins fold in nature. The ideal approach would be using physics based potentials and solving Newton's equation of motion at each step in folding process, but there is a limitation in computing space for normal size of protein. Hence, knowledge-based approaches are

progressing as an alternative and there are successful medium accuracy models for proteins up to 120 amino acids [40].

Although there are increasing numbers of protein sequences, unique folds in nature appear to be limited and also structures of proteins are evolutionarily more conserved than their sequences. These features and the fact that homologous proteins fold into very similar structures lead to homology modeling. However; the range of homology modeling is relatively narrow, because of relying on the presence of homologous template structures [41]. Threading or fold recognition is applied to proteins that have similar folds but evolutionarily distant from each other. Target sequence is aligned with a library of 3D folds and the most compatible structure is assigned by using energy potentials and other similarity scoring methods [42].

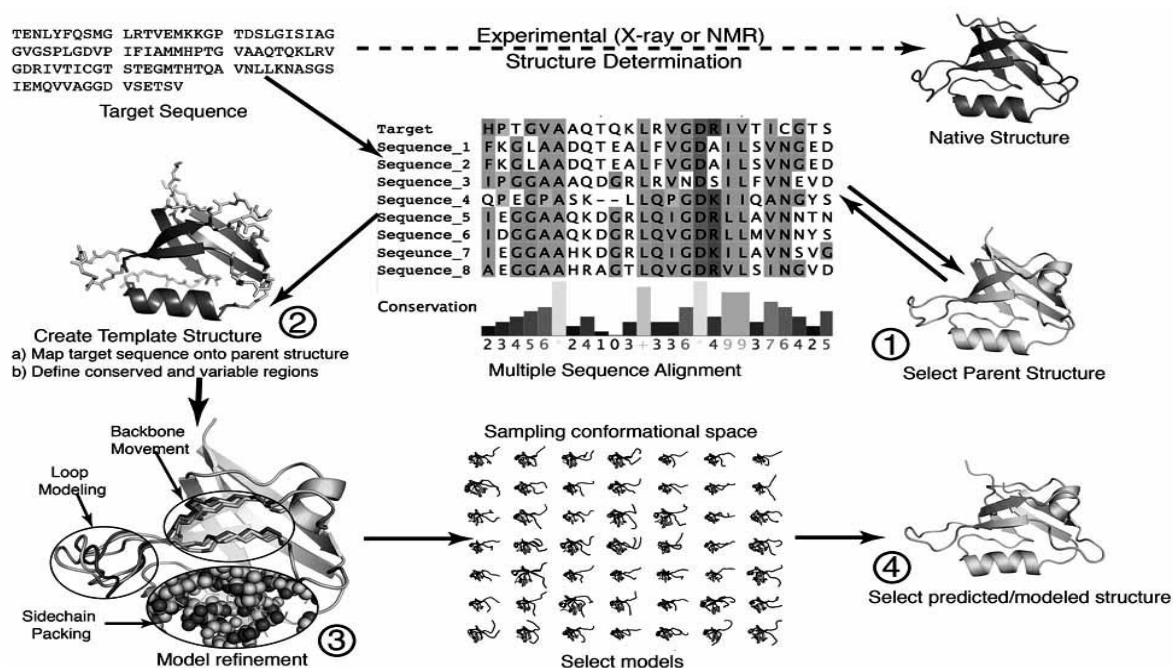


Figure 3.1: Steps in template based protein structure modeling [43]

Advancements in modeling studies have made threading and homology modeling to integrate and they are called as template based modeling (TBM) or comparative modeling, nowadays. The first step in TBM is the identification of the parent structure, which must have been solved before and should be deposited to PDB, with a sequence alignment (Figure 3.1). Proteins with similar sequence show most likely similar folds and this principle is a good start for assigning the parent structure. Once starting parent structures identified, the next step is classifying conserved and variable regions by aligning target sequence with parent structure and building template structure suitably. Third step involves the refinement process through backbone movement, loop modeling and side chain packing. The aim of backbone movement is sampling a backbone conformational space that allows side chain packing but restrains movements to guide the protein to fold into a specific structure. Loop modeling is predicting the structure of the backbone between two ends of termini inside the polypeptide. There are two different methods in loop building, *de novo* approach and database sampling. *De novo* method generates many random samples fitting the geometric constraints by applying a scoring or energy function. Database sampling fits the geometric constraints via conformations from a library of loop fragments. Final phase of refinement, side chain packing uses rotamer libraries that store the possibility of each amino acids rotamers and requires scoring or energy function like loop modeling to check the final samples. During refinement step, thousands of potential models are created that are subjected to final evaluation step to find the most nearest structure to native state [43].

Phyre (Protein Homology/Analogy Recognition Engine) is a widely used server for protein structure prediction and one of the best servers according to recent CASP experiment, CASP8 [44]. Phyre uses a fold library that encompasses about 16500 nonredundant protein sequences mostly taken from Structural Classification of Proteins (SCOP) and enhanced with newer depositions in PDB [45]. When the query is submitted to

the server, five iterations of position specific iterative BLAST (PSI-BLAST) is performed against fold library and a profile is constructed. PSI-BLAST constructs position specific scoring matrix (PSSM) which lists the frequencies of amino acids at specific residue, as a consequence of multiple sequence alignment. Using first profile, search is repeated against entire database and the profile is refined. The iterative searching approach is more sensitive in finding evolutionarily distant homologs. After the profile is constructed, Phyre uses three independent secondary structure prediction algorithms; Psi-Pred, SSPro and JNet [46-48]. These algorithms give confidence values at each residue for three states: alpha helix, beta strand and coil and Phyre calculates the consensus prediction by averaging these confidence values. Phyre also utilizes Disopred program to calculate structurally ordered and disordered regions [49]. The profile and secondary structure prediction are scanned against the fold library by the profile-profile alignment algorithm and search returns 3D structures of the top ten highest scoring alignments. The gaps in the alignment are modeled by using a loop library and reconstruction procedure. The side chains are placed via rotamer library and a fast graph-based algorithm [50].

3.2 Molecular Dynamics

Molecular dynamics (MD) is a computer simulation which presents a system of methods on the molecular scale for microscopic modeling. The simulations compute change of positions, velocities and orientations with time. MD simulations have played pivotal roles in theoretical study of biological molecules and provided comprehensive information on the fluctuations and conformational changes of proteins, nucleic acids, etc. The connection between microscopic information and macroscopic properties such as pressure, energy, heat capacities is achieved via statistical mechanics. Statistical mechanics is fundamental to the study of biological systems by MD simulations.

The protein's 3D geometry might be sufficient for some of the tasks; nevertheless, for most of the functions internal dynamics of the protein is intimately linked. If the researchers try to fully understand protein function, they first need to apprehend its dynamics. Currently, there are no experimental procedures applicable to scrutinize dynamics at the atomic resolution during physiologically relevant time scale ranging from picoseconds to seconds, hence MD simulations are employed to simulate protein dynamics.

Proteins show an extensive range of time scales over which specific processes occur. Local motions (atomic fluctuations, side chain motions and loop motions) range from 0.01 to 5 Å in a timeframe of 10^{-15} to 10^{-1} s whereas rigid body motions (domain motions, helix motions and subunit motions) range from 1 to 10 Å in a timeframe of 10^{-9} to 1s. Large-scale motions (helix coil transitions, dissociation/association processes and folding/unfolding) larger than 5 Å range from 10^{-7} to 10^4 s. The computational power is the decisive force for the design of molecular dynamics simulations. The size of the simulation (number of atoms), the duration and time step should be adjusted in order to make the simulation last for a reasonable time. Time step is the time length between evaluations of the force field and typical time steps in MD simulations are in the order of 1 femtosecond (10^{-15} s). It should also be noted that the calculations must be long enough to study the natural processes of interest. Most of the simulations that are mentioned in scientific publications span about nanoseconds or microseconds that takes several CPU-days to CPU-years to bring the calculation to completion. The first study on protein simulations was done on bovine pancreatic trypsin inhibitor in 1977 and consisted of 9000 steps corresponding to 8.8 ps [51].

The basic idea behind molecular dynamics is the step-by-step solution of the classical equations of motion $\left(m_i \ddot{r}_i = f_i \quad f_i = -\frac{\partial U}{\partial r_i} \right)$. r_i represents complete set of $3N$

atomic coordinates and the forces acting, f_i , are usually derived from potential energy, U s. Non-bonded interactions constitute some part of the energy function where the formula is:

$$U_{non-bonded}(r^N) = \sum_i u(r_i) + \sum_i \sum_{j>i} v(r_i, r_j) + \dots$$

The formula is split into one-body, two-body, three-body, etc. terms; but usually three-body and higher order terms are neglected. One-body term signifies the effects of container walls; however it is generally not taken into consideration because of fully periodic simulations of the bulk system. Non-bonded interactions are including Coulomb interactions and van der Waals interactions which are approximated by Lennard-Jones 6-12 potential; $v^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$ where r is the distance between particles, σ and ϵ are specific parameters, first is determining the position of potential well and the latter represents the well's depth [52]. Potential well is defined as region enclosing a local minimum of potential energy (Figure 3.2).

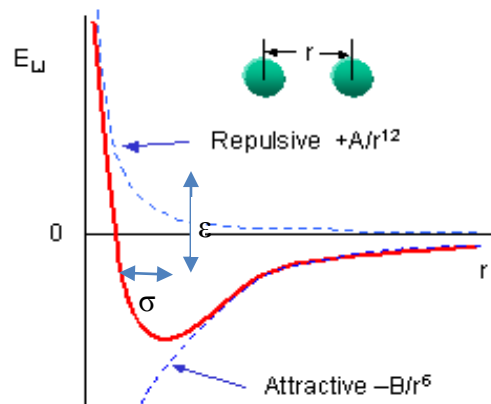
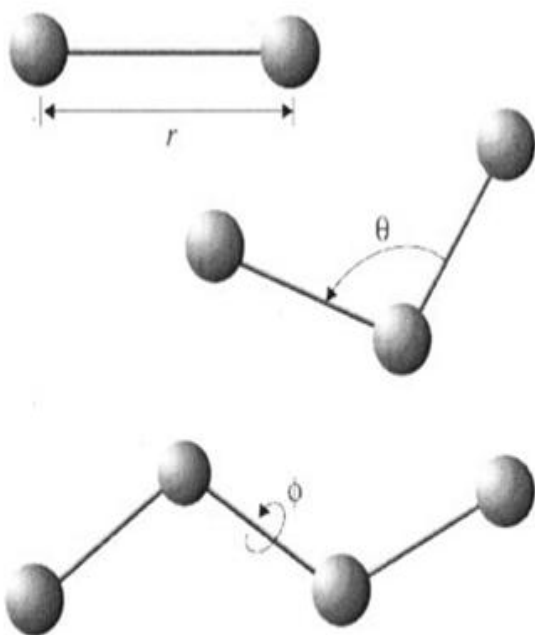


Figure 3.2: Lennard-Jones potential

If electrostatic interactions are present, Coulomb potentials should be added that $v^{Coulomb}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r}$ where Q_1 and Q_2 are charges and ϵ_0 is the vacuum permittivity. During simulations, intramolecular bonding interactions including, stretching, bending and torsional bonded interactions must also be considered (Figure 3.3). r represents the bond

stretching and the potential counts each covalent bond in the system;



$$U_{bond} = \sum_{bonds\ i} k_i^{bond} (r_i - r_{0i})^2$$

θ governs the bond angle term which is the angle between each subsequent pair of covalent bonds residing the shared atom at the vertex.

$$U_{angle} = \sum_{angle\ i} k_i^{angle} (\theta_i - \theta_{0i})^2$$

Φ shows the dihedral or torsion angle that describes the angle occurs between two planes defined by four atoms [53].

Figure 3.3: Internal coordinates for bonded interactions

$$U_{dihedral} = \sum_{dihedral\ i} k_i^{dihedral} [1 + \cos(n_i\phi_i - \gamma_i)]$$

where n represents the periodicity of torsional potential which is periodic through 360° rotation and γ is the reference torsional angle which is mostly 0° or 180° .

There are many MD softwares available. One of the most popular of them, NAMD is a parallel molecular dynamics program that is advantageous for studying large biomolecular systems. It is widely used because of its performance on parallel environment and free of charge distribution. It works with AMBER, CHARMM, X-PLOR and GROMACS input files, parameters and potential functions.

3.3 Gaussian Network Model

The structural architecture of biomolecules determines the intrinsic dynamics and the dynamics affect the conformational mechanisms, most strikingly the enzyme catalysis which is significantly based on transitions from ground states to higher-energy states of both enzyme and substrate. A better understanding of intrinsic dynamics can be achieved through MD simulations and normal mode analysis (NMA). The basic approach for these methods relies upon the idea that low frequency or large amplitude nodes are related to the function. These atomistic methods might have computationally limited applicability for larger systems; so coarse-grained models have emerged. Gaussian Network Model (GNM) is one of the simpler coarse grained-NMA methods utilizing an elastic network model. GNM was firstly proposed in 1997 by Bahar *et al.* by inspiring Flory's previous work on polymer gels [54, 55]. The fluctuation dynamics of proteins are determined firstly by assigning α -Carbon atoms of each residue to the nodes of the elastic network. The nodes are linked by harmonic springs between sufficiently close residue pairs (Figure 3.4.b). GNM provides information about the magnitudes of fluctuations but not directionality; since GNM assumes all the fluctuations are implicitly isotropic. The fluctuations are generally anisotropic in reality; so instead of having $3N-6$ independent modes for 3D structures, GNM is based on $N-1$ modes with N representing each residue on protein. Anisotropic network model (ANM) was developed as an extension of GNM that also includes fluctuation vectors to give idea about mechanism of motion [56].

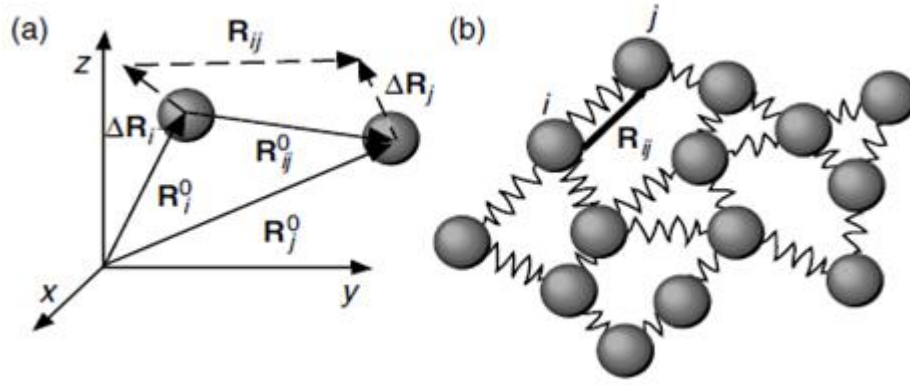


Figure 3.4: GNM description as a minimalist Elastic Network Model [57]

The first step to study dynamics of the elastic network; the equilibrium position of node i , is defined by R_i^0 and the fluctuation from this mean position is defined by ΔR_i where $\Delta R_i = R_i - R_i^0$. Vector R_i gives the instantaneous position of the node. The fluctuations of distance vector between nodes i and j can be expressed as $\Delta R_{ij} = R_{ij} - R_{ij}^0 = \Delta R_j - \Delta R_i$ (Figure 3.4.a). GNM applies Kirchoff matrix to express the connectivity of each residue or node of protein. Kirchoff or Laplacian matrix is used in graph theory to obtain a matrix representation of graph. Kirchoff matrix Γ is defined by,

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_{cutoff} \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_{cutoff} \\ \sum_{i,i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases}$$

r_{cutoff} is the cutoff distance that defines the neighborhood, generally between 6.5-7 Å for proteins. The i th diagonal element of Kirchoff matrix gives the coordination number,

number of its neighbor residues, of i th residue. By considering Hooke's law on springs, potential on the network can be expressed in terms of ΔX_i , ΔY_i and ΔZ_i components of ΔR_i

$$V_{GNM} = \frac{\gamma}{2} \left[\sum_{i,j}^N \Gamma_{ij} \left[(\Delta X_i - \Delta X_j)^2 + (\Delta Y_i - \Delta Y_j)^2 + (\Delta Z_i - \Delta Z_j)^2 \right] \right]$$

γ is force constant which is uniform for all the springs in network. If X , Y and Z components of displacement vectors ΔR_i are expressed as $\Delta X = [\Delta X_1, \Delta X_2, \dots, \Delta X_N]^T$, $\Delta Y = [\Delta Y_1, \Delta Y_2, \dots, \Delta Y_N]^T$, $\Delta Z = [\Delta Z_1, \Delta Z_2, \dots, \Delta Z_N]^T$ the equation above simplifies to

$$V_{GNM} = \frac{\gamma}{2} [\Delta X^T \Gamma \Delta X + \Delta Y^T \Gamma \Delta Y + \Delta Z^T \Gamma \Delta Z]$$

The probability distributions of all fluctuations are Gaussian; so this assumption leads us to

$$p(\Delta X) \propto \exp \left\{ -\frac{\gamma}{2k_B T} \Delta X^T \Gamma \Delta X \right\} \propto \exp \left\{ -\frac{1}{2} \left(\Delta X^T \left(\frac{k_B T}{\gamma} \Gamma^{-1} \right)^{-1} \Delta X \right) \right\}$$

where k_B is Boltzmann constant and T is absolute temperature. Similar forms apply to $p(\Delta y)$ and $p(\Delta z)$. The normalized probability distribution is

$$p(\Delta X) = \frac{1}{Z_x} \exp \left\{ -\frac{1}{2} \left(\Delta X^T \left(\frac{k_B T}{\gamma} \Gamma^{-1} \right)^{-1} \Delta X \right) \right\}$$

Z_x is the partition function given by;

$$Z_x = \int \exp \left\{ -\frac{1}{2} \left(\Delta X^T \left(\frac{k_B T}{\gamma} \Gamma^{-1} \right)^{-1} \Delta X \right) \right\} d\Delta x = (2\pi)^{N/2} \left| \frac{k_B T}{\gamma} \Gamma^{-1} \right|^{1/2}$$

$$NxN \text{ covariance matrix of } \Delta X \text{ is } \langle \Delta X \Delta X^T \rangle = \int \Delta X \Delta X^T p(\Delta X) d\Delta X = \frac{k_B T}{\gamma} \Gamma^{-1}.$$

Since the assumption of isotropic fluctuations requires

$$\langle \Delta X \Delta X^T \rangle = \langle \Delta Y \Delta Y^T \rangle = \langle \Delta Z \Delta Z^T \rangle = \frac{1}{3} \langle \Delta R \Delta R^T \rangle$$

The residue fluctuations are calculated as $\langle \Delta R_i^2 \rangle = \frac{3k_B T}{\gamma} (\Gamma^{-1})_{ii}$ and correlations $\langle \Delta R_i \Delta R_j \rangle = \frac{3k_B T}{\gamma} (\Gamma^{-1})_{ij}$

When singular value decomposition is applied to Kirchoff matrix, $\Gamma = U \Lambda U^T$, U gives eigenvectors of Γ , Λ is a diagonal matrix with eigenvalues of Γ and lastly U^T appears to be a unitary matrix where $U^T = U^{-1}$. Eigenvalues represent the frequencies of individual modes as eigenvectors define the shape of modes. If we combine above correlation calculation with SVD of Kirchoff matrix;

$$\langle \Delta R_i \Delta R_j \rangle = \frac{3k_B T}{\gamma} (\Gamma^{-1})_{ij} = \frac{3k_B T}{\gamma} (U \Lambda^{-1} U^T)_{ij} = \frac{3k_B T}{\gamma} \sum_k [\lambda_k^{-1} u_k u_k^T]_{ij}$$

The correlation contributed by k th mode can be found as [58];

$$\langle \Delta R_i \Delta R_j \rangle_k = \frac{3k_B T}{\gamma} \lambda_k^{-1} [u_k]_i [u_k]_j$$

As a consequence, slow modes contribute to mechanism of global motions and the most constrained residues in these modes play more critical role in the movement such as hinge-bending center. The fastest modes correspond to small amplitude fluctuations and shows rapid, local fluctuations for higher scoring residues. It is also shown that GNM can also be utilized in prediction of binding sites. Anchor residues on ligands are strongly anticorrelated with rest of the residues of ligand in slowest modes and performing large scale fluctuations to efficiently probe its environment. Binding groove residues on receptor are most excited ones by perturbation at fast modes and they transmit the perturbation to rest of the protein acting as “energy gates” [59].

3.4. Molecular Docking

Molecular docking is a structural molecular biology field that has been there for three decades of vigorous study. The striking advances in power of computation, growth in size of protein and small molecule databases and ease of access to those databases have made docking easier and alluring. The goal of molecular docking is to predict binding modes structurally and binding affinity energetically. The docking is generally carried out between a small molecule (ligand) and a target molecule (protein, DNA or RNA); but protein-protein docking is also arousing interest. Molecular docking is widely used in various applications in drug discovery including structure-activity relationships (SAR), lead discovery by virtual screening, lead optimization, aiding X-ray crystallography for fitting of substrates, chemical mechanism studies and combinatorial library design.

Protein-ligand docking studies started as regarding both protein and ligand as rigid bodies and progressed to a stage where full flexibility on the ligand is employed. The next step and major challenge in docking nowadays is handling receptor flexibility effectively. The intrinsic mobility of proteins has frequently been ignored but it is known that proteins are in constant cycling between different conformational states and protein flexibility allows increased affinity between the drug and its target. X-ray crystallography studies of protein-ligand complexes reveal that 70-100% surface areas of ligands are buried that can be achieved as an outcome of receptor flexibility [60].

Docking methods are a combination of a search algorithm and scoring function. There are relatively numerous and ever increasing number of search algorithms and scoring functions available. The goal of search algorithm is finding true binding modes by sampling the degrees of freedom of protein-ligand complex sufficiently. Speed and effectiveness are the critical factors in search algorithm for covering the pertinent

conformational space. Scoring function deals with the thermodynamics of binding, characterizes the true binding modes and ranks them accordingly. Finding the correct pose is the ultimate goal in docking algorithms. The success of the program is generally measured in terms of root mean square deviation (RMSD) of heavy atom (omitting hydrogen atoms) positions between experimentally proved and predicted binding poses. The flexibility of the system drastically increases degree of freedom which in turn affects the searching efficiency. The system includes at least the macromolecule or receptor, the ligand and the solvent molecules; however solvent molecules are either excluded from the calculations or implicitly modeled in scoring function to address the solvent effect because of huge number of degrees of freedom [61].

