**Resubmitted**

# QSAR Analysis, Docking & Molecular Dynamics Simulation of Several Antimicrobial Compounds and Receptors

Thesis submitted for the degree of Doctor of Philosophy in science

(Chemistry) under the

University of North Bengal



By

**Biswajit Das**

Supervisor

**Dr. Asim K. Bothra**

Co-supervisor

**Dr. Pranab Ghosh**

**DEPARTMENT OF CHEMISTRY**

**RAIGANJ COLLEGE (UNIVERSITY COLLEGE)**

RAIGANJ, UTTAR DINAJPUR

**2015**

*Dedicated*

*To*

*My parents*

# DECLARATION

I hereby declare that the research work embodied in this thesis entitled "QSARanalysis, docking & molecular dynamics of several antimicrobial compounds and receptors" has been carried out in the Department of Chemistry, RaiganjCollege (University College) & Department of Chemistry, University of North Bengal, under the joint supervision of Dr. Asim K. Bothra, Associate Professor, Raiganj College (University College) and Dr. Pranab Ghosh, Associate Professor, Department of Chemistry, University of North Bengal. No part of this thesis has formed the basis for the award of any degree or fellowship previously.

Biswajit Das

**Biswajit Das**

# CERTIFICATE

We certify that the thesis entitled, "QSAR analysis, docking & molecular dynamics of several antimicrobial compounds and receptors" submitted by Mr. Biswajit Das for the award of Ph.D degree from Department of Chemistry of Raiganj College (University College), embodies the record of the original investigation carried by him under our supervision. He has been duly registered and the thesis presented is worthy of being considered for the award of Doctor of Philosophy (Science) degree in chemistry. The work has not been submitted for any degree of this or any other university and regulations of the University of North Bengal.

(Dr. Asim K Bothra)

Supervisor

Department of Chemistry

Raiganj College (University College)

DR. ASIM K. BOTHRA
ASSOCIATE PROFESSOR
DEPT. OF CHEMISTRY
RAIGANJ COLLEGE (UNIVERSITY COLLEGE)

Date: 29.06.2015

(Dr. PranabGhosh)

Co- Supervisor

Department of Chemistry

University of North Bengal

Associate Professor
Department of Chemistry
University of North Bengal

Date: 29/6/2015

# ACKNOWLEDGEMENT

I wish to express my deep gratitude to a number of people who giving me guidance, help and support to reach my goal of this thesis.

First of all, I take this opportunity to express my hearty gratitude & sincere thanks to my research supervisor, Dr. Asim Kumar Bothra, Associate Professor, Department of Chemistry, Raiganj College (University College) for his mature, able & invaluable guidance & persistence encouragement. No word can express my deep sincere of reverence for him.

I owe heartfelt thanks to Dr. PranabGhosh, Associate Professor, Department of Chemistry, North Bengal University for his constant support and guidance throughout my research work. I am extremely thankful to him for this invaluable suggestions and advices.

I am thankful to the Head of the Department and all the teachers and staff of the Department of Chemistry, North Bengal University for their inspiration.

I want to express my respect and thanks to the Head of the Department and all the teachers and staff of the Department of Chemistry, Raiganj College (University College), for their help and moral supports during my work.

I am also thankful to Teacher-in-Charge and all the staff of the Raiganj College(University College)for supporting me to complete the work.

I express my deep gratitude, respect and sincere thanks to Dr.SubhasisMukherhjee, Department of Biophysics, Molecular Biology and Bioinformatics University of Calcutta for his help, instructions and support during my work.

I wish to express my warm and sincere thanks to my friends and lab mates Dr. Uttam Kr. Mondal, Shymal Sharma, Dr. AbhikChaterjee, BhaskarBagchi, Ayan Pal, Dipanjan Das for offering me assistance.

I would like to express my gratefulness to all my colleagues who have directly or indirectly helped me to present this work.

Finally I would like to thank my family membersfor their constant support and encouragement. I also thanks to my wife who has providedmoral and emotional support, great encouragement during this work.

# SUMMARY

In this thesis quantitative structure-activity relationship (QSAR), molecular docking and molecular dynamics simulation were performed.