There are two main methods to treat ligand flexibility; systematic methods and stochastic methods. Systematic search methods attempt to investigate all degrees of freedom in ligand and can be categorized further into three types. Conformational search method systematically rotates all rotatable bonds through 360° using a fixed increment till all the possible conformations are covered. This method yields a huge number of structures with increasing number of rotatable bonds, thus needs to be limited with some constraints on the ligand. Fragmentation is one of the commonly utilized methods in ligand docking algorithms. Two approaches are used in fragmentation; first one docks the several fragments of the ligand and links them to create the initial ligand (place and join approach) and the other one firstly docks the rigid core fragment of ligand, then increment the fragment size by adding the flexible regions of ligand and redocks into the active site (incremental approach). LUDI and FlexX are popular docking programs that use fragmentation [62, 63]. Third method is the database method that library of pregenerated conformational ensembles are used. FLOG applies the database method with using 25 different conformations from library per molecule on the basis of distance geometry [64].

Stochastic methods, also called random search algorithms, sample the conformational space by making random changes to ligand or to a pool of ligands at each step. The random change performed is accepted or rejected according to a predefined probability function. Monte Carlo methods, genetic algorithm methods and tabu search methods are three basic types of stochastic algorithms. In Monte Carlo method, the acceptance criterion of a random change is calculated by using the Boltzmann probability function. ICM and DockVision are some of the docking programs that use Monte Carlo method [65, 66]. Genetic algorithm will be mentioned in detail in Autodock topic; but in a nutshell it uses the idea from evolutionary process in biological systems. The most popular programs in molecular docking, Autodock and GOLD, implements genetic algorithm [67, 68]. Tabu search works on a principle that prevents search from previously explored areas of conformational space and promotes the analysis of new regions. The random change is rejected if RMSD of new conformation is less than a cutoff value. PRO_LEADS is the program that uses tabu search algorithm [69].

The accuracy of docking is dictated by scoring functions which distinguishes true binding mode from all other modes generated. Speed and accuracy are two important aspects in scoring algorithms; but some simplifications are needed to be introduced in order to reduce the complexity. There are three major types of scoring: force-field based, empirical and knowledge-based. Force-field scoring functions approach calculation of binding free energy by summing two independent energies; interaction energy between receptor and ligand, the internal energy of ligand. They are based on decomposition into individual interaction terms, such as van der Waals energies, electrostatic energies, bond stretching, torsion and bending by applying CHARMM, AMBER or GROMOS force fields [70-72]. The major challenge in force field scoring is including solvent effect and entropic effect. Empirical scoring functions try to calculate free energy of binding by summing up weighted and uncorrelated empirical energy terms including desolvation term, entropy term

and hydrophobicity term. Experimentally determined binding energies are used to determine the coefficients of weighted energy terms by means of a regression analysis. Knowledge-based scoring functions also consider experimentally determined structures; but they are based on converting structural information into free energies instead of dealing with free energies directly. These functions screen compound databases and form potentials according to different atom-atom pair contacts and other typical interactions.

There is growing number of molecular docking tools uses different ligand sampling methods, scoring functions and deals with protein flexibility issue in various manners. Comparison studies for docking might be useful for determining the accuracy and speed; but there are also docking benchmarks available; for example DUD, to evaluate docking algorithms [73].

3.5 Autodock

Autodock is a well-known docking software that models ligand-receptor binding interactions. The program was first released in 1990 and now with many developments version 4.2 is in use [74]. The algorithm is based on a few approximations for the purpose of predicting the conformation and binding free energy of a docking simulation reliable and fast as much as possible. The ligand in simulation is treated as partly flexible that only torsional degrees of freedom are analyzed, bond angles and bond lengths are remained constant unlike typical molecular mechanics methods. These exclusions lead to rapid transformations of coordinations, but it might be an issue if binding requires significant distortion of ligand. Autodock uses a tree-like representation for ligand flexibility. The fixed portion of ligand is defined as “root”, and branches that define rotatable bonds sprout from the root. There can be zero or more branches and can be nested. The idea of “torsion

tree” does not allow flexibility in cyclic molecules; however with indirect methods ring flexibility is manageable.

Autodock applies grid-based method to allow fast evaluation of ligand-protein interaction. AutoGrid calculates grid of interaction energy around the macromolecular target prior to docking simulation. A three dimensional grid surrounding the coordinates of receptor, given by the user, is constructed and a probe atom is placed at each grid point. The interaction energy between probe atom and the target grid point is computed and stored. Separate maps are constructed for each atom type in the ligand including additional electrostatic and desolvation maps. These maps are later used during docking simulation as a look-up table enabling rapid calculation. Grid based computation requires the receptor molecule rigid; if user chooses specific side chains to be flexible, those parts are evaluated explicitly outside the grid.

There are three main conformational search methods available in Autodock that are simulated annealing, traditional genetic algorithm and Lamarckian genetic algorithm (LGA). Simulated annealing (SA) was the primary method in earlier versions; whereas LGA is shown to be most efficient and reliable search method [75]. SA is based on random walk of the ligand around static protein. A random change is applied to each degrees of freedom of ligand including ligand’s position, orientation and conformation. Energy of the new configuration is calculated and accepted if the energy is lower than the previous configuration. If the energy of new state is higher, acceptance depends on probability $P(\Delta E) = e^{\left(\frac{-\Delta E}{k_B T}\right)}$ where, ΔE is the energy difference between two states, k_B is the Boltzmann constant and T is user defined temperature. The acceptance criterion is known as Metropolis criterion which is first introduced in a seminal paper in 1953 [76]. After a specific number of acceptances or rejections under the current temperature, a new cycle begins in which temperature is decreased by $T_i = gT_{i-1}$ (g is constant between 0 and 1, T_i

is temperature at cycle i) and the initial configuration of new cycle is the minimum energy state of previous cycle. Generally, at high temperatures many of the states are accepted; but with decreasing temperatures the probabilistic selection will be more stringent and majority of the moves will be rejected.

Genetic algorithm approaches molecular docking problem by utilizing the concept of genetics and natural evaluation. The ligand's conformation is codified into "chromosomes" which is composed of real-valued "genes"; three genes accounting for Cartesian coordinates, four genes defining a quaternion specifying ligand orientation and one gene for each ligand torsion. Genetic algorithm is initiated by creating a population of random individuals that are enclosed within a user specified box. Random values between minimum and maximum of the search area are assigned to genes holding Cartesian coordinates; random quaternion consisting of a random unit vector and random rotation angle between -180° and 180° is given to orientation genes and finally random torsion angles, if any, between -180° and 180° are given. The next step is mapping where each "genotype", gene values, is translated into corresponding "phenotype", all atom coordinates, and eventually evaluating the "fitness" of individual; in other words total interaction energy between ligand and protein. Selection procedure determines the individuals that will reproduce in the next generation by the formula $n_0 = \frac{f_w - f_i}{f_w - \langle f \rangle} f_w \neq$

$\langle f \rangle$ where; n_0 represents number of offspring used in the next generation, f_i is fitness of the individual, $\langle f \rangle$ is the mean fitness of population and f_w is the worst fitness in the population; i.e. highest energy. When f_i is lower than $\langle f \rangle$ (free energy of binding of i th individual is lower than mean energy of the population), the value of n_0 will be larger than 1 which means that individual will be represented in the next generation. When $f_w = \langle f \rangle$, the algorithm stops because it is assumed that the population is converged. Once the individuals are selected for next generation; two-point "crossover" is performed.

Crossover swaps equivalent genes of randomly picked chromosomes and original chromosomes are replaced with new offsprings keeping the population size constant. “Mutation” is also used on randomly picked genes that changes translational, orientational and/or torsional values of genes. The fitnesses of new individuals are computed and top genotypes are selected for the next cycle. Genetic algorithm iterates until total number of energy evaluations or generations are reached to user defined values.

LGA is a hybrid method that utilizes local search methods with genetic algorithm. The theory is based on Lamarck’s assertion about evolution that some phenotypic traits acquired can effect on individual’s genotype and become heritable. LGA performs a local search on phenotype of the individual and if it is successful it is translated into corresponding genotype with inverse mapping. Local search is applied to user defined fraction of the population with user defined translational, orientation and torsional step sizes. New conformation’s energy is evaluated and if it is lower than precursor’s energy, local search iterates until energy increases or specific number of steps are performed [75].

Autodock4 applies a semiempirical force field that combines molecular mechanics force fields with empirical weights. Free energy of binding is calculated by finding intramolecular energy changes in ligand and protein through binding and intermolecular energy changes between ligand and protein through binding (Figure 3.5).

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

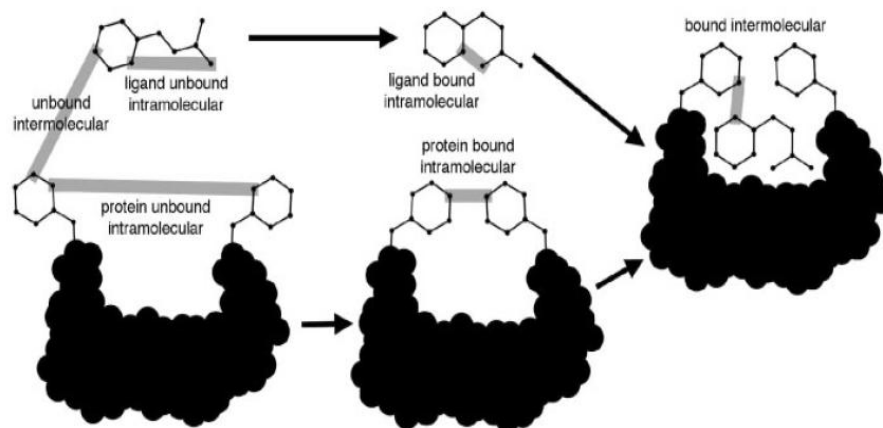


Figure 3.5: Analysis of binding process in Autodock [77]

In these pairwise evaluations (V), L refers to ligand and P refers to protein. It is assumed that ligand and receptor molecules are sufficiently distant from each other that $V_{unbound}^{P-L}$ is zero. If rigid protein docking is the issue, there will not be any motion in the protein; so V_{bound}^{P-P} will be equal to $V_{unbound}^{P-P}$. The pairwise evaluations are calculated as:

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{(r_{ij}^2 / 2\sigma^2)}$$

Weight constants (W) are based on linear regression analysis of experimentally characterized complexes and can be changed in parameter file. The first term in molecular mechanics terms is 6-12 van der Waals potential and the constants A and B are taken from Amber force field [70]. Hydrogen bonding term is 10-12 potential with a directional weight $E(t)$ which is dependent on angle t between probe atom and target atom. Electrostatic interactions are calculated through a screened Coulomb potential. The interactions between

charges are dampened or screened by the presence of water molecules and ions. A distance dependent dielectric function $\varepsilon(r_{ij})$ is used based on the work of Mehler and Solmajer [78]. Desolvation free energy term is expressed in terms of volume of the atoms surrounding a given atom V and solvation parameter S . The exponential term includes the distance between ligand and protein's atoms, r_{ij} . σ is the Gaussian distance constant and taken 3.5Å which corresponds roughly to minimum van der Waals potential for two heavy atoms [79]. Lastly, it is assumed that the term for loss of torsional entropy due to restriction of conformational degrees of freedom, ΔS_{conf} , is directly proportional to number of rotatable bonds in ligand that $\Delta S_{conf} = W_{conf}N_{tors}$.

3.6 Autodock Vina

Autodock Vina is new docking software that offers high performance, ease of use and enhanced accuracy compared to Autodock 4. The program is developed in the same group that is also working on Autodock. There have been two recent studies that comparing Autodock and Autodock Vina (hereafter referred to as Vina) and Vina showed more promising and much faster docking results [80,81].

Vina uses machine-learning type of scoring function and gradient-based local search; whereas Autodock applies more physics based scoring function and stochastic local search utilizing genetic algorithm. The scoring function of Vina consists of conformation-dependent and conformation-independent parts. The conformation-dependent part is the sum of intermolecular and intramolecular contributions and these contributions are calculated as $c = \sum_{i < j} f_{t_i t_j}(r_{ij})$ where $f_{t_i t_j}$ is the interaction function on atom types t_i and t_j and r_{ij} is the interatomic distance. Vina implements its own scoring function that is mostly similar to X-score (formerly known as X-CSCORE). X-score is a consensus score

of three different scores. These scores contain common van der Waals interaction, hydrogen bonding and deformation effect (proportional to number of rotatable bonds) terms but they differ from each other in handling hydrophobicity [82]. X-score neglects intramolecular contributions, but ignoring them sometimes results in clashed structures for search algorithms. Thus, Vina also considers the intramolecular terms. Atom classifications are also similar to X-score; that there are five different types used. H-bond donor atoms are oxygen and nitrogen atoms bonded to hydrogen atoms and metal ions reside in binding site. H-acceptor atoms are oxygen and sp^2 or sp hybridized nitrogen atoms with lone pairs. H-bond donor/acceptor atoms are oxygen and nitrogen atoms which can perform either H-donor or H-acceptor roles such as the oxygen atom in hydroxyl group. Fourth group is polar atoms which are oxygen and nitrogen neither H-bond donor nor H-bond acceptor, sulfur and phosphorous atoms, carbon atoms linked to heteroatoms. The last group is hydrophobic atoms which are halogen atoms and carbon atoms that are not polar. In addition, only heavy atoms contribute in scoring function; hydrogen atoms are not considered explicitly.

The interaction function $f_{t_i t_j}$ actually depends on surface distance rather than interatomic distance. Surface distance is $d_{ij} = r_{ij} - R_{t_i} - R_{t_j}$ where R_{t_i} is the van der Waals radius of atom type t_i . Vina scoring function can also be interpreted as $f_{t_i t_j}(r_{ij}) = h_{t_i t_j}(d_{ij})$ that $h_{t_i t_j}$ is the weighted sum of hydrogen bonding between H-bond donor/acceptor atoms, hydrophobic interactions between hydrophobic atoms and steric interactions between all types of atoms. The hydrogen bonding term is equal to 1 when $d < -0.7\text{\AA}$, 0 when $d > 0$ and linearly interpolated between these d values. The hydrophobic term is equal to 1 when $d < 0.5\text{\AA}$, 0 when $d > 1.5\text{\AA}$ and linearly interpolated between these d values. The steric terms consist of three individual terms:

$$gauss_1(d) = e^{-(d/0.5\text{\AA})^2} \quad gauss_2(d) = e^{-((d-3\text{\AA})/2\text{\AA})^2} \quad repulsion(d) = \begin{cases} d^2, & \text{if } d < 0 \\ 0, & \text{if } d \geq 0 \end{cases}$$

As a limitation value, all the calculations are cut off when surface distance is larger than 8Å. The conformation-independent part of Vina scoring function corresponds to deformation effect in X-score and it is calculated as $g(c_{inter}) = \frac{c_{inter}}{1+wN_{rot}}$ where c_{inter} is intermolecular contribution, w is associated weight and N_{rot} is number of rotatable bonds.

Autodock Vina utilizes iterated local search rather than simulated annealing and genetic algorithm in Autodock as the metaheuristic to search global minimum free energy of binding. The method is an iterative improvement algorithm which is restarting the local optimization by perturbing (mutating) the current solution when the optimization is trapped in local minima. Vina implements the algorithm that uses a succession of steps involves local optimization and mutation. After each step, the Metropolis method is applied as an acceptance criterion like in simulated annealing. Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method is used in the local optimization process.

Quasi-Newton methods are widely used in optimization to find the local maxima and minima of the functions and based on Newton's method. Newton's method in numerical analysis, also known as Newton-Raphson method, is applied for finding the roots of a function. Root finding process is initiated with guessing in the vicinity of suspected root and defining recursively the algorithm $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ until the best approximated root is found. The same approach can also be used in finding local minima and maxima of function in optimization processes. The gradient or the first derivative of function is zero at local extremum points so the algorithm is modified to $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$. Quasi Newton methods do not need Hessian matrix or second derivative of the function, they need only the gradients of objective function to be supplied at each iterate.

They measure the changes in gradients of the objective function and construct a model that is good enough to produce superlinear convergence to local minimum; that is why they are more efficient than Newton's method and BFGS is considered as most effective algorithm for solving large scale optimization problems [83]. Using the gradient based method also gives physical interpretation on ligand. The function to be minimized is energy function and the derivative with respect to position gives clue on total acting force on ligand, derivative with respect to orientation is related to total torque. It may seem that calculating both the energy function and the gradients might be time-consuming; but this approach seems to accelerate optimization process significantly [80].

Chapter 4

RESULTS and DISCUSSIONS

4.1 c-Rel Transcription Factor:

c-Rel is one of the members of Rel family and responsible in controlling B cell proliferation, cell survival and oncogenesis. It is highly homologous to p50 and p65 subunits which both of them form the ubiquitous NF- κ B transcription factor. It mostly exists as in homodimer forms or in heterodimers with p50, and rarely with p65. It was discovered as a human homolog of v-Rel oncogene from avian reticuloendotheliosis virus (REV-T). c-Rel is the only member that can malignantly transform cells among Rel family; thus it is also considered as a proto-oncogene. The first discovery on the role of c-Rel proto-oncogene on oncogenesis was transformation of B and T cells and induction of lymphoma which is the cancer in person's lymph system, or in other words immune system. *REL* (gene name of c-Rel protein) expression is particularly considered in hematopoietic tissues, blood cells forming tissue, but high levels *REL* transcripts are also detected in intestine and lung. Lower levels are detected in heart, testis and kidney tissues and no *REL* transcript is detected in liver tissue [84].

There have been various *REL* deletion and mutation studies on mice. c-Rel null mice develop normally proving that it is not essential for embryonic development. Normal number of hematopoietic cells in knockout (null) mice shows that c-Rel is dispensable for differentiation of hematopoietic precursors; but mice show severely impaired lymphocyte proliferation and function associated with production of diverse cytokines and immune modulatory proteins [85].

Human *REL* gene is located on 2nd chromosome p12-13 region. The region including *REL* gene (2p12-16) is found to be amplified at least four copies in 23% of the cases of diffuse large B-cell lymphomas (DLBL) which is an aggressive type of lymphoma and constitute approximately 50% of B-cell non-Hodgkin's lymphomas, 30% of all types of lymphomas. The characteristic of DLBL, as its name implies, larger than normal B lymphocytes which stopped responding to cell signals that limit the growth and reproduction of cells. The pattern of cancerous lymphoid cells within the lymph node is diffused rather than forming follicles. DLBL is rapidly fatal if it remains untreated. The primary treatment for DLBL patients is R-CHOP chemotherapy. R-CHOP stands for **R**ituximab, a monoclonal antibody that binds to CD20 protein on B cells and destroys both normal and malignant B cells, **C**yclophosphamide, an alkylating agent that adds alkyl groups to guanine base and causes crosslinks in DNA resulting in interfering with its action, **H**ydroxydaunorubicin, a drug that intercalates DNA and prevents its replication, **O**ncovin, a mitotic inhibitor that disrupts assembly of microtubules and arrests dividing cells in metaphase, and **P**rednisone, a steroidal drug that acts as an immunosuppressant. The early stages of DLBL patients, if the disease is limited to only one lymph node, can also benefit from radiotherapy. R-CHOP treatment might sometimes develop serious side effects including fever, febrile neutropenia which is the abnormal low count of neutrophils –immune cells that first respond to infection- and tumor lysis syndrome that is caused by release of toxic break-down products of dying cancer cells. The chance of survival with chemotherapy is lower in older patients. There are also examples of relapsed and refractory patients of lymphoma. Researchers are still trying to increase survival rate of this most common subtype of lymphoma with clinical trials [86].

REL is also shown to be frequently amplified in extranodal DLBLs, lymphomas that arise in non-lymphoid sites such as gastrointestinal tract and skin [87]. In one subtype of DLBL, primary mediastinal B-cell lymphoma which arises in thymus –T lymphocyte

maturation organ-, *REL* locus amplification is found to be a five to tenfold but not with a high incidence, two out of 26 cases. The suspected mechanism is the act of c-Rel transcription factor on B-cell differentiation factors such as Ig κ light chain, class I MHC H-2k and β -microglobulin [88]. Amplification of *REL* locus is also found in another type of non-Hodgkin lymphoma, follicular lymphoma. Follicular lymphoma (FL) is a slow-growing, B-cell effecting cancer and damaged lymphocytes form follicles in cancerous lymph nodes. There is no standard treatment for FL, although immunotherapy with rituximab, radiotherapy and chemotherapy is applied to control the lymphoma and result in regression [89].

As c-Rel has mostly been implicated in lymphomas, the association of the transcription factor with other solid tumors has been observed. In one study, c-Rel is overexpressed in 50% cases of non-small cell lung carcinoma which is most common type of lung cancer, about 75% [90]. c-Rel has also been found overexpressed in many breast cancer specimens; 20 out of 23 cases. The inhibition of functional NF- κ B factors with microinjection of I κ B- α protein leads to death of breast cancer cell lines implying that c-Rel promoting cell survival and mediating tumor progression [91].

The NF- κ B pathway has been a center of focus in therapeutic applications because of its role in tumorigenesis and inflammation. However; most of the proteins in NF- κ B cascade are ubiquitously expressed and the pathway is involved in variety of biological processes in normal cells, blocking the pathway cause serious side effects. The heterodimeric p50/p65 complex is the most abundant one; so these transcription factors are responsible for many of the jobs in NF- κ B system, whereas c-Rel is mostly specialized in immune cells. As mentioned earlier, *REL* knockouts do not affect viability and development in mice; immune responses against antigens or in other words, proliferation and survival of B lymphocytes are impaired. Blocking of c-Rel does not cause severe

global immunodeficiencies, as it has been proved that innate immune responses remain intact. Hence, c-Rel blocking might be generally tolerated in treating B-cell lymphomas. There is a new study that demonstrates silencing of c-Rel with small interfering RNA leads to growth arrest and apoptosis of B cell tumor line [92]. Direct inhibitors of c-Rel might provide effective therapies for B-cell lymphomas and also they can be used for neoadjuvant chemotherapy in other solid tumors in which c-Rel is activated to reduce the size of tumor before surgery or radiotherapy.

4.2 c-Rel Crystal Structure

Human c-Rel crystal structure has not been determined yet. X-ray crystal structure of *Gallus gallus* (chicken) c-Rel bound to CD28 response element at 2.85 Å resolution was solved in 2001 by Ghosh group [93]. The structure is refined as c-Rel homodimer that is bound to enhancer region of cytokine interleukin-2. IL-2 is a small cell-signaling protein that stimulates growth and differentiation of T lymphocytes. The enhancer region around 300 bp upstream of IL-2 promoter region regulates the expression by binding with different transcription factors. One of the elements in enhancer region, CD28 response element, takes role in CD28-mediated expression by binding with c-Rel transcription factor. CD28 protein acts as a co-stimulatory signal for activation of T cells. CD28 initiates a signaling cascade which recruits some transcription factors including c-Rel into IL-2 enhancer region.

The full length of chicken c-Rel protein is 598 amino acid corresponding to about 65 kD in mass. RHD is 290 amino acids long that lie between 7th and 296th amino acid positions. Nuclear localization signal resides as expected on C terminal of RHD between 289 and 294 positions, KAKRQR. Researchers determined only RHD of c-Rel (amino acids 7-281) and 20 bp DNA fragment that includes 9 bp binding site 5`-AGAAATTCC-3`

and 8 bp flanking sequence at 5' and 3 bp at 3' side. Binding site does not carry the consensus κ B sequence but researchers showed that c-Rel homodimer binds with higher affinities to CD28 response element (the dissociation constant K_D is 25 nM) than several κ B sequences (K_D s range from 50 nM to 800 nM). The structure is deposited to Protein Data Bank under PDB code 1GJI (Figure 4.1). The figure was taken from Jmol in PDB with display style cartoon and colored according to secondary structures.



Figure 4.1: c-Rel homodimer bound to DNA view from top (left) and from side (right)

1GJI is folded into two distinct domains that are linked by short flexible linker. Larger domain is 168 amino acid-long and shorter one is 97 amino acid-long. The remaining 10 amino acids correspond to linker region in the crystallized c-Rel. According to CATH structural classification, two domains are mainly in beta sandwich architecture and their topology is Immunoglobulin-like fold (Ig-fold). PROMOTIF program

implemented in PDBsum reveals 22 β -strands and six small α -helices. The global structure c-Rel homodimer is similar to other Rel family dimers, not surprisingly. Carboxy terminal Ig-fold of RHD is smaller and exclusively specialized in dimerization with other subunit (Figure 4.2). A few residues in C-terminal fold also mediate nonspecific interaction with DNA backbone. Nearly 12 residues in each subunit take role in dimerization process. The core of the dimerization interface is comprised of mainly hydrophobic residues including phenylalanine, alanine, valine and leucine. The core is surrounded by charged amino acid residues, arginine, glutamate and aspartate, which form ionic interactions across subunits and mediate stabilization of dimers.

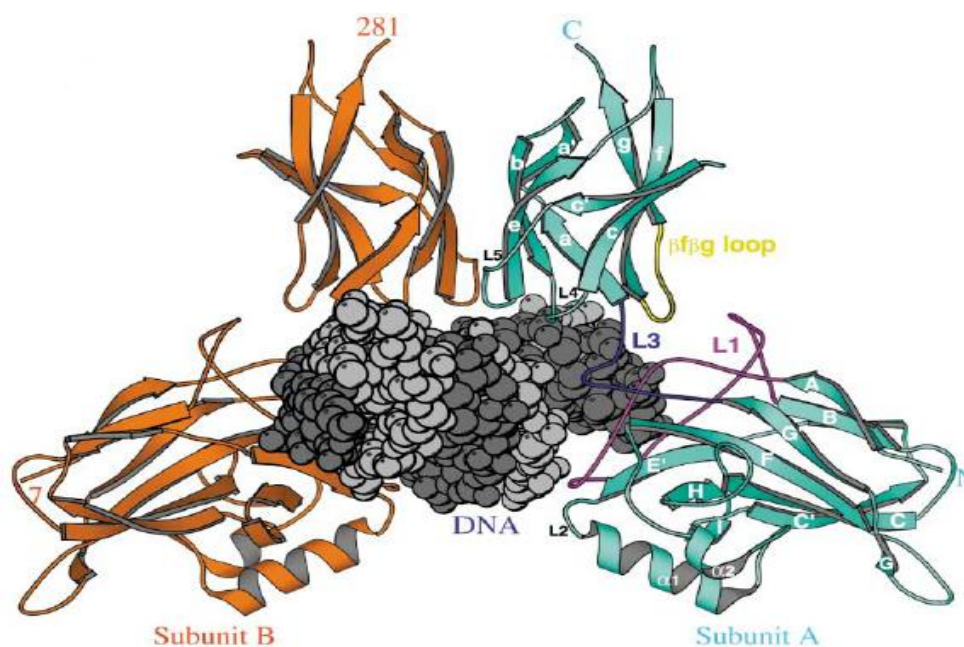


Figure 4.2: Overall structure c-Rel-DNA complex [93]

Larger N terminal Ig fold of RHD is primarily responsible for direct base contacts with DNA. Four of the contacts; Arg21, Arg23, Tyr24 and Glu27 reside in loop L1 in N terminal domain and the other contact Arg178 resides in loop L3 in the linker region (Figure 4.2). The residues on different domains also take role in interdomain interactions to stabilize DNA bound c-Rel structure. Arg18 is shown to make contacts with Thr182, Ala183, Glu184, Arg264, Asp267 and Glu269. However; these interactions do not seem to make N terminal domains stable in the absence of DNA. Like other dimers of Rel family in the absence of DNA, N terminal domains of c-Rel are more dynamic than C terminal domains which are stabilized through dimerization interactions.

```

gi|268370281_7-178      PYIEIFEQPRQGMRFYKCEGRSAGSIPGEHSTDNNKTFPSIQILNLYFGKVKIRITTLVT 60
gi|402649_8-179       PYIEIIEQPRQGMRFYKCEGRSAGSIPGEHSTDNNRTYPSIQIMNYYGKGVKIRITLVT 60
*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****

gi|268370281_7-178      KNEPYKPHPHDLVGDGCRDGYEAEFGPERRVLSFQNLGIQCVKKDLKESISLRISKKI 120
gi|402649_8-179       KNDPYKPHPHDLVGDGCRDGYEAEFGQERRPLFFQNLGIRCVKKKEVKEAIITRIKAGI 120
**:*:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****

gi|268370281_7-178      NPFNVPEEQLHNIDEYDLNVVRLCFQAFLPDEHGNYTLALPPLISNPIYDNR 172
gi|402649_8-179       NPFNVPEKQLNDIEDCDLNVVRLCFQVFLPDEHGNTLALPPVVSNPIYDNR 172
*****:*****:*****:*****:*****:*****:*****:*****:*****:*****

```

Figure 4.3: Pairwise alignment of chicken and human RHD of c-Rel

Human and chicken RHD are highly homologous. Above you can see the pairwise alignment of *Gallus gallus* and *Homo sapiens* RHD of c-Rel protein. The amino acid sequences are retrieved from NCBI Protein database. GenInfo Identifier 268370281 is chicken c-Rel accession number and GI 402649 corresponds to human c-Rel. Larger fold of RHD lies between 7th and 178th amino acid residues in chicken proteins; whereas it lies between 8th and 179th residues in human c-Rel. The alignment is performed in ClustalW2

(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) [94] program that reveals 81% identity between two sequences. Asterisk (*) in the alignment output indicates the residues are identical, semi colon (:) indicates conserved substitutions that residues have similar properties, period (.) indicates semi-conserved substitutions that residues have similar steric conformations but does not share chemical properties. Amino acid residues are also colored according to their chemical properties in the output file. Red colored amino acids are hydrophobic, blue are acidic, magenta are basic and green ones have polar side chains.

4.3 Homology modeling of human c-Rel

High similarity between human and chicken c-Rel RHD sequences offers promising prospect for homology modeling. There are plentiful of 3D structure prediction tools in literature. The main criterion for selection of these programs is CASP8 (2008) results. Four of the highest scoring servers, I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) [95], HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>) [96], pro-sp3-TASSER (<http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html>) [97] and Phyre (<http://www.sbg.bio.ic.ac.uk/~phyre/>) [50], have been used in the first step of modeling study. Figure 4.4 shows secondary structures of modeling results for each tools such that (b) is I-TASSER output, (c) is HHpred, (d) is pro-sp3-TASSER, (e) is Phyre server output and (a) is larger fold of chicken c-Rel RHD. The images are captured from Accelrys Discovery Studio Visualizer and blue colored arrows are β -strands, red tubes are α helices and green regions are loops. Sander and Schneider proposed that homology threshold is 24.8% with alignment length is larger than 80 residues [98]. It is largely accepted that protein pairs with more than 30% identical residues are structurally similar.

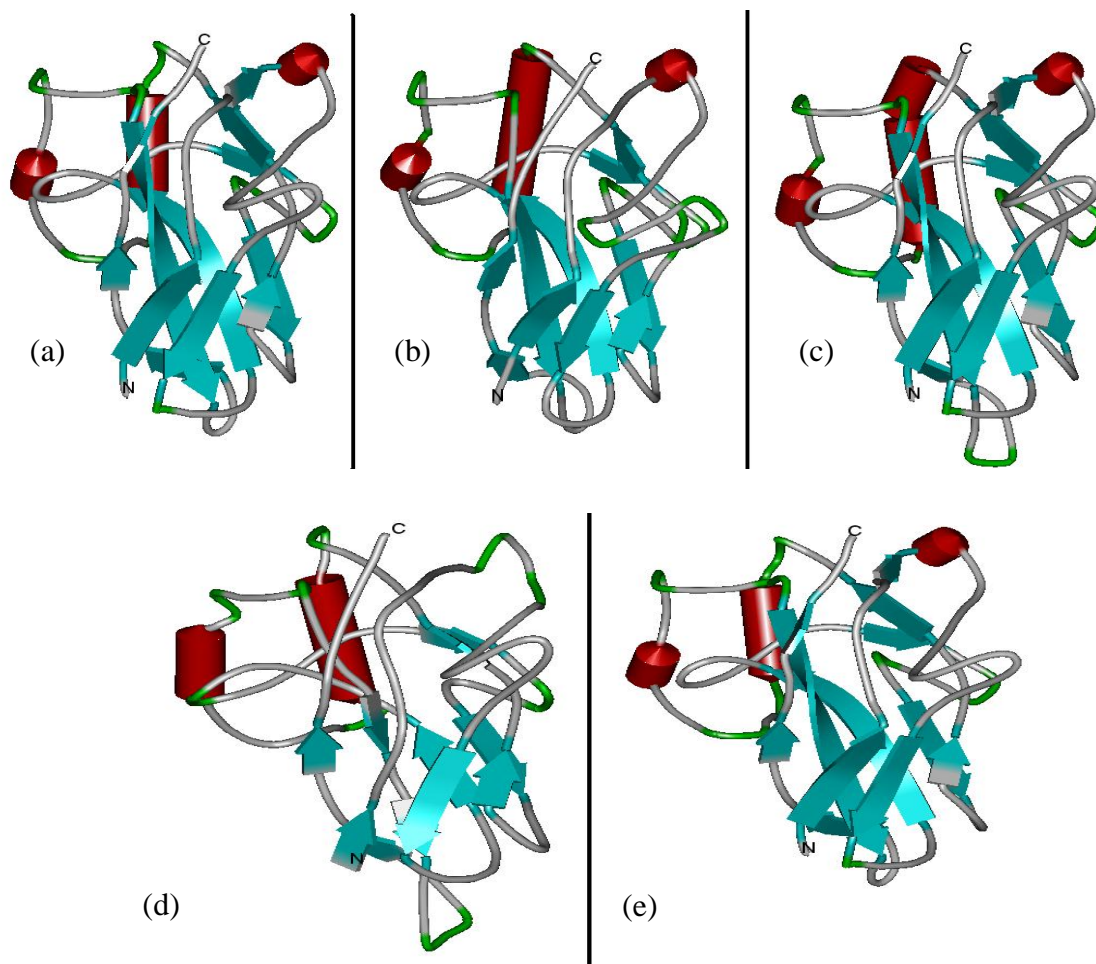


Figure 4.4: Schematic representations of modeled structures

81% identity between human and chicken c-Rel RHD is the main reason behind comparing secondary structures of modeled human c-Rel proteins and chicken c-Rel assuming that the structures might be similar overall. The coarse view of the figure shows that I-TASSER and Phyre outputs are more similar to X-ray structure and pro-sp3-TASSER is more distant. To get more detailed information; some features of DS Visualizer are used and root

mean square deviation (RMSD) values of each modeling tools' outputs against larger fold of RHD of chicken c-Rel are calculated. Two biopolymer structures are superimposed by their center of geometry and tube representations are captured (Figure 4.5). Green tube

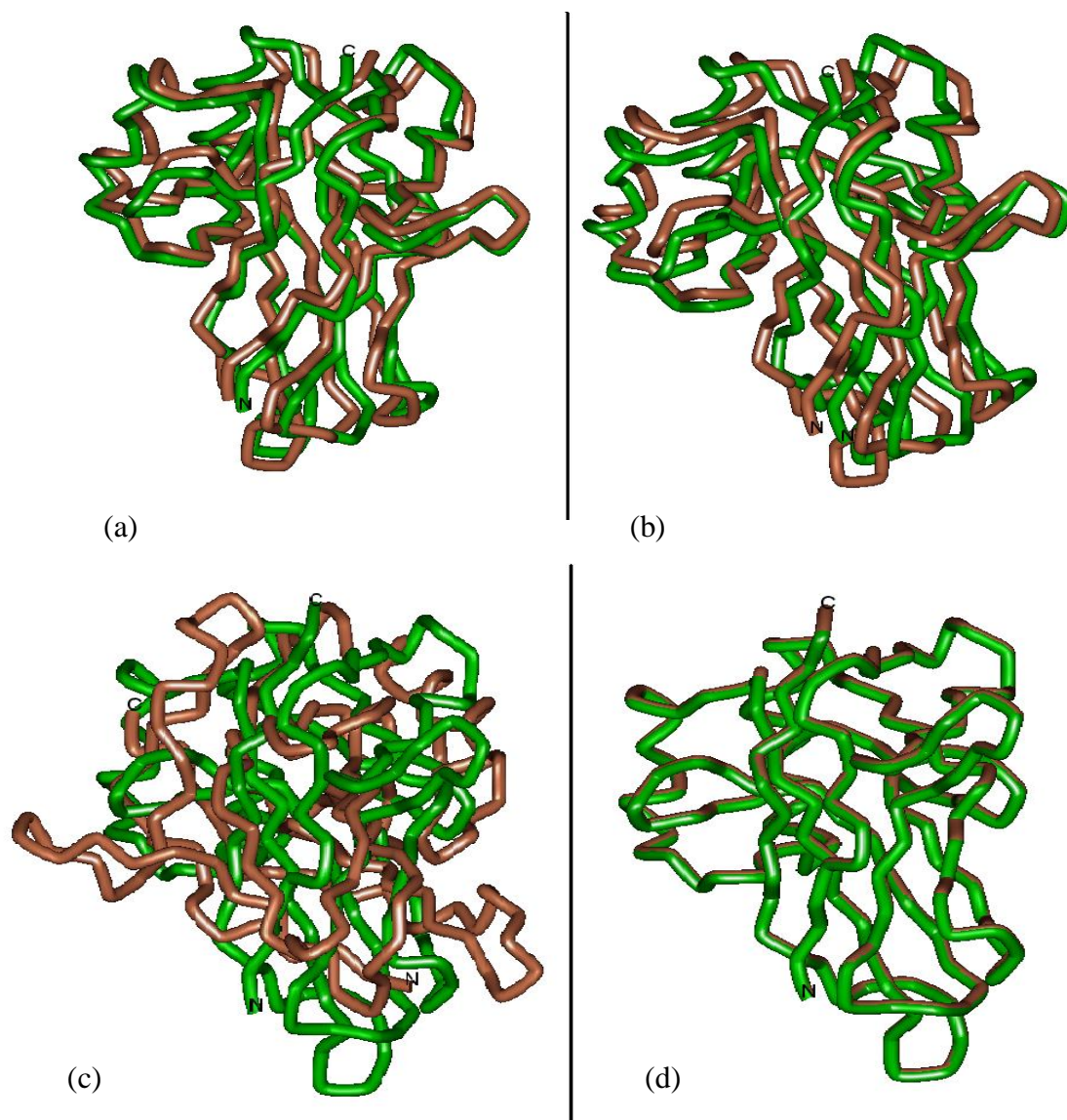


Figure 4.5: Superimposed images of modeled polypeptides vs. chicken c-Rel RHD

represents determined and refined structure of chicken c-Rel and pink tube represents I-TASSER output in (a), HHpred output in (b), pro-sp3-TASSER output in (c) and phyre output in (d). Although Figure 4.4(a) and 4.4(d) shows that overall 3D structures, not the topologies, might be a little similar, superimposed image (Figure 4.5.c) is much messier which can be explained by the change in the position of center of geometry. RMSD values of main chain atoms for these superimposed images are 2.126 Å for (a), 2.606 Å for (b), 18.369 Å for (c) and 0.241 Å for (d). Phyre server gives very similar output to 1GJI justifying some reports in literature which claims 50% or more sequence similarity tends protein backbones to be around 1 Å RMSD [99]. Since Phyre output is the closest one to known c-Rel structure, it is chosen for subsequent studies.

4.4 Energy Minimization of Modeled Structure

The energy minimizations are performed with molecular dynamics program NAMD 2.7 using the CHARMM22 all-hydrogen topology file for proteins and nucleic acids [100] and the whole system is visually examined by Visual Molecular Dynamics (VMD) tools [101]. The system that is subjected to minimization is modeled larger fold of c-Rel RHD bound to DNA. Since Phyre output and pre-determined structure has only 0.24 Å RMSD difference in backbone atoms and DNA contacting residues (Arg21, Arg23, Tyr24, Cys26, Glu27, Lys110, Lys111, Arg178) lie on highly conserved site; modeled c-Rel is placed in the position of c-Rel homodimer in 1GJI. Furthermore, the docking studies are primarily based on DNA-bound form of c-Rel and its inhibition and/or release from DNA; so it is significant to incorporate DNA into the minimized system. Before the minimization, protein-DNA complex is solvated by a water sphere with a radius around 42.4 Å, using the Solvate plugin of VMD. The system consists of total 31050 atoms including 8982 water

molecules and 4104 atoms from protein-DNA complex. 10000 steps of conjugate gradient minimization are applied to the system in NAMD. 1-4 interactions are excluded from non-bonded interactions (i.e., if atom A is bonded to atom B, and atom B is bonded to atom C, and atom C is bonded to atom D, then the atom pair A-D would be excluded including also 1-2 and 1-3 interactions). The cutoff distance for non-bonded van der Waals interactions and non-bonded electrostatic interactions is set to 12 Å. Switch distance is set to 10 Å where the switching function starts to take effect. The switching function smoothly truncates van der Waals potential at the cutoff distance. Pair list distance is set to 13.5 Å so that atom pairs within this distance are included into pair list but energies are calculated for the members of pair list that are inside the cutoff distance and this pair list is updated for each cycle which is determined with *stepspercycle* parameter and set to 20 steps. Other parameters and the configuration file are available in Appendix A.

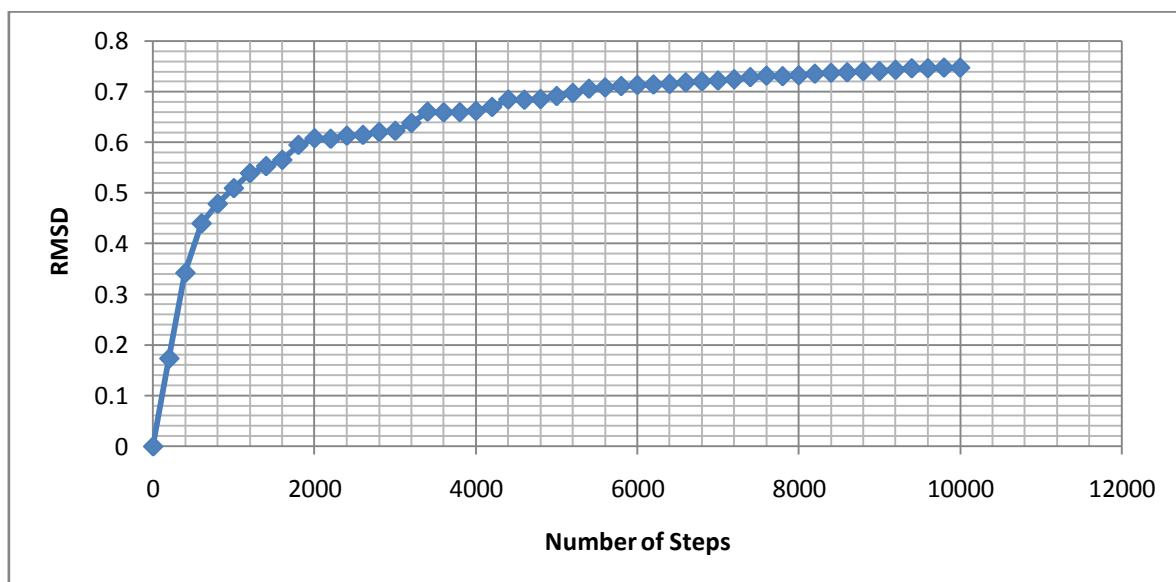


Figure 4.6: RMSD graph of entire protein over the course of minimization

The coordinate trajectory .dcd file is written for every 200 steps leading to 50 frames that gives idea about the behavior of the system during the minimization process. Protein structure from each frame is superimposed onto first frame's structure and RMSD values of backbone atoms are measured using VMD's *measure* function. The graph of these RMSD values over minimization steps is given in Figure 4.6. As it can be seen from the figure; the RMSD curve converges to a value around 0.75 Å. It would be more unequivocal if the minimization is proceeded for additional 1000 or 2000 steps that the RMSD value has reached its ultimate value, ergo the energy of the system is minimum. Figure 4.7 shows superimposition of secondary structures of larger fold of RHD of minimized and modeled human c-Rel (colored green) and X-ray crystallized chicken c-Rel (colored pink) and demonstrating that there no major changes in the secondary structure of the polypeptide. RMSD of the main chain atoms is found to be 0.91 Å.