This thesis contains total nine chapters. First and second and third chapter describe introduction, review of this work, Materials and Methods respectively.

Chapter four comprises MD simulation of rennin and molecular docking of its inhibitors. Molecular docking studies were carried out with AutoDock 4.2. and molecular dynamics simulation was performed using GROMACS. Inhibitor 72X was successfully docked into the active site of human renin. There are many hydrogen bonds formed between inhibitor and protein. It is observed that Asp226 and Gly228 residues are important for binding. All docking results show that hydrogen bond formed between P1 moiety and Asp226 of 72X. Also Gly228 makes a hydrogen bond with the P3' moiety of the 72X. A careful inspection of the binding pocket indicated that the inhibitor in hydrophobic cage surrounded by mainly hydrophobic residues of renin. It is examined that two aspartic acid residues Asp38 and Asp226 placed at the P1 moiety and maximum number residues in chain A involve in the binding process whereas only two residues Leu252 (B) and Phe253 (B) of chain B in the binding process. Region having amino acids Asp226, Gly228 and Ser230 are most important residues for initiating the interaction with ligand and good binding.

Chapter five describes molecular dynamics simulation of chick Type IIa receptor protein tyrosine phosphatases sigma to understand the folding and structural behavior of cheek RPTP σ. Molecular dynamics simulation is performed using software GROMACS to understand the motional properties of the protein. Root Mean Square Deviation (RMSD), Radius of gyration (Rg) and Principal Component Analysis(PCA) has been carefully

done. The residues of binding site are also studied extensively. From the analysis it is clear the motion of the protein is distrusted among the PCAs. Hydrogen bonds formed between the hydroxyl groups of SER50 and TYR216 (HB6). Another hydrogen bond formed between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB7) in Ig1-Ig2 pro-rich loop. It was found that hydrogen bond between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB6) remain intact during the whole simulation time.

Chapter six describes the QSAR studies of Hydroxamate inhibitors of Anthrax Lethal toxin. Constructed QSAR models by stepwise regression analysis is developed using quantum chemical descriptors. After considering training set and test set we have also designed seventeen compounds and predicted their activity. This work may be helpful in screening and synthesis of Anthrax inhibitor.

Chapter seven comprises molecular dynamics simulation of human bifunctional glutamyl-prolyl-tRNA synthetase to understand the motional properties and mode of action of the human bifunctional glutamyl-prolyl-tRNA synthetase. Molecular dynamics simulation of human bifunctional glutamyl-prolyl-tRNA synthetase, in aqueous environment was carried out using the software, GROMACS. From the time evolution, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) and Radius of gyration (Rg), it was found that the toxin was relatively flexible. Principal Component Analysis (PCA) was also performed for better understanding of motional properties in reduced dimension. All these observations help us to understand the structure and function of human bifunctional glutamyl-prolyl-tRNA synthetase

Chapter eight describes screening of Triazine derivatives (MAP-kinase inhibitors) through mathematical modeling and molecular Modeling. Triazine derivatives possess

various pharmacological actions against breast, lung, and ovarian cancers.These derivatives bind with MAP-kinase p-38. In this work four QSAR models were developed using 16 compounds and its predictive ability was assessed using a test set of 9. We computed several graph theoretical indices along with quantum chemical parameters and constructed regression equations. We have also designed a series of triazine derivatives and predicted their activity. We intend to suggest some compounds, which have high predicted activity.

In the last chapter describe molecular docking of Triazine analogues. Docking of MAP inhibitors were performed using AutoDock and binding energy for the inhibitors are calculated and regression equation is formed using HT29. Effect of substitution is analyzed. It is found presence of morpholino or anilino ring is essential for inhibition. Some compounds are designed and their binding energy is calculated. It is seen that designed compound also good binding energy and inside the binding pocket of MAP-kinase p-38.

# LIST OF PUBLICATIONS

1. Biswajit Das, Abhik Chatterjee, Uttam Kumar Mondal, Pranab Ghosh, Asim Kumar Bothra. Molecular Docking and MD simulation of human renin: implication on its binding site. Manuscript communicated to Current Protein and Peptide Science (under review)

2. Biswajit Das, Uttam Kumar Mondal, Pranab Ghosh, Asim Kumar Bothra. Molecular dynamics simulation of chick Type IIa receptor protein tyrosine phosphatases. International Journal of Integrative Biology, 2013, Vol. 14(1), 1-6.