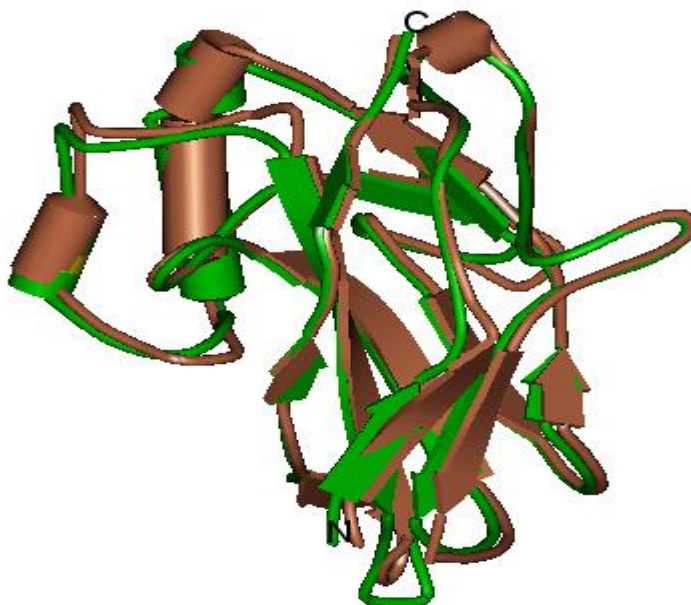


Figure 4.7: Superimposition of minimized modeled human c-Rel and chicken c-Rel

4.5 Drug Interaction Site Finding

The study of c-Rel as a drug target is founded on only one available crystal structure which is refined from chicken c-Rel sequence as bound to a response element. The crystal structure contains only RHD of c-Rel, 3D structure of transactivation domain that resides on C-terminal of c-Rel is not solved in that study. As stated earlier in literature review chapter and crystal structure of c-Rel section; RHD is composed of two immunoglobulin-like folds. Smaller C-terminal fold takes role in dimerization, inhibition by I κ B proteins and nuclear localization. Most of the approaches to NF- κ B signaling system members are focused on preventing initiation of NF- κ B response, but our approach heads towards termination of NF- κ B response. The hypothesis is succeeding in cessation of signal by inducing conformational change around DNA binding region upon binding of drug molecule. Since C-terminal transactivation domain and smaller fold of RHD domain are far away from DNA binding region and located on different domains, they are presumed to have no effect on DNA binding. Due to that, larger fold of RHD, which is fundamentally responsible for DNA binding, is taken into consideration throughout the study.

GNM comes into play in the process of determination of binding site. Fastest modes are known to be associated with energetically active residues as mentioned earlier. In a recent study with HLA Class I proteins showed that total mean square fluctuation of a residue at fastest mode is proportional to incoming energy from surroundings. If residues are on the surface of protein, it is called “energy gate”; if it is buried inside the protein, it is a “hub protein” which transfers the energy along interaction pathway [102]. Figure 4.8 shows these mean square fluctuations called D_i versus residue index for larger fold RHD of c-Rel protein. GNM is applied to Phyre result of c-Rel which is enumerated from amino acid #1 to amino acid #172, actually the region lies between 8th and 179th residues on native protein.

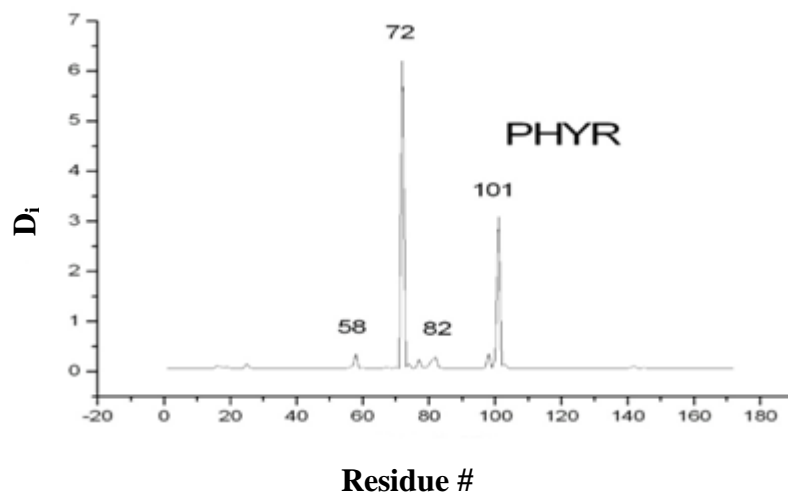


Figure 4.8: Energetically active residues on c-Rel

The energetically active residues correspond to Leu65, Leu79, Tyr89 and Arg108. Leu65 is buried inside the protein; whereas Leu79, Tyr89 and Arg108 are on the surface of c-Rel (Figure 4.9). They are displayed in CPK coloring schema. Leu79 and Arg108 reside inside a pocket that is also close to DNA binding site. Tyr89 is also on the surface of protein but it faces to towards a divergent side. Higher D_i values also support Leu79-Arg108 pocket to be used in docking studies. Figure 4.9 shows also residues Arg22, Arg24, Tyr25 and Glu28, which are in direct base contact, in ball and stick representation. Arg24, Tyr25 and Glu28 are also in close proximity of Arg108, around 4 Å. Our hypothesis is established on the assumption that a drug molecule interacts with energy gates Leu79 and Arg108 residues, which induces conformational change around DNA binding residues and ultimately the interaction between DNA and c-Rel is disrupted. One might ask why these DNA contacting residues do not give higher D_i values if they are in interaction pathway or respond to energy perturbation. The reason might be application of GNM method to DNA-

bound form of c-Rel in which these residues are already in contact with DNA, unfluctuating; so lower values of D_i .

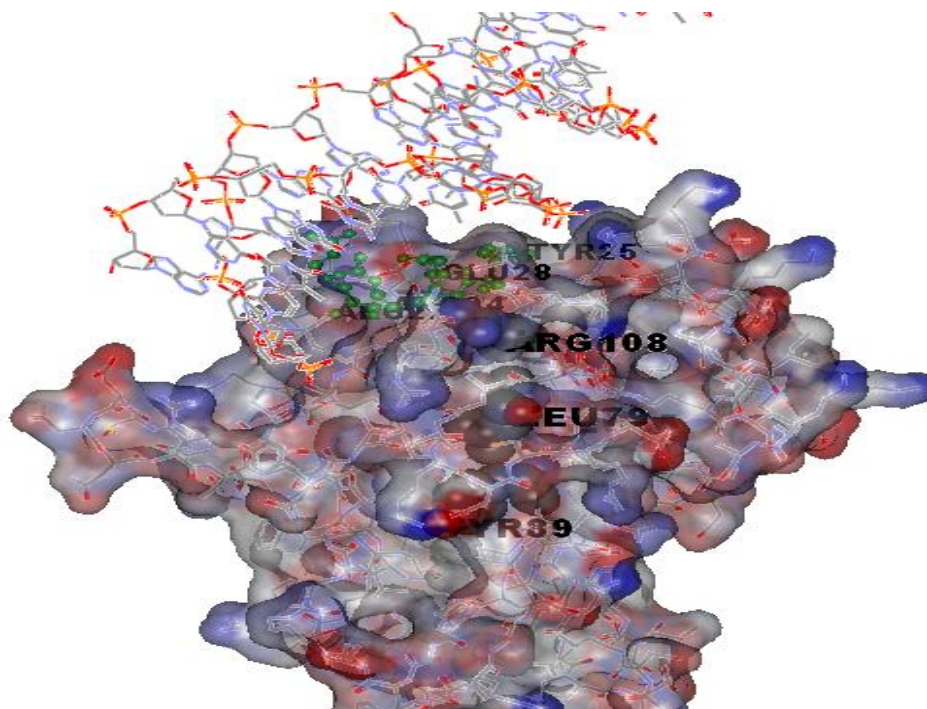


Figure 4.9: Surface representation of c-Rel bound to DNA

4.6 Docking Studies

4.6.1 Zinc Fragment-Like Set Docking

The first approach in docking studies is using small fragment-like molecules and identifying lead compounds for further modification. During these studies, all the molecules that are examined abide by Lipinski's rule of five. The rule of five suggests that chemical compounds with more than five hydrogen-bond donors, more than ten hydrogen-bond acceptors, molecular weight is over 500 and log P value is over five have more likely poor absorption and permeability properties [103].

There are several chemical databases that provide millions of compound structures, including ChEMBL, DrugBank, chemexper, ChemSpider, eMolecules, PubChem, SPRESI. ZINC (<http://zinc.docking.org>) is also one of the favorite compound databases that are used for ligand discovery. It is a free database and contains over 19 million compounds as of January 2011. The real advantage lies on downloading the compounds ready-to-dock 3D MOL2 and SDF file formats [104]. The database has also multifaceted search feature that you can limit your searches according to some molecular properties (e.g. net charge, log P, rotatable bonds, H-bond donors, H-bond acceptors, molecular weight and polar surface area), use similarity search by using SMILES strings or drawing the substructures on molecular editor. Since ZINC database is prepared using compound catalogs from vendors, search results also return vendor and purchasing information. The database also provides multiple protonation models and tautomeric forms of compounds. The researchers created ready-to-download subsets based on three physical properties; number of rotatable bonds, log P and molecular weight. Three main subsets are drug-like ($\log P \leq 5$, $150 < \text{molecular weight} \leq 500$, rotatable bonds < 8), lead-like ($2.5 < \log P < 3.5$, $250 < \text{molecular weight} < 350$, $5 < \text{rotatable bonds} \leq 7$) and fragment-like subsets ($\log P \leq 2.5$, molecular weight ≤ 250 , rotatable bonds ≤ 5).

432661 fragment-like compounds were downloaded in February 2010 from ZINC database. Molecules are merged into multi-structure MOL2 files and these files are split into individual files by the built-in function of Autodock Raccoon. Raccoon is a graphical interface to facilitate virtual screening and developed by Autodock team. Raccoon also converts MOL2 files to PDBQT format which is required for Autodock and Vina dockings. PDBQT format is used for both receptor and ligand molecules and it stores atomic coordinates, partial charges and Autodock atom types. Ligand PDBQT files also include information about rigid and flexible parts of the ligand. Receptor PDBQT file is prepared by AutodockTools which is a graphical user interface sets up parameter and input files,

launches Autodock and analyze docking results. Vina uses compatible file formats with Autodock; so AutodockTools becomes handy when preparing input files, choosing the search space and viewing docking results.

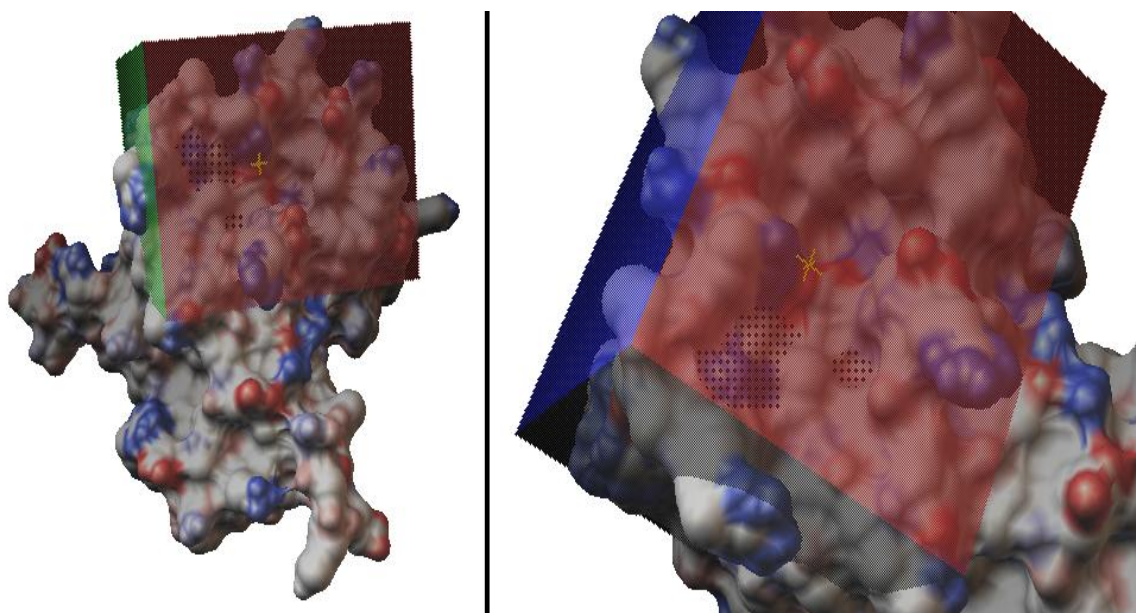
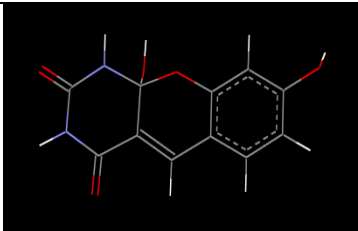
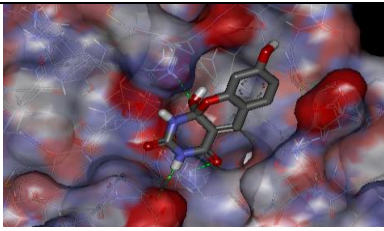
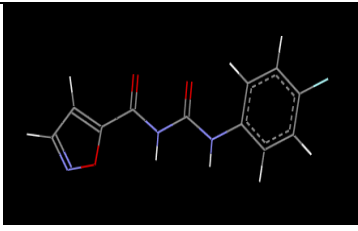
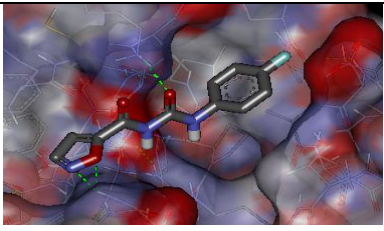


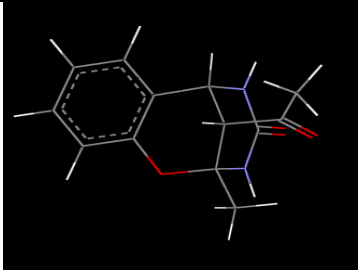
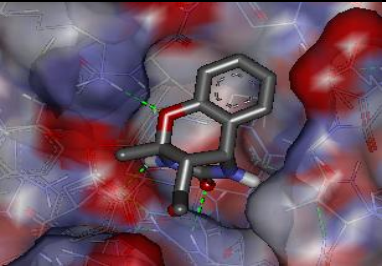
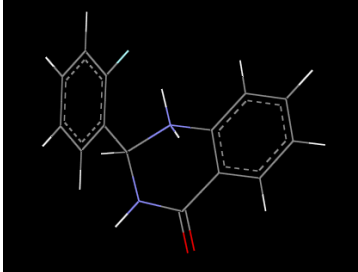
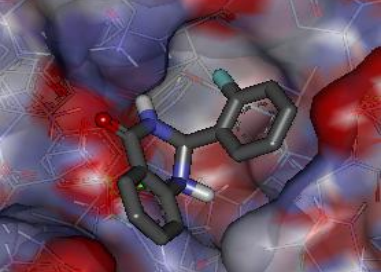
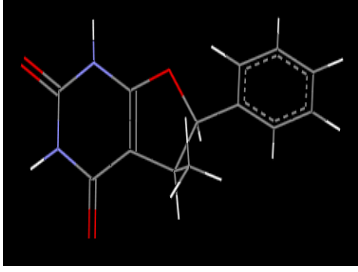
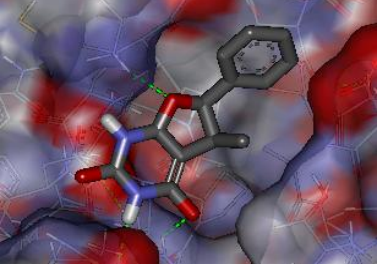
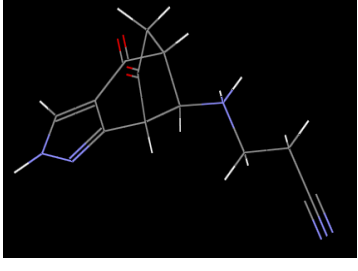
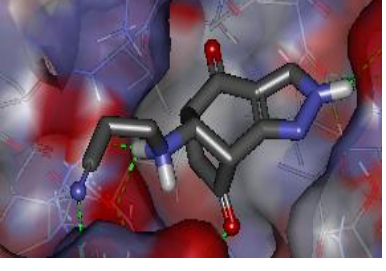
Figure 4.10: Grid box of receptor used during docking studies

Search space or grid box of c-Rel polypeptide is shown in figure 4.10. The box is 24 Å in x-direction, 28 Å in y-direction, 22 Å in z-direction and centered near Leu79 and Arg108 residues. Yellow cross represents center of the box; red faces of the box depict x-direction, green faces y-direction and blue faces z-direction. The images are captured from AutodockTools and the receptor is displayed in molecular surface representation, colored according to atom charges. Black dotted regions show Leu79 and Arg108 regions. Side chain of Arg108 is mostly exposed to the surface; however only carboxy terminal of Leu79

is on the surface c-Rel. The grid box is kept a little wider to cover the groove where also Leu79 resides.

Vina is utilized as the docking software in virtual screening studies because of its speed and taking advantage of multiple CPU cores. The exhaustiveness level, which determines the thoroughness of search algorithm, is set to eight as in the default. It is thought to be adequate exhaustiveness value for small fragment-like ligands. When the ligands are modified, i.e. new groups are added to lead compound, the exhaustiveness level is increased. Although it is assured that Vina algorithm itself changes time spent on the search depending on the size of ligand and flexibility, meticulousness of search is strengthened. Best ten docking results of fragment-like dataset are provided in Table 4.1

ZINC Code	Binding Energy (kcal/mol)	Structure	Docked Figure
ZINC00082871	-7.5	 The chemical structure of ZINC00082871 is a complex polycyclic molecule. It features a central six-membered ring with two carbonyl groups (C=O) and a nitrogen atom. This central ring is fused to a five-membered ring and a six-membered ring. The six-membered ring has several methyl groups and a dashed line indicating a specific conformation or feature.	 The docked figure shows the chemical structure of ZINC00082871 (in stick representation) bound within a protein's binding pocket. The protein's surface is shown as a semi-transparent mesh with a color-coded electrostatic potential map, where red indicates negative charge and blue indicates positive charge. The ligand is positioned to interact with the pocket's residues.
ZINC00121681	-7.5	 The chemical structure of ZINC00121681 is a complex polycyclic molecule. It features a central six-membered ring with two carbonyl groups (C=O) and a nitrogen atom. This central ring is fused to a five-membered ring and a six-membered ring. The six-membered ring has several methyl groups and a dashed line indicating a specific conformation or feature.	 The docked figure shows the chemical structure of ZINC00121681 (in stick representation) bound within a protein's binding pocket. The protein's surface is shown as a semi-transparent mesh with a color-coded electrostatic potential map, where red indicates negative charge and blue indicates positive charge. The ligand is positioned to interact with the pocket's residues.

ZINC01092567	-7.3	 A 2D chemical structure of ZINC01092567, showing a central benzene ring with a dashed circle indicating its aromaticity. It is substituted with a methyl group, a methoxy group, and a complex side chain containing a nitrogen atom, a carbonyl group, and a methyl group.	 A 3D molecular docking model of ZINC01092567. The molecule is shown in a stick representation, with carbon in grey, oxygen in red, and nitrogen in blue. It is docked into a protein binding pocket, with red and blue electrostatic potential surfaces visible in the background.
ZINC02807066	-7.3	 A 2D chemical structure of ZINC02807066, featuring two benzene rings connected by a nitrogen-containing bridge. Both rings are substituted with methyl groups.	 A 3D molecular docking model of ZINC02807066. The molecule is shown in a stick representation, docked into a protein binding pocket with electrostatic potential surfaces.
ZINC01645508	-7.3	 A 2D chemical structure of ZINC01645508, showing a central benzene ring with a dashed circle. It is substituted with a methyl group, a methoxy group, and a complex side chain containing a nitrogen atom, a carbonyl group, and a methyl group.	 A 3D molecular docking model of ZINC01645508. The molecule is shown in a stick representation, docked into a protein binding pocket with electrostatic potential surfaces.
ZINC05286149	-7.1	 A 2D chemical structure of ZINC05286149, featuring a benzene ring with a dashed circle. It is substituted with a methyl group, a methoxy group, and a complex side chain containing a nitrogen atom, a carbonyl group, and a methyl group.	 A 3D molecular docking model of ZINC05286149. The molecule is shown in a stick representation, docked into a protein binding pocket with electrostatic potential surfaces.

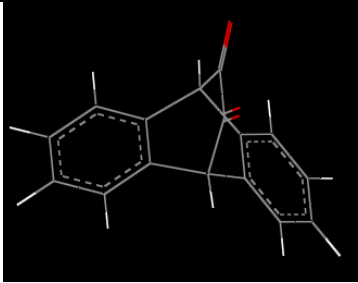
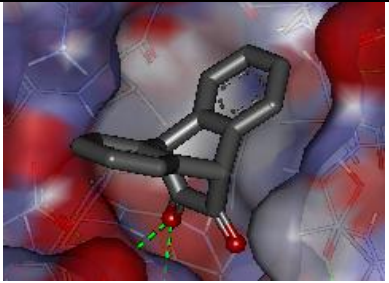
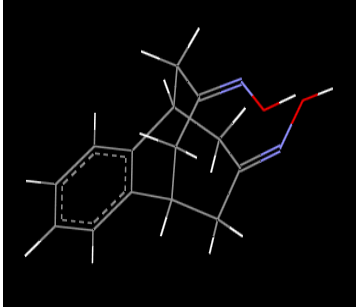
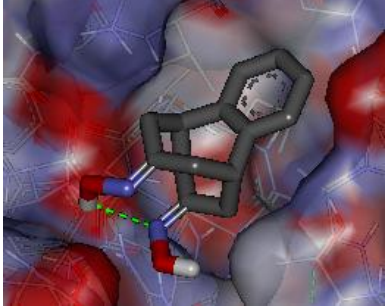
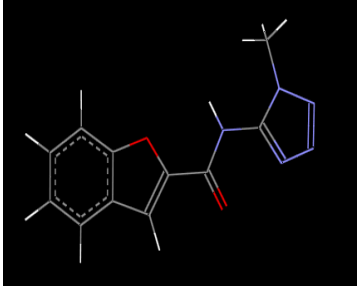
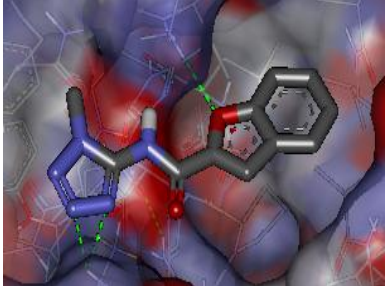
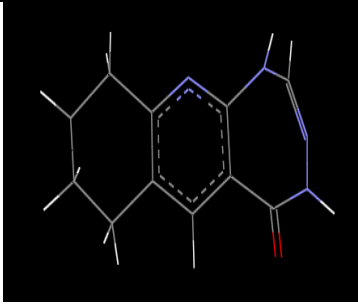
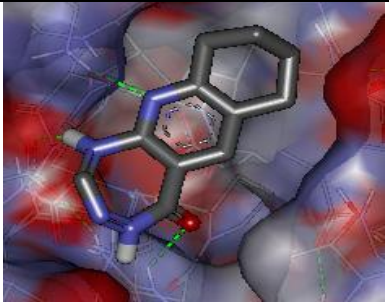
ZINC01872076	-7.1	 A 2D chemical structure of ZINC01872076, showing a central benzene ring with a dashed circle indicating its aromaticity. It is substituted with a methyl group, a methoxy group, and a complex side chain containing a quaternary carbon, a methyl group, and a carbonyl group.	 A 3D molecular docking model of ZINC01872076. The ligand is shown in a stick representation with a grey carbon skeleton, red oxygen atoms, and blue nitrogen atoms. It is docked within a protein's binding pocket, which is visualized as a semi-transparent surface colored by electrostatic potential (red for negative, blue for positive, grey for neutral). Green dashed lines represent hydrogen bonds between the ligand and the protein.
ZINC00367077	-7.1	 A 2D chemical structure of ZINC00367077, featuring a central benzene ring with a dashed circle. It has a methyl group, a methoxy group, and a side chain containing a quaternary carbon, a methyl group, and a hydroxyl group.	 A 3D molecular docking model of ZINC00367077. The ligand is shown in a stick representation with a grey carbon skeleton, red oxygen atoms, and blue nitrogen atoms. It is docked within a protein's binding pocket, visualized as a semi-transparent surface colored by electrostatic potential. Green dashed lines represent hydrogen bonds.
ZINC00514909	-7.1	 A 2D chemical structure of ZINC00514909, showing a central benzene ring with a dashed circle. It is substituted with a methyl group, a methoxy group, and a side chain containing a quaternary carbon, a methyl group, and a pyridine ring.	 A 3D molecular docking model of ZINC00514909. The ligand is shown in a stick representation with a grey carbon skeleton, red oxygen atoms, and blue nitrogen atoms. It is docked within a protein's binding pocket, visualized as a semi-transparent surface colored by electrostatic potential. Green dashed lines represent hydrogen bonds.
ZINC08655548	-7.0	 A 2D chemical structure of ZINC08655548, featuring a central benzene ring with a dashed circle. It is substituted with a methyl group, a methoxy group, and a side chain containing a quaternary carbon, a methyl group, and a pyridine ring.	 A 3D molecular docking model of ZINC08655548. The ligand is shown in a stick representation with a grey carbon skeleton, red oxygen atoms, and blue nitrogen atoms. It is docked within a protein's binding pocket, visualized as a semi-transparent surface colored by electrostatic potential. Green dashed lines represent hydrogen bonds.

Table 4.1: Top ten docking results of fragment like ligands

Results seem encouraging for these small compounds. Modifications and addition of groups to these lead compounds apparently result in lower binding energies. Nevertheless, the presence of RHD in other NF- κ B raises the question whether these compounds are specific to c-Rel. Multiple alignments of N-terminal larger folds of RHD is done in ClustalW2 server and is shown in Figure 4.11. GI 402649 is c-Rel sequence, GI 48775023

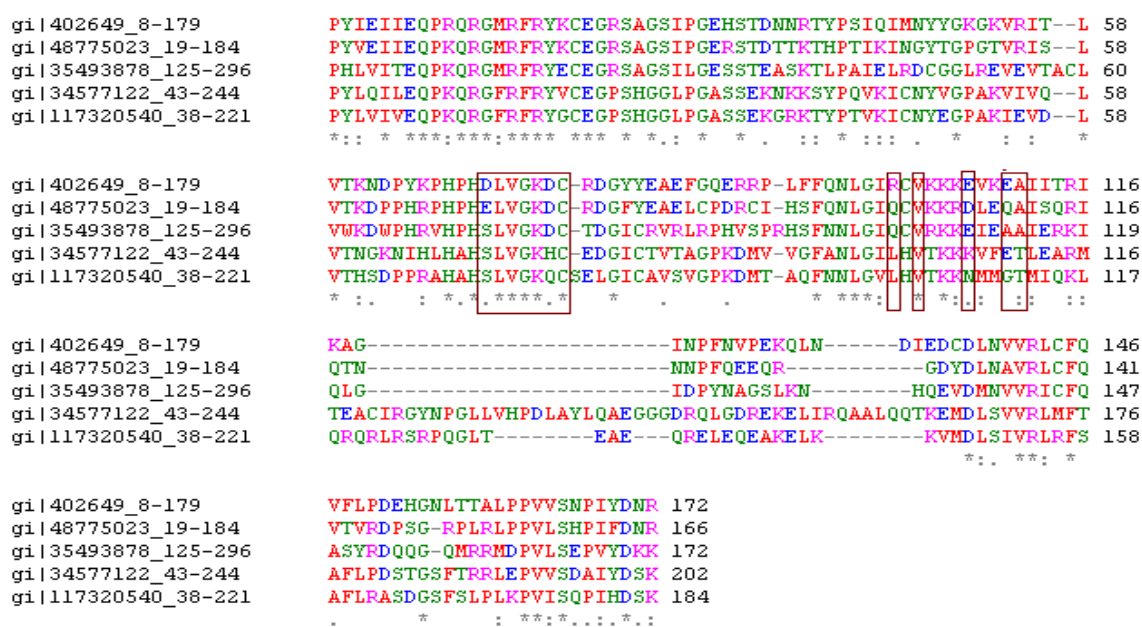


Figure 4.11: Multiple sequence alignment of NF- κ B members

is RelA sequence, GI 35493878 is RelB sequence, GI 34577122 is p105 (non-degraded p50) sequence, GI 117320540 is p100 (non-degraded p52) sequence. Other numbers next to GenInfo IDs denote the residue numbers which they lie in between in native protein. The residues show 27% identity to c-Rel (calculated as total number of identical residues divided by total length of N-terminal RHD of c-Rel). The red boxes enclose the residues

which reside in the pocket of interest. Asp78, Leu79, Val80, Gly81, Lys82, Asp 83, Cys84, Arg108, Val109, Glu110, Glu117 and Ala118 are the pocket residues of c-Rel. Residues around Leu79 seems to be conserved, yet we cannot say the same thing for Arg108. The conservation of Leu72 and its immediate surroundings over NF- κ B members might give us a clue about its functional importance. The highly conserved region between Gly20 and Ser34 is DNA binding region of NF- κ B transcription factors.

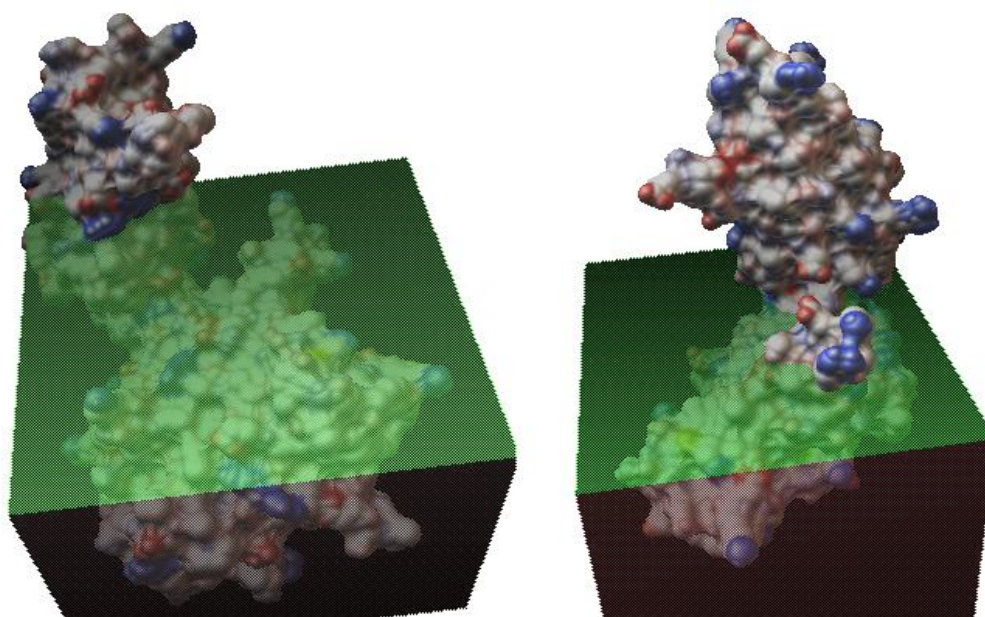


Figure 4.12: Grid boxes of p50 (left) and p52 (right)

Best ten compounds are docked to four other members of Rel family with Vina. The grid boxes are kept such large that the entire N-terminal fold of RHD is encompassed (Figure 4.12 - 4.13). This phenomenon is known as blind docking in which binding site is unknown and whole receptor or domain is searched. The exhaustiveness level of Vina is increased to 25 for more thorough search over highly expanded search space. Receptor

structures are extracted from PDB, p50: 1SVC, p52: 1A3Q, RelA: 3GUT, RelB: 3DO7.

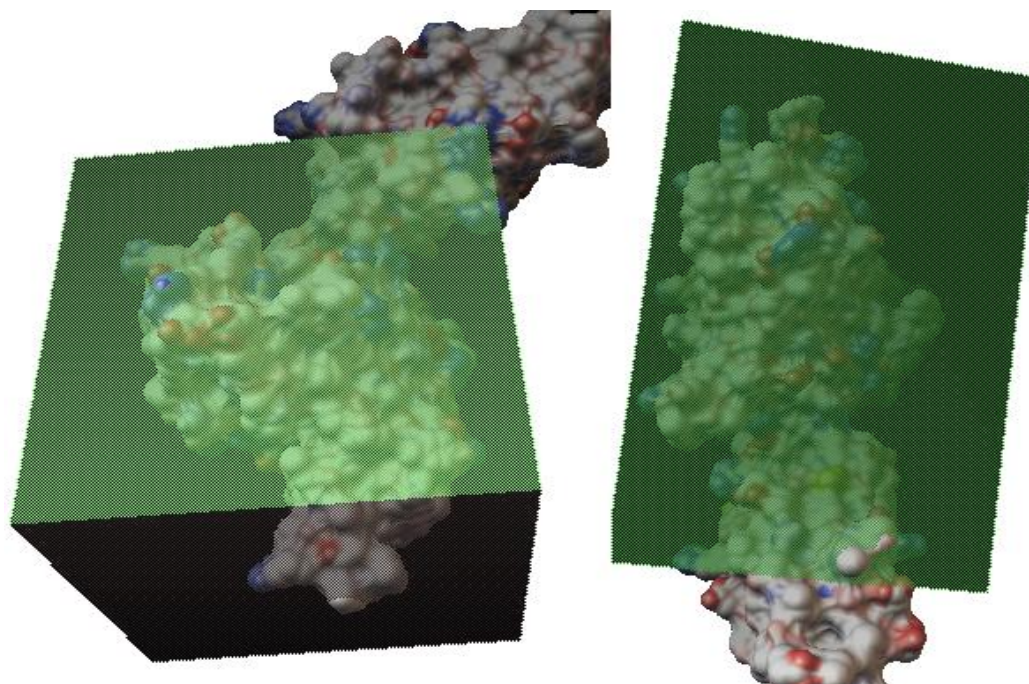


Figure 4.13: Grid boxes of RelA (left) and RelB (right)

All the sequences except RelB are human origin, but 3DO7 contains *Mus musculus* (mouse) RelB sequence. Phyre server used for homology modeling and the output is treated as the receptor for these selectivity studies of small compounds.

Table 4.2 lists the binding energies of small fragment-like ligands of interest. According to these energies, ZINC00121681 is the most favorable compound for lead modification such that it is one of the best bound ligand to c-Rel and deviation over other NF- κ B proteins is the largest. ZINC00121681 is catalogued from MolPort database (www.molport.com/) with compound number MolPort-002-924-612. The IUPAC name of the compound is 1-(4-fluorophenyl)-3-[(1,2-oxazol-5-yl)carbonyl]urea with molecular

ZINC #	p50	p52	p65 (relA)	relB	c-Rel
ZINC00082871	-7.0	-7.3	-7.0	-7.1	-7.5
ZINC00121681	-6.6	-6.7	-5.8	-6.7	-7.5
ZINC01092567	-5.8	-6.5	-6.6	-6.9	-7.3
ZINC02807066	-6.1	-6.7	-6.0	-7.2	-7.3
ZINC01645508	-7.1	-6.3	-6.2	-6.9	-7.3
ZINC05286149	-5.6	-6.7	-5.9	-6.9	-7.1
ZINC01872076	-6.2	-8.0	-6.6	-6.9	-7.1
ZINC00367077	-5.4	-6.9	-6.3	-6.6	-7.1
ZINC00514909	-7.2	-6.6	-6.0	-7.1	-7.1
ZINC08655548	-6.2	-6.4	-6.2	-7.4	-7.0

Table 4.2: Binding energies (kcal/mol) of top ten compounds over Rel family members

formula $C_{11}H_8FN_3O_3$. The molecular mass of compound is 249.2 Dalton; log P value is 1.15 with two H-bond donors and five H-bond acceptors. log P value (partition coefficient) is commonly used in pharmacology to assess the drug-likeness of the molecule. It is the measure of lipophilicity of molecule and calculated as the logarithm of ratio of concentrations water-immiscible solvent and water, $\log P_{oct/wat} = \log \left(\frac{[solute]_{octanol}}{[un-ionized]_{water}} \right)$. The molecular mass, partition coefficient, H-bond donor and acceptor counts are determined with MarvinView application.

Before starting modification, similarity feature of ZINC database is utilized. There are 107 compounds in the database that are at least 60% similar to ZINC00121681. These compounds are docked to c-Rel with Vina at exhaustiveness level 15. There are two compounds that have better binding energies than ZINC0012681, ZINC42093947 (Figure 4.14-left) and ZINC37143096 (Figure 4.14-right). ZINC42093947 binds to c-Rel with -7.9

kcal/mol binding energy and ZINC37143096 binds with -7.6 kcal/mol. However; these compounds are bigger than ZINC00121681 and do not offer much improvement in the binding energy, they are not pursued for further studies.

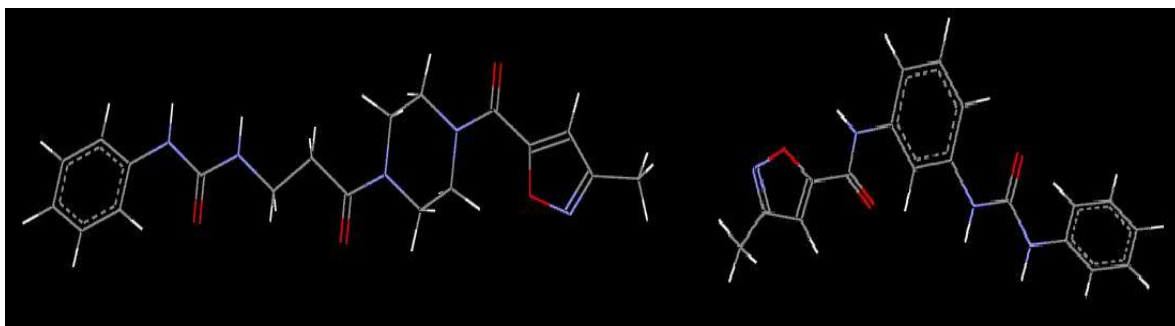


Figure 4.14: Similar compounds of best docked ZINC00121681

The modification of the lead compound ZINC0012681 is achieved with combinatorial virtual library generation software iLib Diverse (Inte:Ligand GmbH[®]) [105]. iLib Diverse builds drug-like molecules from an extensive fragment set and provide users to augment these fragments to their compounds of interest. The reactive sites on the compounds can be changed in order to restrict modification to specific sites. Predefined filters are available in the program to aid selection according to blood-brain barrier permeability, oral bioavailability, lead and high-drug likeness. The users can also define custom filters by changing various parameters of chemical properties. Easy-to-use software generates outputs either in SMILES string format or in 3D SD format. iLib Diverse implements CORINA by Molecular Networks for high-quality 3D structure generation.

1500 modified ZINC00121681 compounds are generated with iLib Diverse. Filter properties are mostly like Lipinski rule of five, except upper limit of molecular weight is set to 375. The reason for this restriction is adding chemical groups to lead molecule

gradually; so there might be higher improvements than building compounds up to 500 Da at once. iLib Diverse gives a single multi-structure output file with SD format which is converted to MOL2 format with MarvinView. The MOL2 file is then split into single MOL2 structures and PDBQT files are generated with Autodock Raccoon. This procedure is always used when ligand PDBQT files are created from multi-structure file. Each ligand is docked to c-Rel in Vina with exhaustiveness level 15.

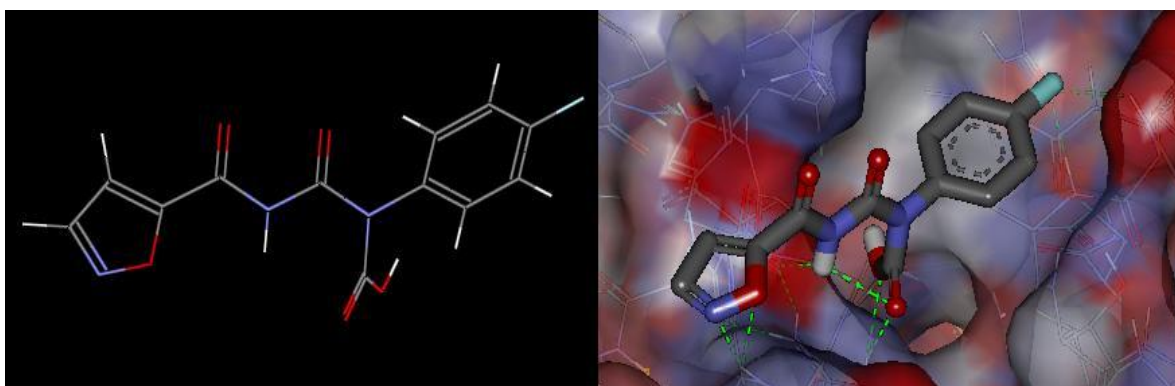


Figure 4.15: Compound I and its docked position

The highest binding score is obtained from the compound in Figure 4.15. The ligand is bound with -8.8 kcal/mol to the receptor and newly-added hydroxyl group looks like a good fit to the groove in binding pocket when comparing with the figure in Table 4.1. Addition of carboxyl group to N1 atom of urea improves binding energy without dramatically increasing molecular weight. New compound (compound I), 1-(4-fluorophenyl)-1-carboxy-3-[(1,2-oxazol-5-yl)carbonyl]urea, has 293.21 Da molecular mass, 1.25 log P value, two H-bond donor atoms and eight H-bond acceptor atoms. Second phase of modifications with iLib Diverse and docking of them to c-Rel are not successful. The lowest binding energy is -8.9 kcal/mol that is not favorable with larger molecule size.

When the ligand is fit to groove after carboxyl group attachment, the main purpose of next phase of modifications is the attachment of radical groups to oxazole ring so as to engage also Arg101 residue in binding. Hence, the reactivity of carbon atoms in oxazole ring is increased and reactivity of carbon atoms in phenyl ring is decreased in iLib Diverse to restrict the augmentation to one side of compound. 750 new molecules are created for third phase of modification according to Lipinski rule of five with iLib Diverse and docked with Vina.

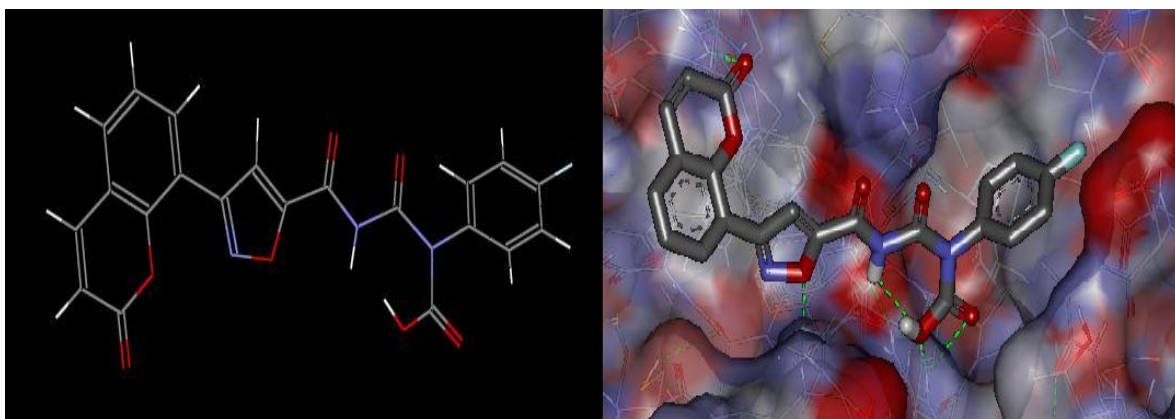


Figure 4.16: Compound **II** and its docked pose

Best binding energy is obtained from the molecule which is modified by coumarin group attachment to carbon-3 atom of oxazole ring (Figure 4.16). New molecule (compound **II**) binds to c-Rel with -10.4 kcal/mol energy, which is huge improvement in binding energy. Compound **II**, $C_{21}H_{12}FN_3O_7$, has 437.33 Da molecular mass, 3.09 log P value, two H-bond donor and nine H-bond acceptor sites. Coumarins are bicyclic oxygen heterocycles and form a large group of plant secondary metabolites. Coumarin derivatives are shown to have anti-leukemia, anti-inflammation, anti-cancer, anti-coagulant, and antibiotic properties. Because of their diverse pharmacological aspect, they attract intense research interest in recent years [106].