3. Biswajit Das, Uttam Kumar Mondal, Shyamal Sharma, Pranab Ghosh, Asim Kumar Bothra. Molecular Docking of Triazine analogues. J. Chem. Pharm. Res, 2012, 4(3), 1595-1600.

4. Biswajit Das, Uttam Kumar Mondal, Pranab Ghosh, Asim Kumar Bothra. Molecular dynamics simulation of human bifunctional glutamyl-proyl-tRNA synthetase. J. Chem. Pharm. Res., 2011, 3(4), 964-973

5. Biswajit Das, PranabGhosh, Asim Kumar Bothra. QSAR studies of Anthrax Lethal inhibitors through QuantumChemical Indices. J. Chem. Pharm. Res., 2011, 3(5), 443-449.

6. Biswajit Das, Shyamal Sharma, Pranab Ghosh, Subhasis Mukherhjee, Asim Kumar Bothra. Screening of Triazine Derivatives, Inhibitors of MAP-kinase p-38 Alpha, Through Mathematical Modeling and Molecular Modeling. The IUP Journal of Chemistry, 2009, 2(4), 7-39.

# CONTENTS

# Chapter I

# Introduction

## 1.1: Brief History of QSAR

Quantitative structure-activity relationships represent an attempt to correlate structural or property descriptors of compound with activities. The QSAR modeling is to derive empirical models that relate the biological activity of compounds to their chemical structure. The biological potency of a compound may be revealed by its toxicity or activity.

In 1863 A.F.A. Cros introduced a relationship which existed between the toxicity of primary aliphatic alcohols with their water solubility [1]. Crum-Brown and Fraser (1968) expressed the idea that the biological activity of alkaloids ($\Phi$) was a function (f) of their molecular constitution (C) [2].

$$\Phi = f(C) \qquad 1$$

This equation is considered to be the first general formulation of a Quantitative structure-activity relationship.

In 1893, Richet correlated toxicities of alcohols, ethers, and ketones with their corresponding water solubility and suggested that their toxicities are inversely related to their water solubilities [3]. This relationship showed in equation 2, where $\Delta\Phi$ are the differences in biological activity values and their corresponding changes in the chemical and especially the physicochemical properties, $\Delta C$.

$$\Delta\Phi = f(\Delta C) \qquad 2$$

Now only the differences in biological activities are quantitatively correlated with changes in lipophilicity and other physicochemical properties of the compounds under investigation. Therefore it is assumed that all QSAR equations correspond to eq. 2.

Meyer (1899) and Overton (1901) independently suggested that the liner relationship between the narcotic action of a group of organic compounds and their olive oil/water partition coefficients [4-5].

In 1935 Louis Hammett introduced a method to account for substituent effects on reaction mechanisms through the use of an equation which took two parameters into consideration namely the (i) substituent constant and the (ii) reaction constant [6-7]. Hammett correlated electronic properties of a substituent on benzoic acid with the dissociation constants and reactivity. On the basis of free energy relationships he proposed a relation while studying the dissociation constant of substituted benzoic acid and its benzoate ion and postulated electronic σ-ρ constant. Hammett observed that a linear relationship resulted when substitutions of different groups were made to benzoic acid.

$$\text{Log } K/K_0 = \rho \text{ Log } K'/K'_0 = \rho\, \sigma 3$$

Where $K_0$ and $K_0'$ represents equilibrium constants for unsubstituted compounds where K and K' for substituted compounds. The σ denotes the properties of the substitution groups i.e. descriptor. If an electron withdrawing group attached to the benzene ring, σ is positive. But the presence of an electron donating group, σ is negative. The magnitude of σ indicates the degree of these effects. If the effect of substituents is proportionally larger than on the benzoic acid equilibrium, then $\rho > 1$, but the effect is less than on the benzoic acid equilibrium, $\rho < 1$. By definition, for benzoic acid is equal to 1.