Compound **II** has also been docked to other Rel members to check its interaction with other Rel family proteins. Same grid boxes that cover whole polypeptide are used as in the previous selectivity study. Compound **II** is bound to the nearly the same pocket in c-Rel, p50, p65 and relB proteins in its lowest energy configuration; however ligand is bound to that pocket in p52 with -7.8 kcal/mol binding free energy. The lowest binding energies of Compound **II** to NF- κ B members are given below.

	p50	p52	p65	relB	c-Rel
Binding Energy (kcal /mol)	-8.6	-8.4	-7.5	-8.7	-10.4

The free energy of binding (ΔG) for compound **II** is estimated -10.4 kcal/mol. The dissociation constant for drug and receptor can be calculated as $\Delta G = RT \ln K_i$ where R is gas constant (1.9858775 cal/Kmol), T is temperature (we consider body temperature, 310.15K). The dissociation constant for compound **II**-c-Rel complex is calculated as 46.43 nM and for compound **II**-relB complex it is 733.67 nM.

IC_{50} values are generally used in pharmacology to assess the effectiveness of drug. IC_{50} is known as half maximal inhibitory concentration; in other words the concentration of drug that is required for 50% inhibition. When there is no competition between substrate and drug, as in our case, IC_{50} value becomes equal to K_i value. IC_{50} of ZINC00121681 for c-Rel (referring to Table 4.2) is 5.15 μ M and for RelB is 18.87 μ M. The reason for presenting these values is showing how the modification improves binding affinity and how selective the compounds are. Compound **II**, derived from ZINC00121680, has 110-fold higher affinity to c-Rel than ZINC00121680. The binding affinity of compound **II** to relB might seem higher; but it should be noted that RHD of c-Rel and relB shows 47% sequence identity. However; while ZINC00121680 has 3.5-fold higher affinity to c-Rel

than relB, compound **II** has 15.8-fold higher affinity. This fact also proves the success of modification process.

Autodock 4.2 is also used as a secondary docking program to get an auxiliary opinion about binding pose and binding energy of compound **II**. The grid spacing is set to 0.375Å; however Vina uses 1Å spacing. Therefore; grid box size is increased to cover the same space with Vina. Lamarckian genetic algorithm is used as the search algorithm with population size 200, number of energy evaluations 5000000 and 30 run. Other parameters are used as in the default value of Autodock. Figure 4.17 shows docking poses of Autodock (green molecule) and Vina (yellow). RMSD of heavy atoms of these poses is 1.092 Å. However; this concordance in docked positions is not achieved in binding energy estimation. The free energy of binding is calculated as -8.52 kcal/mol in Autodock.

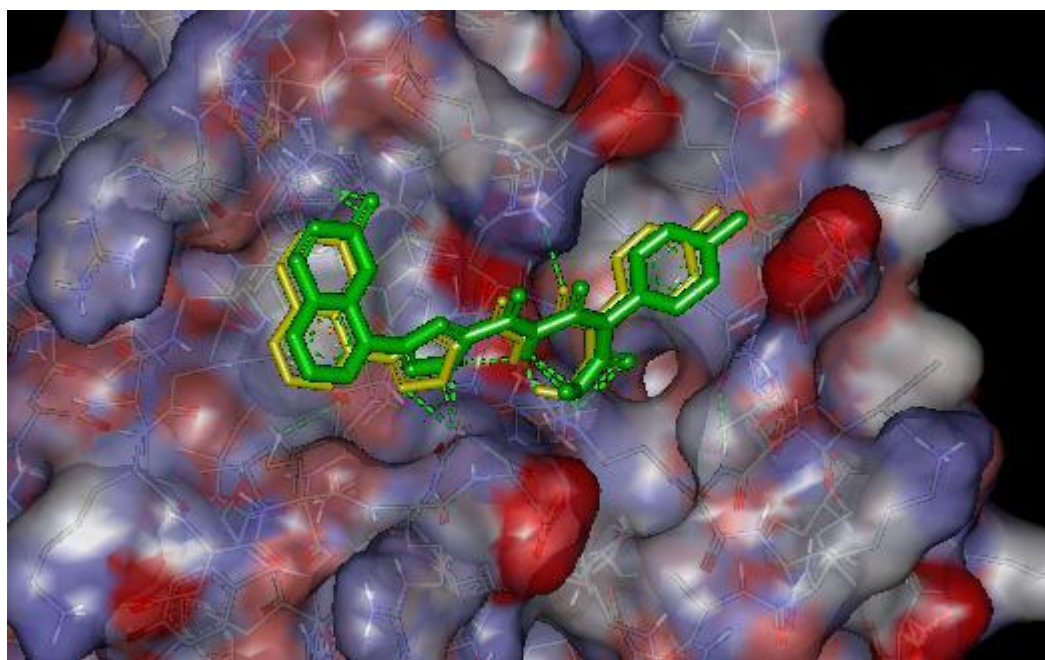


Figure 4.17: Superimposition of Vina and Autodock docked poses of compound **II**

Compound **II** forms hydrogen bonds with Asp78, Gly87, Arg108 and Arg122 residues (Figure 4.18). Guanidinium hydrogen of Arg108 interacts with oxo group of coumarin part of compound **II**. Carboxylic oxygen atom of Asp78 forms H-bond with hydrogen atom of carbamide group. Amino hydrogen of Gly87 interacts with oxygen atom of oxazole group and guanidinium hydrogen of Arg122 interacts with carboxy oxygen of compound **II**.

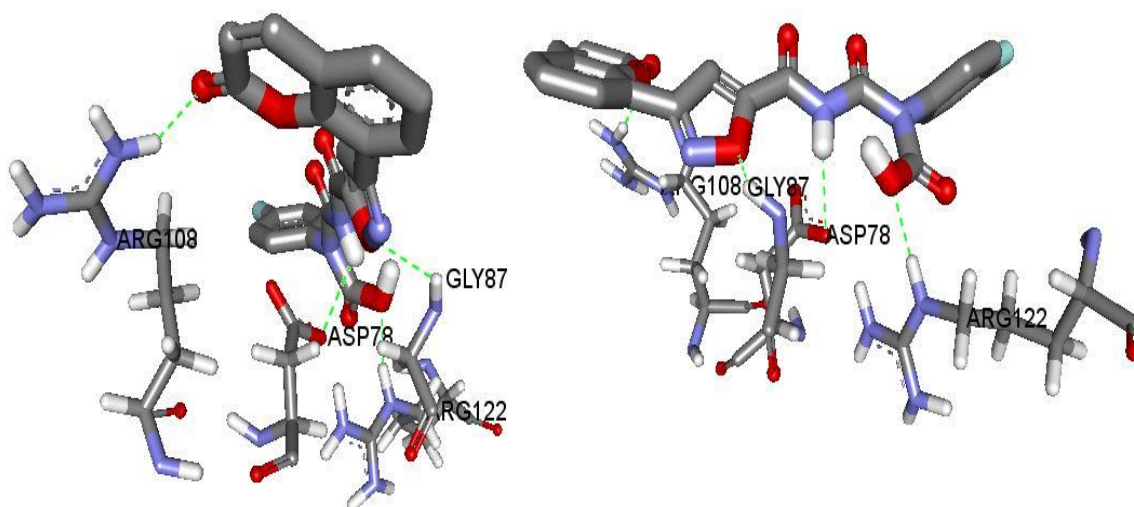


Figure 4.18: Compound **II** and hydrogen bonding residues

Besides, there are also unconventional hydrogen bonds that contribute to nonbonded interactions between compound **II** and protein. The attraction between C-H bond and Π system is referred to as CH--- Π interaction which is weaker than conventional hydrogen bonding. The interaction results from dispersion attractions between carbon atom of CH group (carbon atom has larger electric polarizability than hydrogen atoms) and carbon atoms of Π system. The distance between center of aromatic ring and carbon atom must be less than 4 Å for CH--- Π interaction. Methyl hydrogen of Ala118 and fluorophenyl ring of compound **II** are involved in CH--- Π interaction (Figure 4.19).

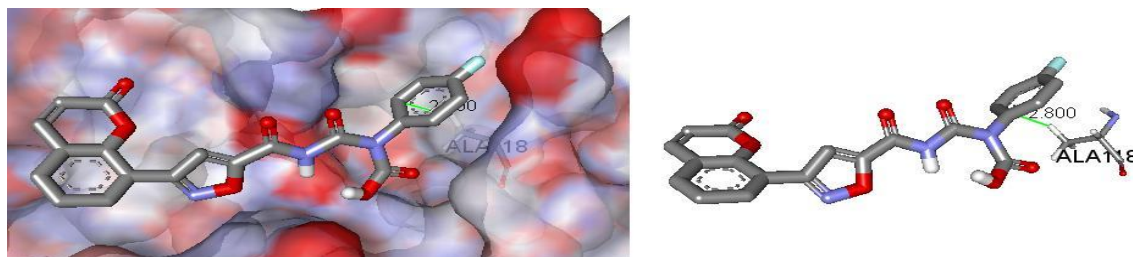


Figure 4.19: CH--- π interaction between Ala118 and compound **II**

C-H---O interactions are another type of weak hydrogen bonds. They resemble to genuine hydrogen bonds regarding that electrostatic interactions are largest contributor of attraction rather than dispersion, charge transfer and polarization components. However; when CH₄ acts as a proton donor, C-H---O interaction becomes quite weaker. Replacement of hydrogen atoms with electronegative substituents, e.g. fluorine, strengthens the interaction. C---O distance can be up to 4.2 Å in C-H---O interaction [107]. Side chain carbon of Val80 interaction with oxygen atom of dihydropyran ring, carboxyl oxygen of Arg30 interaction with carbon atom of dihydropyran ring, carboxyl oxygen of Cys84 interaction with phenyl group, side chain carbon of Arg108 interaction with keto oxygen of coumarin group, side chain carboxyl oxygen of Glu117 interaction with phenyl group and methyl group of Ala118 interaction with carboxyl oxygen of compound **II** are C-H---O interactions between ligand and receptor molecules (Figure 4.20).

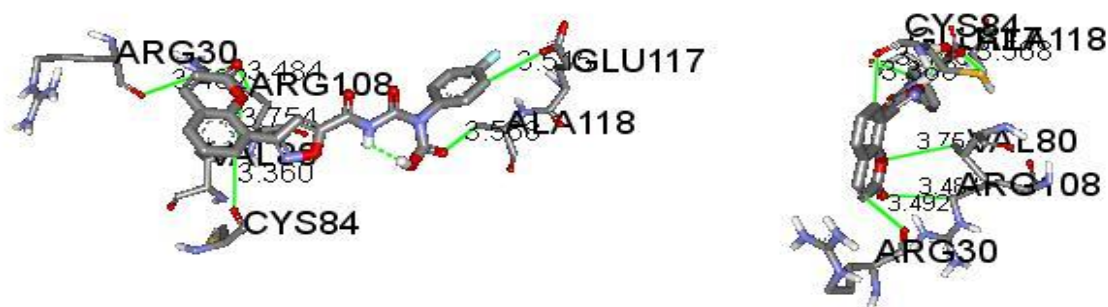
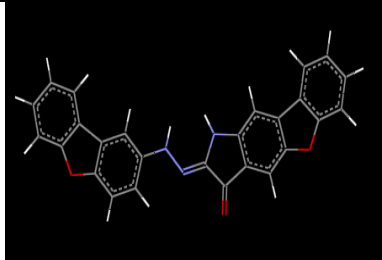
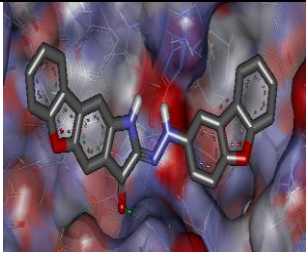

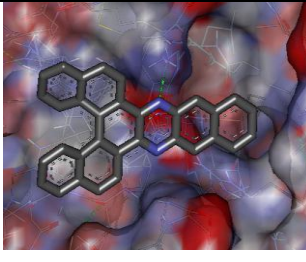
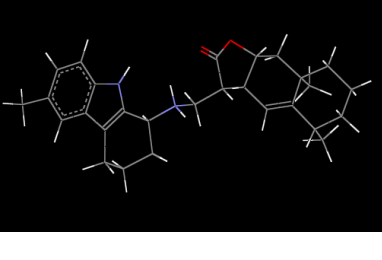
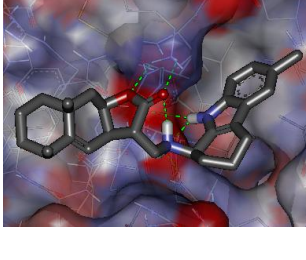
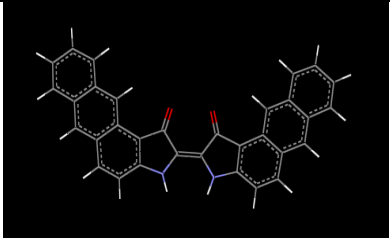
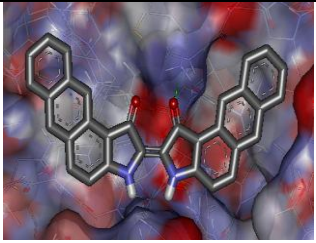
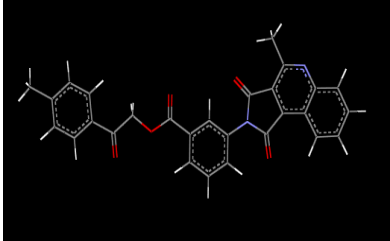
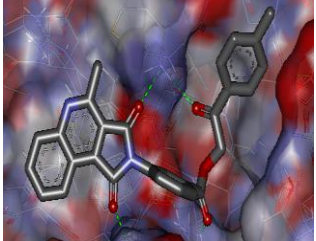
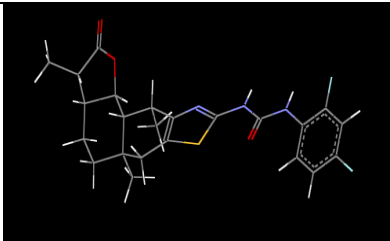
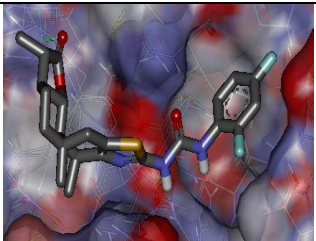
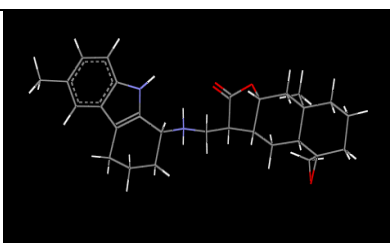
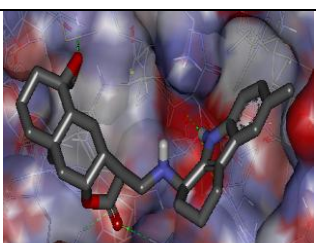
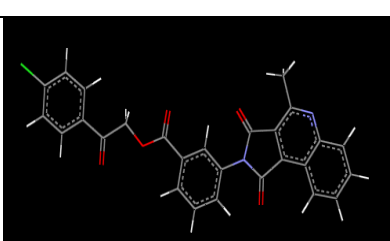
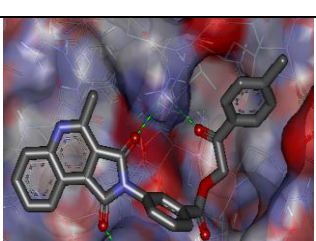


Figure 4.20: C-H---O interaction c-Rel residues and compound **II**

4.6.2 Zinc Natural Products Docking

The second set that is used for virtual screening is natural products or natural product derivatives that are downloaded from ZINC database. The dataset can be found under meta-subsets tab on database site. It contains almost 90000 compounds that are commercially available from seven vendors. The whole dataset is docked to c-Rel with Vina by using the same configuration file from previous fragment-like set dockings. Grid box size is never changed for c-Rel throughout the whole study.

ZINC Code	Binding Energy (kcal/mol)	Structure	Docked Figure
ZINC02121309	-9.7		
ZINC03847098	-9.6		
ZINC08918490	-9.5		

ZINC04265785	-9.5		
ZINC01321694	-9.4		
ZINC03839954	-9.4		
ZINC08952607	-9.3		
ZINC01094583	-9.3		

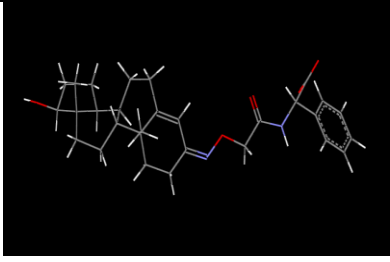
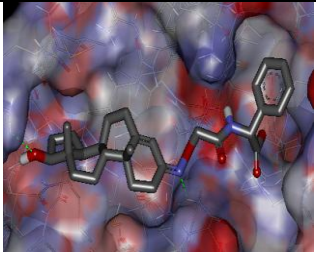
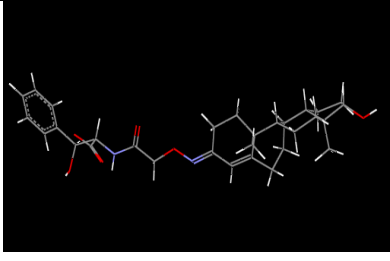
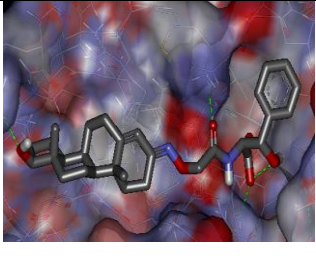
ZINC12659436	-9.3		
ZINC12662663	-9.2		

Table 4.3: Best ten natural compounds bind to c-Rel

Table 4.3 gives top ten lowest binding energies of natural compound dataset to c-Rel. Some of the structures contain too many aromatic rings which might result in lower binding energies. Docking algorithms are also known to be biased towards size of the molecule. Relative size of these molecules is larger; some of them exceed 500 Da, which might also cause in lower binding energies. The best approach to confirm specificity and high affinity of these natural compounds to c-Rel is to perform selectivity study over other Rel family members. Free energy of binding values are listed in Table 4.4. As suspected, binding energies of other NF- κ B transcription factors are generally as lower as c-Rel energies. In fact, top 40 compounds, which have lowest c-Rel binding energy values, are analyzed for their free energy of binding scores over other NF- κ B members and ZINC12659436 is chosen as the most favorable study further on. The compound is catalogued from InterBioScreen Company under catalog ID STOCK1N-20552. The molecular formula of the compound is $C_{29}H_{37}N_2O_5^-$ with molecular mass 493.61 Da, and -1 formal charge. Log P value of the molecule is 3.90 and have two H-bond donor, seven H-bond acceptor atoms.

ZINC #	p50	p52	p65	relB	c-Rel
ZINC02121309	-9.5	-8.7	-8.3	-8.2	-9.7
ZINC03847098	-9.0	-9.5	-7.9	-8.6	-9.6
ZINC08918490	-8.7	-8.6	-7.2	-9.0	-9.5
ZINC04265785	-8.9	-9.6	-8.0	-8.5	-9.5
ZINC01321694	-8.7	-7.8	-7.0	-8.0	-9.4
ZINC03839954	-8.1	-8.6	-7.2	-8.0	-9.4
ZINC08952607	-8.7	-8.5	-7.3	-8.5	-9.3
ZINC01094583	-7.7	-8.3	-6.9	-7.6	-9.3
ZINC12659436	-8.1	-6.4	-6.2	-7.9	-9.3
ZINC12662663	-8.2	-7.8	-5.8	-7.1	-9.2

Table 4.4: Binding energies (kcal/mol) of top ten natural compounds over Rel family

There is no information about the compound in the literature other than chemical properties in various databases. The company offers the molecule as a screening compound. A series of modifications are applied to the molecule including fluorine attachment to phenyl group, keto, hydroxyl and amine group attachments to various sites of the compound to induce H-bonding, and removal of methyl groups to identify excess groups in the compound. These trials to improve binding energy of the compound are unsuccessful. However; removal of phenyl group dramatically increases the binding energy that shows the importance of phenyl group in binding.

The company does not provide any information how the compound is derived or the precursor molecule of the compound; but the presence of steroid moiety (brown box in Figure 4.21) is highly noticeable. Testosterone, a steroid male sex hormone, is highly similar with the mentioned moiety. To improve the binding energy, the steroid moiety and the phenyl group is used to generate new molecules by iLib Diverse program. Three flasks are created in iLib Diverse; steroid moiety (fragment in brown box) is placed in the first

flask, carboxyl group attached to phenyl group (fragment in green box) is placed in the third flask. Carboxyl group is also added because it is seen from the docked pose that oxygen atoms form hydrogen bonds on the groove of binding pocket. The second flask is filled with fragment library of iLib Diverse. The reactivity of carbon atoms that are pointed out with arrows is increased and reactivity of other atoms in the first and third flasks is decreased; so that the attachments only occur in interested carbons. By this way; only the linker region between two groups are changed.

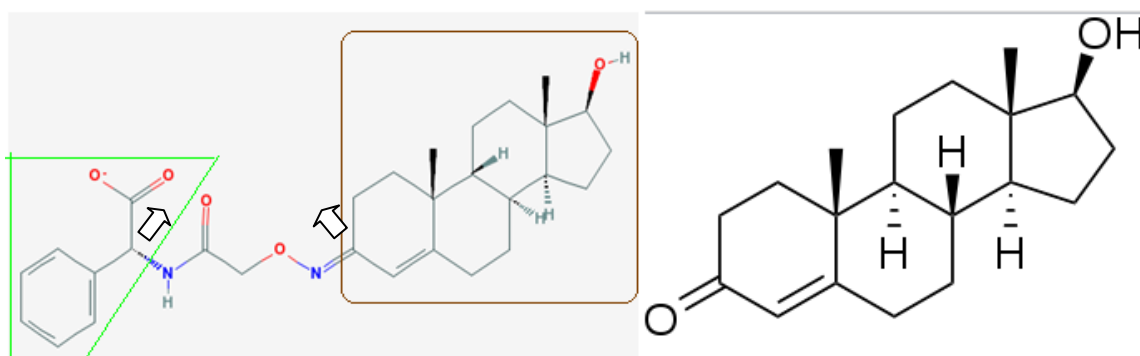


Figure 4.21: Molecular structure of ZINC12659436 (left) and testosterone (right)

400 new compounds are created by combinatorial approach and docked to c-Rel with Vina by using previously used parameters. Best binding energy obtained is -9.8 kcal/mol with a smaller size (compound **III**, Figure 4.22). Compound **III** is formed by linking the steroid moiety and 2-phenyl-acetic acid (carboxyl group linked to phenyl group) with urea group. Urea group takes also place in compound **II**, that might tells us the significance of urea as well as phenyl group in c-Rel docking studies. Compound **III**, with molecular formula $C_{28}H_{38}N_2O_4$, has a mass of 466.61 Da with a logP value of 3.9. It has four hydrogen bond acceptor and four hydrogen bond donor sites.

Selectivity study also performed on compound **III**, it is seen that compound **III** binds more specifically to c-Rel than ZINC12659436. The binding energy difference

between c-Rel and next lowest binding energy is 1.2 kcal/mol for ZINC12659436 (Table 4.4); whereas it is 1.6 kcal/mol for compound **III**.

	p50	p52	p65	relB	c-Rel
Binding Energy (kcal /mol)	-7.6	-7.9	-8.1	-8.2	-9.8

Dissociation constants are also calculated for compound **III** and ZINC12659436 to indicate the improvement on binding affinities. The calculations, as mentioned previously, show that K_i value for ZINC12659436-c-Rel complex is 276.96 nM and 122.99 nM for compound **III**-c-Rel complex that manifest more than two-fold increase in binding affinity. K_i or IC_{50} value for compound **III**-relB complex is 1.65 μ M that infers compound **III** binds more than 13-fold stronger to c-Rel.

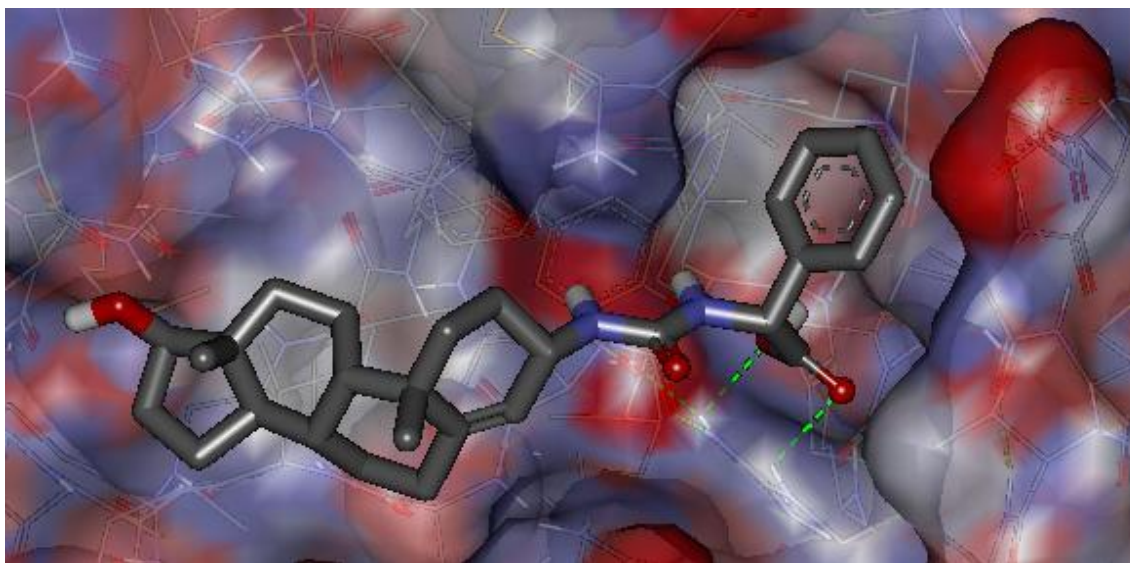


Figure 4.22: Docked pose of compound **III**

Compound **III** is also docked with Autodock 4.2 with the same parameters that are introduced with compound **II**. Figure 4.23 shows the superimposed image of binding poses that are generated from Autodock (green molecule) and Vina (yellow molecule). The poses seem to fit each other, such that RMSD of heavy atoms are calculated as 0.474 Å. The binding energy is estimated as -8.73 kcal/mol; the binding pose and energy agreement of Autodock and Vina is higher for compound **III** than compound **II**.

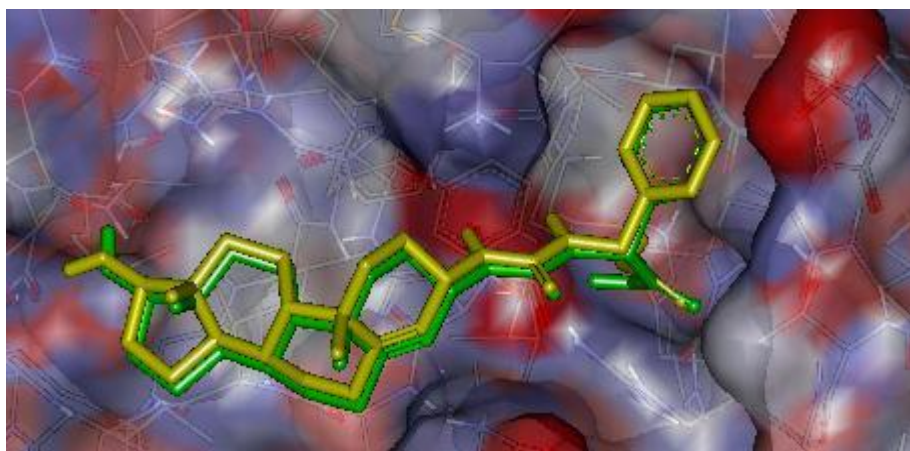


Figure 4.23: Superimposition of Vina and Autodock docked poses of compound **III**

Compound **III** forms hydrogen bonds with Pro76, Asp78 and Arg122 residues (Figure 4.24). Oxygen atom of carboxyl group of Pro76 interacts with hydrogen atom of carboxyl group of compound **III**, oxygen atom of side chain carboxyl group of Asp78 interacts with hydrogen atom of carbamide (urea) group of compound **III** and guanidine group hydrogens of Arg122 form hydrogen bonds with carboxyl group oxygens. CH--- π interaction between methyl group of Ala118 and phenyl ring of compound **III** also takes role in binding (Figure 4.25). There are also more weaker C-H---O interactions between side chain of Arg30 and cyclopentane and -OH group of steroid part of compound **III**, carboxyl oxygen of Cys84 and cyclohexane ring of steroid part, main chain carbon of

Gly81 and cyclopentane ring of steroid part, side chain carboxyl oxygen of Glu114 and phenyl group of compound **III**, side chain carboxyl oxygen of Glu117 and phenyl group,

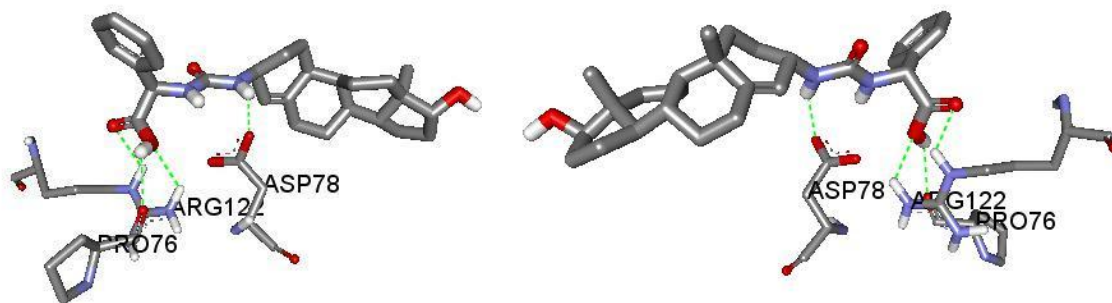


Figure 4.24: Compound **III** and hydrogen bonding residues

methyl group of Ala118 and phenyl group, and lastly side chain carbon of Arg122 and carboxyl oxygen of compound **III** (Figure 4.25).

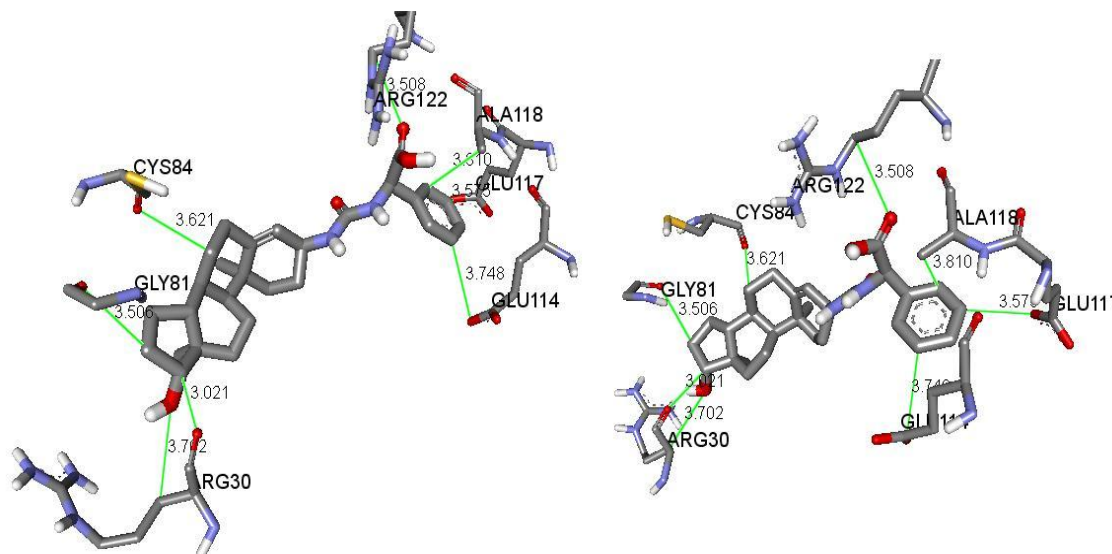


Figure 4.25: C-H...O and C-H...II interactions c-Rel residues and compound **III**

4.6.3 Amino Acid Derived Drug Discovery

There are numerous amino acid derived drugs in the market including valaciclovir (valine derivative), tetrazolyglycine (glycine), lisinopril (proline), dihydroxyphenylglycine (phenylalanine) and acetyl-leucine (leucine). The third approach in drug design against c-Rel transcription factor is based upon D- and L- enantiomeric amino acids. 22 proteinogenic amino acids (20 standard amino acids, and also genetic-coded pyrrolysine and selenocysteine) with all the possible stereoisomers result in 51 different structures. These structures are drawn with MarvinSketch tool and minimized with Dreiding-like forcefield that is implemented in DS Visualizer program. The molecules are docked to c-Rel and the ones with best binding scores are docked to other NF- κ B proteins. Table 4.5 lists top five ligands according to their binding energies against c-Rel.

	c-Rel	p50	p52	p65	relB
D-TYR	-5.5	-5.5	-5.6	-5.1	-5.7
D-TRP	-5.5	-6.2	-5.7	-5.4	-6.5
2S-3R-L-PYL	-5.4	-5.5	-5.6	-5.0	-5.5
L-TRP	-5.4	-5.8	-5.7	-5.4	-5.9
D-PHE	-5.4	-5.5	-5.3	-4.8	-5.5

Table 4.5: Free energy of binding (kcal/mol) values of best five amino acids

As it is seen on the table, none of the amino acids specifically bind to c-Rel even some of them bind stronger to p50 and relB than c-Rel. It is hard to choose which amino acid as the starting fragment for modification analysis. D-phenylalanine is the smallest among five of them and the binding energy difference between c-Rel and other members is

one of the least for D-PHE; therefore D-PHE is picked for combinatorial chemistry studies. 200 new structures are created with iLib Diverse and docked with Vina. The best binding energy that is obtained is -7.6 kcal/mol with benzothiadiazine dioxide attached to amino terminal of D-PHE, compound **IV** (Figure 4.26).

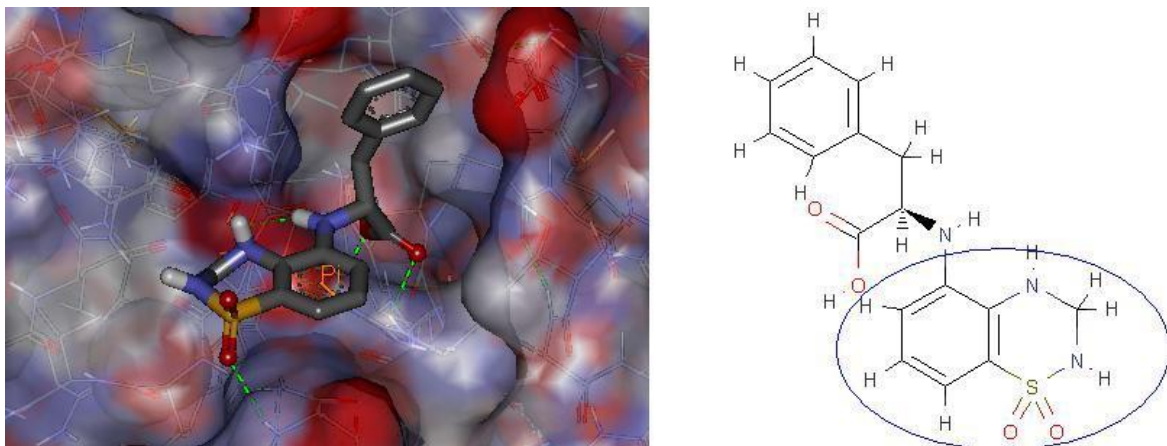


Figure 4.26: 2D structure of modified D-PHE and its docked figure

Benzothiadiazine dioxide is a bicyclic benzene derivative with the heterocycle containing two nitrogens, two oxygens and one sulphur atom. It is known to have antiviral activity and its derivatives have been focus of research for therapeutic implications in viral infections. Compound **IV** with molecular formula $C_{16}H_{17}N_3O_4S$ has molecular weight 347.39 Da, four H-bond donor atoms, six H-bond acceptor atoms with logP value 1.39. However; compound **IV** has also shown affinity to other Rel proteins, as reported below.

	p50	p52	p65	relB	c-Rel
Binding Energy (kcal /mol)	-6.5	-6.7	-7.0	-7.2	-7.6

Another round of modification is performed to increase the ligand's specificity to c-Rel and also lower the binding energy. 250 new compounds are derived from compound **IV** with the help of iLib Diverse and docked to c-Rel with Vina using the same settings. Compound **V** is determined as not only just considering binding energies to c-Rel but also binding energies to other NF- κ B proteins are kept in mind. Binding energy of compound **V** is not the lowest one among 250 compounds, in fact fourth lowest one; but it is the most specific one to c-Rel among top ranked ligands.

	p50	p52	p65	relB	c-Rel
Binding Energy (kcal /mol)	-7.5	-8.0	-7.2	-8.1	-9.1

Compound **V** is generated by attaching phthalazine group to nitrogen atom of benzothiadiazine group (Figure 4.27). Phthalazine derived compounds also have biological activities such as anti-inflammatory, anticonvulsant, anticancer and antimicrobial activities. The molecular weight of compound **V** is 475.52 Da with 2.66 log P value, three H-bond donor and eight H-bond acceptor atoms.

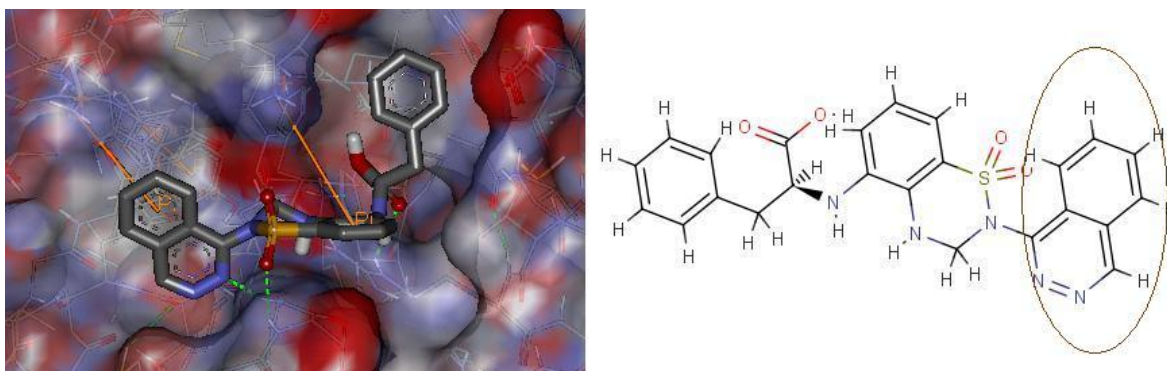


Figure 4.27: 2D structure of compound **V** and its docked pose

IC_{50} value of phenylalanine-derived compound **V** is also determined. According to previous formulations, 383.24 nM compound **V** is required to at least occupy 50% of c-Rel binding sites. Dissociation constant for relB is 1.94 μ M which is approximately five-fold higher than c-Rel. This binding ratio between relB and c-Rel is not remarkable; however developing this much specific molecule to c-Rel from promiscuous D-PHE is still a success for the amino acid derived drug.

Compound **V** forms three Hydrogen bonds with Asp86, Gly87 and Arg122 residues (Figure 4.28). Amino hydrogen of Gly87 interacts with nitrogen atom of phthalazine group, amino hydrogen of Asp86 interacts with one of the oxygen atoms of benzothiadiazine dioxide group and side chain guanidinium hydrogen of Arg122 interacts with carboxy oxygen of phenylalanine moiety of compound **V** to form H-bonds.

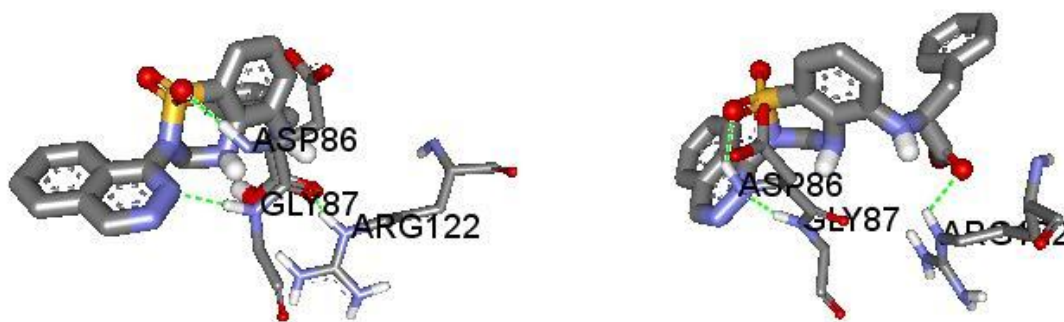


Figure 4.28: Compound **V** and hydrogen bonding residues

There are also cation- Π interactions specific to compound **V** (Figure 4.29). Cation- Π interactions that occur within 6 \AA have significant effect on stability of proteins like hydrogen bonding and salt bridges. Carbon atoms are more electronegative than hydrogens and this fact leads to a build-up of negative charge in the center of aromatic ring and

positive charge around the edge. Cations with their positive charge are electrostatically attracted to the face of ring that results in the phenomenon known as cation- π interaction. Side chain of Arg108 interacts with benzene ring of phthalazine group and side chain of Lys26 interacts with benzene ring of benzothiadiazine dioxide group.

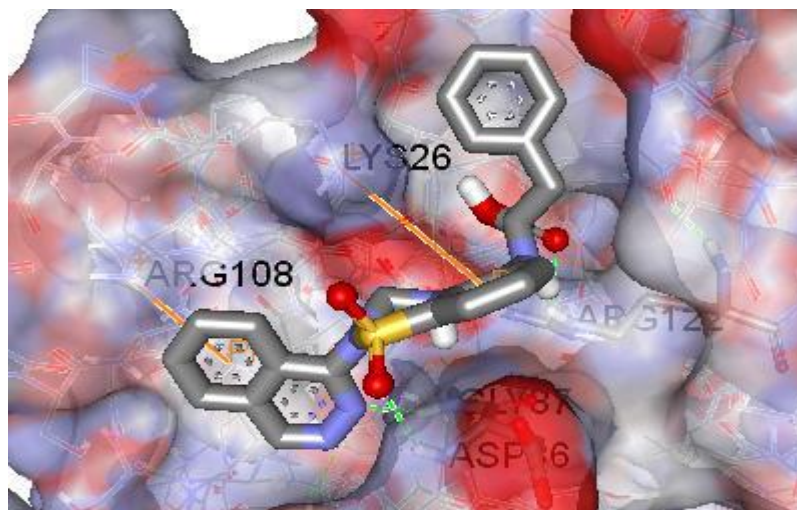


Figure 4.29: Cation- π interactions between compound **V** and positively charged residues

Compound **V** forms CH--- π and C-H---O interactions with surface residues of c-Rel (Figure 4.30). Alpha carbon of Asp86 residue (represented as yellow sphere) forms CH--- π interaction with benzene ring of benzothiadiazine dioxide group. C-H---O interactions are the weakest non-covalent interactions but more abundant ones between compounds and c-Rel. Side chain carboxyl oxygen of Asp78 interacts with carbon atom of benzothiadiazine dioxide ring, carboxyl oxygen of Cys84 interacts with carbon atom of phthalazine ring, carboxyl oxygen of Asp86 interacts with benzene ring carbon of benzothiadiazine dioxide group, side chain carbon of Val110 interacts with carboxyl oxygen of D-PHE, side chain carboxyl oxygen of Glu114 interacts with benzene ring carbon of D-PHE and side chain carboxyl of Glu117 interacts again with benzene ring carbon of D-PHE to form C-H---O interactions.