In 1939 Ferguson correlated depressant action with the relative saturation of volatile compounds in their vehicle and gave a thermodynamic generalization to the toxicity [8].

Bruce, Kharasch, and Winzler were constructed a group contribution to biological activity values in a series of thyroid hormone analogs. This study may be considered as a first

Free Wilson-type analysis [9]. Zahradnik constructed an equation similar to the Hammett equation, which was used for three decades to describe the reactivity of organic compounds in a quantitative manner [10-12].

The extensive work of Bell and Roblin (1942) established the importance of ionization of a series of sulfanilamides with their antibacterial activities [13].

Taft (1956) proposed an approach for separating polar, steric, and resonance effects of substituents in aliphatic compounds. He first introduced the steric parameter [14].

In 1962 Hansen constructed relationship between the toxicities of substituted benzoic acids and the electronic $\sigma$ constants of their substituents, which is the first real Hammett type relationship [15].

In 1962, Hansch and Mour correlated the biological activities of the plant growth regulators with Hammett constant and hydrophobicity using the octanol/water system [16].

In 1969 Hansch introduced a new hydrophobic scale using the octanol/water system and a whole series of partition coefficients were calculated. The parameter $\pi$ is the relative hydrophobicity of a substituent, which was analogous to the definition of sigma [17].

$$\pi = \log P_X - \log P_H \qquad 4$$

$P_X$, and $P_H$ are the partition coefficients of a derivative and the parent molecule respectively.

The contributions of Hammett and Taft together laid the basis for the development of the QSAR model by Hansch and Fujita (1964), which combined the hydrophobic parameters

with Hammett's electronic constants to yield the linear Hansch equation and its many extended forms [18].

$$\text{Log } 1/C = a\sigma + b\pi + c\,k \qquad 5$$

After this equation over hundreds of equations were constructed and the failure of these linear equations in cases with extended hydrophobicity ranges led to the development of the Hansch parabolic equation (eq.5) [19].

$$\text{Log } 1/C = a \log P - b (\log P)^2 + c\sigma + k \qquad 6$$

The delineation of these models led to explosive development in QSAR analysis and related approaches.

Besides the Hansch approach, other methodologies were also developed structure activity relationship. The Free and Wilson (1964) constructed an additive model, where the activity is a simple sum of contributions from different substituents [20].

$$BA = \Sigma a_i x_i + u \qquad 7$$

BA is the biological activity, u is the average contribution of the parent molecule, and $a_i$ is the contribution of each structural feature; xi denotes the presence Xi = 1 or absence Xi = 0 of a particular structural fragment.

In 1971, Fujita and Ban simplified the Free-Wilson equation estimating the activity for the series of non substituted compound and derived an equation that used the logarithm of activity, which brought the activity parameter in line with other free energy related terms [21].

$$\text{Log } BA = \Sigma G_i X_i + u \qquad 8$$

In equation 8, u is the calculated biological activity value of the unsubstituted Parent compound of a particular series. $G_i$ denotes the biological activity contribution of the substituents, whereas $X_i$ is the presence ($X_i = 1$) or absence ($X_i = 0$) of a particular structural fragment structure.

In 1960s several molecular descriptor were proposed in the development of QSAR modeling, which signaled the beginning studies on molecular descriptors based on graph theory [22-25].

In 1976, Kubinyi examined the transport of drugs via aqueous and lipoidal compartment systems and further refined the parabolic equation of Hansch to develop an advanced bilinear (non-linear) QSAR model [26].

$$\text{Log } 1/C = a \log P - b \log (\beta P+1) + k \qquad 9$$

In the early 1980s, Hans Konemann and Gilman Veith developed multi-class-based, hydrophobic dependent models for industrial organic chemicals, must share credit for the revival of QSAR [27, 28]. Simon developed the quantitative structure property/activity relationships by the minimum topological difference (MTD) method and also Kier and Hall studies on molecular connectivity. Connectivity indices based on hydrogen suppressed molecular structures are rich in information on branching, 3-atom fragments, the degree of substitution, proximity of substituents and length, and hetero atom of substituted rings [29,30]. In 1997, Hansch and Gao developed comparative QSAR (C-QSAR), incorporated in the C-QSAR program. Various attempts have been made to use in QSAR model recognition and successful applications have been reported [31]. The partial least squares (PLS) method offer better opportunity in QSAR model development.