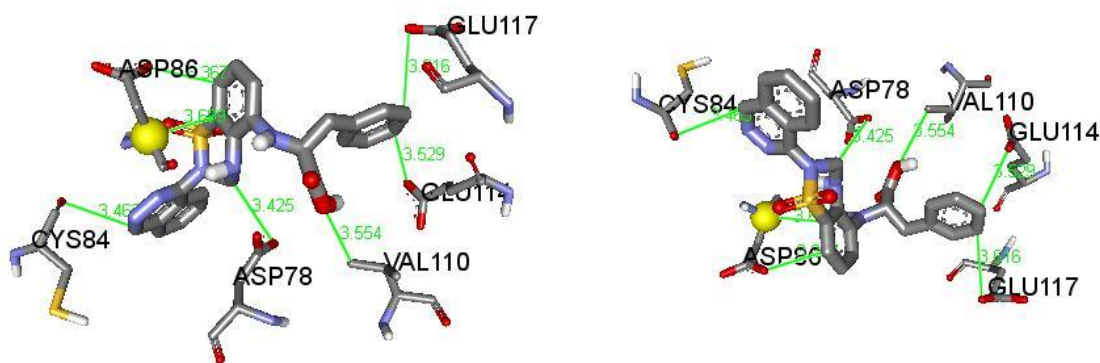


Figure 4.30: C-H...O and C-H... Π interactions c-Rel residues and compound **V**

Compound **V** is also docked with Autodock 4.2 like previous compounds. The binding poses for Vina (yellow one) and Autodock are far from each other (Figure 4.28). Autodock estimates the lowest binding energy of compound **V** as -7.61 kcal/mol. RMSD of heavy atoms of these poses is calculated 8.178 Å. There is also “epdb” command in Autodock that calculates the binding energy of a specific pose. The binding pose that is estimated by Vina is used an input for epdb command and Autodock estimates the free energy of binding of Vina’s pose as -5.53 kcal/mol. This discrepancy in binding energies and poses is confusing whether compound **V** has good potential against c-Rel as a drug.

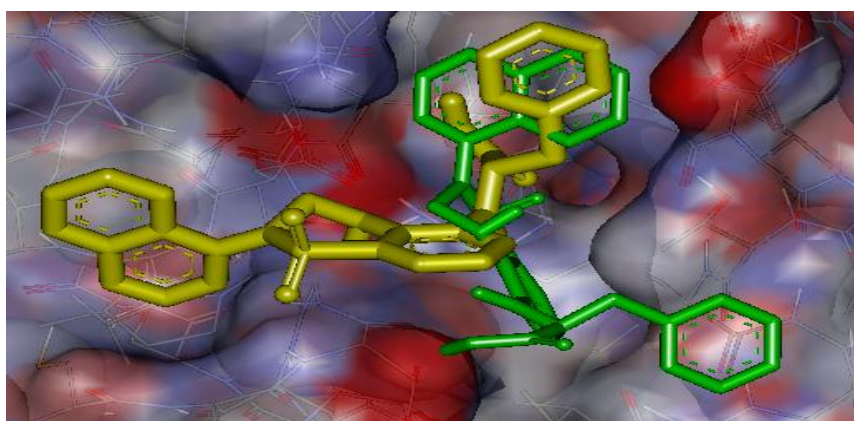


Figure 4.31: Superimposition of Autodock and Vina docked poses

Chapter 5

CONCLUSIONS

This study is concluded with offering three molecules as a drug for c-Rel inhibition. The reason of studying c-Rel is having a direct oncogenic activity of c-Rel on B cell lymphomas. The secondary reason is avoiding serious side effects of NF- κ B inhibition. Many of the generic functions of NF- κ B are carried out by p50/p65 heterodimer and c-Rel is the most specialized subunit. 3D structure of human c-Rel has not been determined yet; so we start the analysis by modeling from pre-determined chicken c-Rel structure. Four different template based modeling servers are used and the output which is the closest to c-Rel is chosen to further study on. The reason behind choosing Phyre result is nearly 80% identity between two cross-species proteins and the relation between structure and sequence.

Docking simulations are carried out by new docking software Vina. It remarkably increases the speed of docking and some reports show that prediction of true binding modes is better than Autodock. However; the resultant molecules are also docked with Autodock 4.2 and binding modes of compound **II** and compound **III** are nearly the same. Free energy of binding varies between Autodock and Vina with Vina comes up with lower binding energies. Still, the binding energies that Autodock estimated are considerably low for compound **II** and **III**. Estimated binding poses for compound **V** do not correlate and selectivity of compound **V** to c-Rel is not promising that shows us not a good drug potential for compound **V**. The selectivity docking is performed after screening and modification studies to see how the drug is specific to c-Rel. Compound **V** also points out how the selectivity for lead compound is important. The modification of promiscuous molecules does not result in high selectivity in the case of highly homologous protein family.

When compound **II** and compound **III** are scrutinized, the presence of some groups is important for binding interactions between protein and drug. Phenyl group in a channel-like region of the pocket and carboxyl group in the groove of the binding region form hydrogen bonds with c-Rel. The presence of cyclic groups in the plateau-like region of binding site also increases the binding affinity.

Our hypothesis is based upon novel inhibition approach on NF- κ B signaling cascade. Inducing unbinding of NF- κ B from DNA has never been studied before. The unknowns about the termination of NF- κ B signaling, results of GNM and high conservation ratio of Leu79 region among NF- κ B members encourage us to find the missing piece of NF- κ B puzzle. If the inhibition of c-Rel is demonstrated with experimental techniques, such as NF- κ B binding assays or gel shift assays, it might also be a leading example of blocking the activity of other transcription factors. The experimental evidence can also offer a different way for understanding termination of NF- κ B signals in cells.

APPENDIX A

Configuration File of NAMD Minimization

```
#####  
## JOB DESCRIPTION ##  
#####  
  
# Minimization of c-Rel in a Water Sphere  
  
#####  
## EXECUTION SCRIPT ##  
#####  
  
numsteps 10000  
minimization on  
  
#####  
## ADJUSTABLE PARAMETERS ##  
#####  
  
structure phyreDNA_ws.psf  
coordinates phyreDNA_ws.pdb  
  
outputname phyreDNA_ws_min  
  
outputEnergies 100  
binaryoutput no  
restartfreq 1000  
dcdfreq 200  
  
#####  
## SIMULATION PARAMETERS ##  
#####  
  
# Input  
paraTypeCharmm on  
parameters par_all127_prot_na.prm
```

```
# Force-Field Parameters
exclude          scaled1-4
1-4scaling       1.0
cutoff           12.
switching        on
switchdist       10.
pairlistdist     13.5
dielectric       1.0
margin           0.0
stepspercycle    20
```

```
#####
```

BIBLIOGRAPHY

1. Sen R, Baltimore D. Inducibility of kappa immunoglobulin enhancer-binding protein NF-kappa B by a posttranslational mechanism. *Cell* 1986; **47**: 921-928
2. Sethi G, Tergaonkar V. Potential pharmacological control of the NF- κ B pathway. *Trends Pharmacol Sci* 2009; **30**: 313-320
3. Oeckinghaus A, Ghosh S. The NF- κ B family of transcription factors and its regulation. *Cold Spring Harb Perspect Biol* 2009; **1**:a000034
4. Baldwin AS Jr. The NF- κ B and I κ B proteins: New discoveries and insights. *Annu Rev Immunol* 1996; **14**: 649-683
5. Zhong H, MayMJ, Jimi E, Ghosh S. The phosphorylation status of nuclear NF- κ B determines its association with CBP/p300 or HDAC-1. *Mol Cell* 2002 **9**: 625–636
6. Dobrzanski P, Ryseck RP, Bravo R. Both N- and C-terminal domains of RelB are required for full transactivation: Role of the N-terminal leucine zipper-like motif. *Mol Cell Biol* 1993; **13**: 1572–1582.
7. Lin L, Ghosh S. A glycine-rich region in NF-kappaB p105 functions as a processing signal for the generation of the p50 subunit. *Mol. Cell. Biol.* 1996; **16**: 2248-2254
8. Beinke S, Belich MP, Ley SC. The Death Domain of NF- κ B1 p105 is essential for signal-induced p105 proteolysis. *J. Biol. Chem.* 2002; **277**: 24162-24168
9. Wang Y, Cui H, Schroering A, Ding JL, Lane WS, McGill G, Fisher DE, Ding HF. NF- κ B2 p100 is a pro-apoptotic protein with anti-oncogenic function. *Nature Cell Biol.* 2002; **4**: 888-893
10. Ghosh S, May MJ, and Kopp EB. NF- κ B and REL proteins: Evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* 1998; **16**: 225-260
11. Ghosh S, Karin M. Missing pieces in the NF- κ B puzzle. *Cell* 2002; **109**: S81-S96

12. Mortier J, Frederick R, Ganeff C, Remouchamps C, Talaga P, Pochet L, Wouters J, Piette J, Dejardin E, Masereel B. Pyrazolo[4,3-c]isoquinolines as potential inhibitors of NF-kappaB activation. *Biochem. Pharmacol.* 2010; **79(10)**: 1462-1472
13. Baeuerle PA, Baltimore D. IκB: a specific inhibitor of the NF-κB transcription factor. *Science* 1988; **242**: 540-546
14. Huxford T, Ghosh G. A structural guide to proteins of the NF-κB signaling module. *Cold Spring Harb Perspect Biol* 2009; **1**:a000075
15. Schwarz EM, Van Antwerp D, Verma IM. Constitutive phosphorylation of IκBα by casein kinase II occurs preferentially at serine 293: requirement for degradation of free IκBα. *Mol Cell Biol.* 1996; **16(7)**: 3554–3559
16. Hoffmann A, Natoli G, Ghosh G. Transcriptional regulation via the NF-κB signaling module. *Oncogene* 2006; **25**: 6706-6716
17. Wertz IE, Dixit VM. Signaling to NF-κB: Regulation by ubiquitination. *Cold Spring Harb Perspect Biol* 2010; **2**: a003350
18. Hayden MS, Ghosh S. Signaling to NF-κB. *Genes Dev* 2004; **18**: 2195-2224
19. Carmody RJ, Ruan Q, Palmer S, Hilliard B, Chen YH. Negative regulation of Toll-like receptor signaling by NF-κB p50 ubiquitination blockade. *Science* 2007; **317**: 675-678
20. Scheidereit C. IκB kinase complexes: Gateways to NF-κB activation and transcription. *Oncogene* **25**: 6685-6705
21. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science* 2002; **298**:1241-1245
22. Tanaka T, Grusby MJ, Kaisho T. PDLIM2-mediated termination of transcription factor NF-κB activation by intranuclear sequestration and degradation of the p65 subunit. *Nat Immunol* 2007; **8**: 584–591.

23. Maine GN, Mao X, Komarck CM, Burstein E. COMMD1 promotes the ubiquitination of NF-kappaB subunits through a cullin-containing ubiquitin ligase. *EMBO J* 2007; **26**: 436-447
24. Lawrence T, Bebien M, Liu GY, Nizet V, Karin M. IKK α limits macrophage NF-kB activation and contributes to the resolution of inflammation. *Nature* 2005; **434**: 1138–1143
25. Gilmore TD, Herscovitch M. Inhibitors of NF-kB signaling: 785 and counting. *Oncogene* 2006; **25**: 6887-6899
26. Song XR, Torphy TJ, Grisowold DE, Shealy D. Coming of age: anti-cytokine therapies. *Mol. Interv.* 2002; **2**: 36-46
27. Walker EH, Pacold ME, Perisic O, Stephens L, Hawkins PT, Wymann MP, Williams RO. Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol Cell* 2000; **6**: 909-919
28. Karin M, Yamamoto Y, Wang QM. The IKK NF-kB system: a treasure trove for drug development. *Nat Rev Drug Discov* 2004; **3**: 17–26
29. Burke JR, Pattoli MA, Gregor KR, et al. BMS-345541 is a highly selective inhibitor of I κ B kinase that binds at an allosteric site of the enzyme and blocks NF-kB-dependent transcription in mice. *J Biol Chem* 2003; **278**: 1450–1456
30. di Meglio P, Ianaro A, Ghosh S. Amelioration of acute inflammation by systemic administration of a cell-permeable peptide inhibitor of NF-kB activation. *Arthritis Rheum.* 2005; **52**: 951-958
31. Zhou H, Monack DM, Kayagaki N, Wertz I, Yin J, Wolf B, Dixit VM. Yersinia virulence factor YopJ acts as a deubiquitinase to inhibit NF-kB activation. *J Exp Med* 2005; **202**:1327-1332

32. Swinney DC, Xu YZ, Scarafia LE, Lee I, et al. A small molecule ubiquitination inhibitor blocks NF-kappa B-dependent cytokine expression in cells and rats. *J Biol Chem* 2002; **277**: 23573–23581
33. Adams, J. Proteasome inhibition: a novel approach to cancer therapy. *Trends Mol. Med.* 2002; **8**: S49-S54
34. Letoha T, Somlai C, Takacs T, Szabolcs A, et al. A nuclear import inhibitory peptide ameliorates the severity of cholecystokinin-induced acute pancreatitis. *World J Gastroenterol* 2005; **11**: 990-999
35. Umezeawa K. Inhibition of tumor growth by NF-κB inhibitors. *Cancer Sci.* 2006; **97**: 990-995
36. Zhang S, Won YK, Ong CN, Shen HM. Anti-cancer potential of sesquiterpene lactones: bioactivity and molecular mechanisms. *Curr. Med. Chem. Anticancer Agents* 2005; **5**: 239-249
37. Khaled AR, Butfiloski EJ, Sobel ES, Schiffenbauer J. Use of phosphorothioate-modified oligodeoxynucleotides to inhibit NF-κB expression and lymphocyte function. *Clin Immunol Immunopathol* 1998; **86**: 170-179
38. Bergmann M, Hart L, Lindsay M, Barnes PJ, Newton R. IκBα degradation and nuclear factor-κB DNA binding are insufficient for interleukin-1β and tumor necrosis factor-α-induced κB-dependent transcription. Requirement for an additional activation pathway. *J Biol Chem* 1998; **273**: 6607–6610
39. Moulton J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. *Proteins* 1995; **23**: 2-4
40. Lee J, Wu S, Zhang Y. Ab initio protein structure prediction. *From Protein Structure to Function with Bioinformatics* Chapter 1, Edited by D. J. Rigden, (Springer-London, 2009), P 1-26

41. Kihara D, Chen H, Yang YD. Quality Assessment of Protein Structure Models. *Current Protein and Peptide Science* 2009; **10**: 216-228
42. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000; **29**: 291-325
43. Qu X, Swanson R, Day R, Tsai J. A guide to template based structure prediction. *Curr Protein Pept Sci* 2009; **10**: 270-285
44. CASP 8 special issue. *Proteins* 2009; **77 (Suppl. 9)**: 1-228
45. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 1995; **247**: 536-540
46. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000; **16**: 404-405
47. Pollastri G, Przybylski D, Rost B, Baldi P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 2002; **47**: 228-235
48. Cole C, Barber JD, Barton GJ. The JPred 3 secondary structure prediction server. *Nucleic Acids Res.* 2008; **36 (suppl 2)**: W197-W201
49. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for prediction of protein disorder. *Bioinformatics* 2004; **20**: 2138-2139
50. Kelley LA and Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 2009; **4**: 363-371
51. McCammon JA, Gelin BR, and Karplus M. Dynamics of folded proteins. *Nature* 1977; **267**: 585-590
52. Allen MP. An introduction to molecular dynamics. *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, 2004; **23**: 1-28

53. Phillips JC, Braun R, Wang W, *et al.* Scalable molecular dynamics with NAMD. *J. Comp. Chem*, 2005; **26**: 1781-1802
54. Bahar I, Atilgan AR and Erman B. Direct evaluation of thermal fluctuations in protein using a single parameter harmonic potential. *Folding Des* 1997; **2**: 173-181
55. Flory PJ Statistical thermodynamics of random networks. *Proc. R. Soc. Lond. A*, 1976; **351**: 351-380
56. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, 2001; **80**: 505-515
57. Chennubhotla C, Rader AJ, Yang LW, Bahar I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.*, 2005; **2**: 173-180
58. Rader AJ, Chennubhotla C, Yang LW, Bahar I. The Gaussian Network Model: Theory and applications. In "Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems" Eds Qiang Cui and I Bahar, Chapman & Hall / *CRC Mathematical and Computational Biology Series*, 2006, 41-64.
59. Haliloglu T, Seyrek E, Erman B. Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model. *Phys. Rev. Lett.*, 2008; **100**: 228102
60. Teague SJ. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2003; **2**: 527-541
61. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. *Proteins* 2006; **65**: 15-26
62. Bohm HJ. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* 1992; **6**: 61-78
63. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996; **261**: 470-489

64. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select “quasi-flexible” ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* 1994; **8**: 153–174
65. Abagyan R, Totrov M, Kuznetsov D. ICM – a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 1994; **15**: 488-506
66. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. *Proteins* 1992; **13**: 206–222
67. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997; **267**: 727–748
68. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009; **30**: 2785-2791
69. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* 1998; **33**: 367–382
70. Weiner PK, Kollman PA. AMBER—assisted model building with energy refinement—a general program for modeling molecules and their interactions. *J Comput Chem* 1981; **2**: 287–303
71. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983; **4**: 187–217
72. van Gunsteren WF, Berendsen HC. Computer simulations of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew Chem Int Ed* 1990; **29**: 992–1023

73. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J. Med. Chem.* 2006; **49**: 6789-6801
74. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Prot. Struct. Funct. Genet.* 1990; **8**: 195–202
75. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998; **19**: 1639–1662
76. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculation by fast computing machines. *J Chem Phys* 1953 **21(6)**: 1087–1092
77. Huey R, Morris GM, Olson AJ, Goodsell DS: A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 2007; **28**: 1145-1152
78. Mehler EL and Solmajer T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Engineering* 1991; **4**: 903-910
79. Stouten PFW, Frommel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. *Mol Simulations* 1993; **10**: 97–120
80. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comp Chem* 2010; **31**: 455-461
81. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. *PLoS ONE* 2010; **5(8)**: e11955
82. Wang R, Lai L, Wang S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput. Aided Mol. Des.* 2002; **16**: 11-26
83. Nocedal J, Wright SJ. Numerical Optimization *Springer-Verlag* 2006; **2nd ed**: 193-201

84. Carrasco D, Weih F, Bravo R. Developmental expression of the mouse c-rel proto-oncogene in hematopoietic organs. *Development* 1994; **120**: 2991–3004
85. Gerondakis S, Grossman M, Nakamura Y, Pohl T and Grumont R. Genetic approaches in mice to understand Rel/NF- κ B and I κ B function: transgenics and knockouts. *Oncogene* 1999; **18**: 6888-6895
86. Armitage JO. How I treat patients with diffuse large B-cell lymphoma. *Blood* 2007; **110**: 29-36
87. Rao PH, Houldsworth J, Dyomina K, Parsa NZ, Cigudosa JC, Louie DC, Popplewell L, Offit K, Jhanwar SC and Chaganti RSK. Chromosomal and Gene Amplification in Diffuse Large B-Cell Lymphoma. *Blood* 1998; **92**: 234-240
88. Joos S, Otano-Joos MI, Ziegler S, Brüderlein S, du Manoir S, Bentz M, Möller P and Lichter P. Primary mediastinal (thymic) B-cell lymphoma is characterized by gains of chromosomal material including 9p and amplification of the REL gene. *Blood* 1996; **87**: 1571-1578
89. Goff LK, Neat MJ, Crawley CR, Jones L, Jones E, Lister TA, Rupta GK. The use of real-time quantitative polymerase chain reaction and comparative genomic hybridization to identify amplification of the REL gene in follicular lymphoma. *Br J Haematol* 2000; **111**: 618-625
90. Mukhopadhyay T, Roth JA and Maxwell SA. Altered expression of the p50 subunit of the NF- κ B transcription factor complex in non-small cell lung carcinoma. *Oncogene* 1995; **11**: 999–1003
91. Sovak MA, Bellas RE, Kim DW, Zanieski GJ, Rogers AE, Traish AM, Sonenshein GE. Aberrant nuclear factor- κ B/Rel expression and the pathogenesis of breast cancer. *J Clin Invest* 1997; **100**: 2952–2960

92. Tian W, Liou HC. RNAi-Mediated c-Rel Silencing Leads to Apoptosis of B Cell Tumor Cells and Suppresses Antigenic Immune Response In Vivo. *PLoS ONE*. 2009; **4(4)**: e5028
93. Huang DB, Chen YQ, Ruetsche M, Phelps CB, Ghosh G. X-Ray crystal structure of proto-oncogene product c-Rel bound to CD28 response element of IL-2. *Structure* 2001; **9**: 669-678
94. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 2003; **31(13)**: 3497–3500
95. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 2010; **5**: 725-738
96. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005; **33**: W244-248
97. Zhou H, Skolnick J. Protein structure prediction by pro-sp3-TASSER. *Biophys. J.* 2009; **96(6)**: 2119-2127
98. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 1991; **9**: 56-68
99. Randall AZ, Baldi P, Villarreal LP. Structural proteomics of the poxvirus family. *Artif Intell Med* 2004; **31**: 105-115
100. MacKerell Jr. AD, Feig M, Brooks III CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 2004; **25(11)**: 1400-1415
101. Humphrey W, Dalke A., Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* 1996; **14**: 33-38

102. Haliloglu T, Gul A, Erman B. Predicting important residues and interaction pathways in proteins using Gaussian Network Model: binding and stability of HLA proteins. *PLoS Comput Biol* 2010; **6(7)**: e1000845
103. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* 1997; **23**: 3-25
104. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005; **45(1)**: 177-182
105. Wolber G, Langer T. CombiGen: A novel software package for the rapid generation of virtual combinatorial libraries. In H.-D. Höltje and W. Sippl, editors, *Rational approaches to drug design*, 2000. **Prous Science**: 390-399
106. Wu L, Wang X, Xu W, Farzaneh F, Xu R. The structure and pharmacological functions of coumarins and their derivatives. *Curr Med Chem* 2009; **16(32)**: 4236-4260
107. Gu Y, Kar T, Scheiner S. Fundamental Properties of the CH...O Interaction: Is It a True Hydrogen Bond? *J. Am. Chem. Soc.* 1999; **121(40)**: 9411-9422

VITA

Mümin Öztürk was born in Kardzhali, Bulgaria on October 31, 1986. His family moved to İzmir, Turkey in 1989. He attended Bornova Anadolu Lisesi (B.A.L) for secondary education. He received his Bachelor of Science degree in Molecular Biology and Genetics Department, from Bilkent University, Ankara in June 2008. He worked as a teaching and research assistant in Chemical and Biological Engineering Department at Koç University from 2008 to 2010. His research covers designing small molecule inhibitors against c-Rel transcription factor.

Contact e-mail: mumin.ozturk@gmail.com