## 1.2: Molecular descriptors in QSAR

To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used. These descriptors are generally understood as being any term, index or parameter conveying structural information. The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment. The descriptors have been used in the formulation of QSAR for the prediction of a chemical properties, biological activity or toxicity. Several descriptors are obtained directly from the chemical structure such as constitutional, geometrical, and topological and electrotopological descriptors, although some of the descriptors represent physicochemical properties such as lipophilicity and hydrophilicity descriptors, electronic descriptors. An additional feature of the descriptors is that the values of some of them depend only on the 2D chemical formula (constitutional, topological and electro topological descriptors), whereas the values of others are influenced additionally by the 3D molecular conformations (geometrical descriptors, electronic descriptors, energies of inter actions). The main grouping of descriptors is given in Table 1

**Table 1: Classification of Descroptors**

| Descriptors | Examples |
|---|---|
| Lipophilicity (hydrophobicity), and Hydrophilicity | aqueous solubility, oil-water partition coefficient (logP), oil-water distribution coefficient (logD) |
| Constitutional | molecular weight, total number of atoms, number of individual types of atoms, number of rings, total number of bonds, number of individual types of bonds |
| Geometrical | van der Waals volume, molecular volume, molecular surface area, solvent accessible molecular surface area, shadow indices |
| Topological and electrotopological-state | Wiener index, Balaban index, Randic indices, Kier and Hall, connectivity indices, Kappa shape indices; E-state atom type indices, E-state extended atom type indices |
| Electronic and quantum mechanical | dipole moments, polarizability, hydrogen bonding parameters, Hammett constant, HOMO and LUMO energies, orbital electron densities, super delocalisabilities |
| Energies of interaction | steric, electrostatic, hydrophobic energies of interaction |

**Descriptors of lipophilicity (hydrophobicity) and hydrophilicity**

Lipophilicity or hydrophobicity is more attracted than the other physicochemical property in QSAR studies, due to its direct relationship to solubility in aqueous phases, to

membrane permeation and to its contribution to ligand binding at the receptor site. Lipophilicity is defined by the partitioning of a compound between an aqueous and an organic phase. Meyer and Overton made their seminal discovery on the correlation between oil/water partition coefficients and the narcotic action of small organic molecules [4, 5]. Ferguson extended this analysis by placing the relationship between depressant action and hydrophobicity in a thermodynamic context; the relative saturation of the depressant in the biophase was a critical determinant of its narcotic potency [8]. Hansch et al. formulated a multi parameter approach that included both electronic and hydrophobic terms, to establish a QSAR for a series of plant growth regulators [18].This study laid the basis for the development of the QSAR model and also firmly established the importance of lipophilicity in biosystems. Hydrophobic interactions are of vital importance in many areas of chemistry such as enzyme-ligand interactions, the assembly of lipids in biomembranes, aggregation of surfactants, coagulation and detergency [33-36]. Molecular recognition depends strongly on hydrophobic interactions between ligands and receptors. The classical model for hydrophobic interactions was defined by Kauzmann to describe the van der Waals attractions between the nonpolar parts of two molecules immersed in water [37]. Hydrophobicities of solutes can be calculated by measuring partition coefficients designated as P.

**Constitutional descriptors**

Constitutional descriptors are the most simple and commonly used descriptors in QSAR analysis and produced from the chemical composition of a compound without any information about its molecular geometry or atom connectivity. This includes molecular weight (MW), number of atoms (nAT), number of Hydrogen atoms (nH), number of Carbon atoms (nC), number of Nitrogen atoms (nN), number of Oxygen atoms (nO), number of halogen atoms (nX), number of rings (nCIC). They encode the size

of molecules (molecular weight, number of atoms, number of rings), and chemical properties (type and number of functional groups).

**Geometrical descriptors**

Geometrical representation involves knowledge of the relative positions of the atoms in 3D space, i.e. the (x; y; z) atomic coordinates of the molecule atoms. These descriptors require access to the 3D coordinates of all the atoms in the given molecule. Geometrical descriptors have more information and discrimination power than topological descriptors for similar molecular structures and molecule conformations. Examples for such descriptors include distances between particular points of the molecular surface (the two Farthest points, the two closest points), and distances between given chemical groups. Several classes of geometrical descriptors descriptor can be distinguished, e.g. quantum-chemical, grid-based, volume and surface descriptors, 3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction) descriptors, etc. The most widely used geometrical descriptors are molecular surface area and molecular volume.

**Topological descriptors**

The topological descriptors treat the structure of the compound as a graph, with atoms as vertices and covalent bonds as edges. Topological descriptors were divided in to two different subsets: topostructural and topochemical. Topostructural descriptors encode information about the adjacency and distances of atoms in moleculer structures irrespective of the chemical nature of the atoms. But topochemical descriptors encode regarding the connectivity as well as specific chemical properties of the atom making up the molecule. The most popular and widely used topostructural indices are Wiener index [38], Randic connectivity index [39], Balaban distance connectivity index [40], Schultz molecular topological index [41], Kier shape descriptors [42], Kier and Hall valence

connectivity index etc [43,44]. Topochemical descriptors are Burden eigen values, BCUT descriptors [45,46], mean information content ($IC_r$) [47], Structural information content ($SIC_r$) [48], Complementary information content ($CIC_r$) [49].

**Electrotopological state indices**

In 1990, Kier and Hall was introduced the electrotopological state (E-state) as a new approach to molecular structure representations [50]. A new method for molecular structure description is presented in which both electronic and topological characteristics are combined. The method makes use of the hydrogen-suppressed graph to represent the structure. The focus of the method is on the individual atoms and hydride groups of the molecular skeleton. The characteristics of the electrotopological state values are indicated by: examples of various types of organic structures, including chain lengthening, branching, hetero atoms, and unsaturation. Electrotopological state indices are E-state atom type indices, E-state extended atom type indices, E-state bond indices etc.

**Electronic and quantum mechanical descriptors**

Electronic descriptors are of significant importance in determining the types of intermolecular forces that lead to drug-receptor interaction. The three major types of parameters that were initially suggested and still hold sway are electronic, hydrophobic, and steric in nature. Electronic descriptors are calculated by the special distribution of the electrons in a molecule. The descriptor characterizes molecular properties such as dipole moment [51], polarizability [52].

The quantum chemical methodology applied to QSAR, by direct derivation of electronic descriptors from the molecular wave functions [53]. The two most accepted quantum chemical methods used for the calculation of quantum chemical descriptors are ab initio

(Hartree-Fock) and semiempirical methods. Like other electronic parameters, QSAR models incorporating quantum chemical descriptors will include information on the nature of the intermolecular interactions involved in the biological response and can easily be derived from the theoretical structure of the molecule. Quantum chemical calculations have no statistical error. The errors are generally made in the assumptions that are established to facilitate calculation [54]. Important quantum chemical descriptors such as net atomic changes, highest occupied molecular orbital/lowest unoccupied molecular orbital (HOMO-LUMO) energies, frontier orbital electron densities and super delocalizabilities have been shown to correlate well with various biological activities [55].

**Energies of interaction**

Major forces encountered in the drug-receptor intermolecular interactions including electrostatic, hydrogen bonding, steric/van der Waals and hydrophobic. The electrostatic and hydrogen bonding interactions are responsible for ligand-receptor specificity. Hydrophobic interactions generally provide the strength for binding. These interactions are used as descriptors in QSAR (3D) analysis [56].

### 1.3: Antimicrobial drugs

Antimicrobial drugs are chemicals used to prevent and treat microbial infections. Antimicrobial drugs can be classified as antibacterial, antifungal, antiviral or antiparasitic depending on the type of microbe the drug targets. Antibiotics fight against bacteria by inhibiting certain vital processes of bacterial cells or metabolism. There are five major modes of antibiotic mechanisms of activity.

(i) **Interference with cell wall synthesis**

β-lactam antibiotics such as penicillins and cephalosporins inhibit the synthesis of the enzymes peptidoglycan layer.Glycopeptides (vancomycin, teicoplanin, oritavancin) target the bacterial cell wall by binding to the D-alanyl-D-alanine termini of the peptidoglycan chain, thereby preventing the cross-linking steps [57, 59].

(ii) **Inhibition of protein synthesis**

Neomycin, gentamicin, tobramycin, kanamycin, and amikacin are common aminoglycoside antibiotics, with antibacterial effect. Their mechanism of action involves inhibition of bacterial protein synthesis via inhibition of the 30S bacterial ribosome. Macrolides interfere with the elongation of nascent polypeptide chains bythe 50S ribosomal subunit. Chloramphenicol binds to the 50S ribosomal subunit blocking peptidyl transferase reaction [57, 58, and 60].

(iii) **Interference with nucleic acid synthesis**

Nalidixic acid and related fluoroquinolones (FQN) are antibacterial effects via action on bacterial DNA synthesis. Quinolone antibiotics such asnorfloxacin, ciprofloxacin, ofloxacin, gatifloxacin interrupt DNA synthesis by interference with type II topoisomerases DNA gyrase and topoisomerase IV during replication and by causing double strand breaks. [57, 58]

(iv) **Inhibition of a metabolic pathway**

Sulfonamides competitively inhibit the first step of bacterial folate synthesis i.e. conversion of para-aminobenzoic acid to dihydrofolic acid. But pyrimethamine and trimethoprim reversibly inhibit the second step via inhibition of the enzyme dihydrofolate reductase i.e. reduces dihydrofolic acid to tetrahydrofolic acid. Antibiotics include

sulfonamides, pyrimethamine, and trimethoprim, and are generally considered as an antibacterial inhibition by a metabolic pathway [57, 58].

**(v) Disorganizing of the cell membrane**

Bacterial and mammalian cell membranes are close to each other. Therefore few antibiotics are able to selectively and effectively target this cellular component without inducing patient toxicity. It is assumed that polymyxins exert their inhibitory effects by increasing bacterial membrane permeability, causing leakage of bacterial content [58, 61, and 62].

The first antimicrobial drug was synthesized by Ehrlich (1910). This was the arsenic containing drug salvarsan. It was not effective against a wide range of bacterial infections but effective against the protozoal disease sleeping sickness (trypanosomiasis) and the spirochaete disease of syphilis. The drug was replaced by penicillin in 1945.

In1928, Alexander Fleming discovered penicillin. Howard Florey and Ernst Chain performed first clinical trials of penicillin in the 1940s.It was not effective against all types of infection and therefore need for new antibacterial agents still remained. Penicillin is an example of a toxic chemical produced by a fungus to kill bacteria.It was realized that fungi might be a source for novel antibiotics. Penicillin is an excellent agent in terms of safety and efficiency. Penicillin saved the lives of many wounded soldiers during World War II [63].

In 1935 with the discovery of prontosil was effective against streptococci infections in vivo. It was finally accepted asprodrugs for a new classes of antibacterial agent the sulfonamides. The discovery of sulfa drugs was the first drugs to be effective against

bacterial infections carried in the bloodstream and they were the only effective drugs until penicillin became available in the early 1940s.

In 1944, the antibiotic streptomycin, an amino glycoside antibiotic, was discovered from a systematic search of soil organisms. It extended the range of chemotherapy to Tubercle bacillus and a variety of Gram-negative bacteria. Several antibiotics were discovered from the soil bacterium Streptomyces griseus such as chloramphenicol, tetracycline, macrolide, and glycopeptide.

After the Second World War, the effort continued to find novel antibiotic structures. This led to the discovery of the peptide antibiotics (e.g. bacitracin (1945)), chloramphenicol (1947), the tetracycline antibiotics (e.g. chlortetracycline (1948)), the macrolide antibiotics (e.g. erythromycin (1952)), the cyclic peptide antibiotics (e.g. cycloserine (1955)), and in 1955 the first example of a second major group of (3-lactam antibiotics, cephalosporin C. The synthesized antimicrobial agent isoniazid, a pyridine hydrazide structure, was found to be effective against human tuberculosis in 1952.

In 1962 nalidixic acid, the first of the quinolone antibacterial agents was discovered. A second generation of this class of drugs was introduced in 1987 with ciprofloxacin [63].

Many antibacterial agents are now available and the vast majority of bacterial diseases have been brought under control (e.g. syphilis, tuberculosis, typhoid, bubonic plague, leprosy, diphtheria, gas gangrene, tetanus, gonorrhoea) [64].

Bacterial resistance to antimicrobial drugs is one of the most serious threats to global public health. Initially treatment with antibiotics is sensitive to microorganism. But prolonged treatment with antibiotics can lead to the development of resistance in a microorganism and later it can adapt gradually and develop resistance to antibiotics.

Previously in some cases, antimicrobial agents are effective but now they are no longer useful. The bacterium S. aureus is the resistant bacterium most common in the clinical setting and rapidly acquired resistance to sulfonamides drugs. Penicillin was initially effective to S. aureus but resistant strain sof S. aureus produce penicillinase increased in the 1950s. In 1960, penicillinase stable methicillin was developed. However, in1961, methicillin resistant S.aureus (MRSA) was isolated in the UK [65]. Nosocomial infection with MRSA became a social problem in 1990s.During this period, the targets of new antimicrobial agents including second and third-generation cephems were widely used. The antimicrobial activity shifted from Gram-positive to Gram-negative bacteria, and agents with wide spectra but weaker activity against Gram-positive bacteria. MRSA obtained resistance to most of the β-lactam antibiotics through its acquisition of the penicillin-binding protein (PBP) 2' gene. PBP2' is an enzyme, which involved in cell wall synthesis that has low binding affinity for β-lactam antibiotics.

S. pneumoniae was originally susceptible to penicillin. But, in the latter half of the 1960s, penicillin-intermediate S. pneumoniae (PISP) strains were found and penicillin-resistant S. pneumoniae (PRSP) strains in the latter half of the 1970s. In Japan, PRSP was isolated in the 1980s and the detection of PRSP strains began to increase around 1990. Plenty use of oral cephem antibiotics seems to be responsible for the increase in PRSP.

Initially antimicrobial activity of Ampicillin was effective for Haemophilusinfluenzae. Some of this species were found to produce β-lactamasein the 1980s and thereby proper resistant to ampicillin. However, in the 1990s, such β-lactamase-producing strains decreased in Japan. Strains that acquired highly resistance to β-lactam through mutations in PBP genes, increased instead. These are known as β-lactamase-negative ampicillin-resistant (BLNAR). It has been considered that use of a large amount of oral cephems

antibiotics is also responsible for antimicrobial resistant, similar to the situation with PRSP.

Many antimicrobial agents produced antimicrobial effects towards the P. aeruginosa. The emergence of P. aeruginosa strains resistant to all of three classes of antimicrobials, i.e., carbapenems, quinolones, and aminoglycosides. This multidrug resistant P. aeruginosa (MDRP) sometimes seems to cause an epidemic in a body. MDRP has complex mechanisms of drug resistance. Antibiotic resistance may be classified in to Genetics of antibiotic resistance and Biochemistry of antibiotic resistance

**Genetics of antibiotic resistance**

Mutations

**Spontaneous mutations**

These mutations occur as errors replication or an incorrect repair of a damaged DNA in actively dividing cells. They are called the growth spontaneous mutations or growth-dependent mutations. These mutations present an important mode of generating antibiotic resistance [66]. Quinolone resistance in Escherichia coli is a result of change in at least seven positions in the gyrA gene or three positions in the parC gene [67, 68], where as only a single point mutation in the rpoB gene is associated with a complete resistance to rifampin [69]. There are several biochemical mechanisms of antibiotic resistance related to Spontaneous Mutations. Different types of genes can be involved in antibiotic resistance. Reduced the affinity of sulfonamides is due to the chromosomal mutation in dihydropteroate synthetase [70].

**Hypermutators**

A small bacterial population during a prolonged nonlethal selection of microorganisms may achieve a transient state when the population mutates at a very high rate [71]. Hypermutators are observed in many bacteria species such as E. coli, S. enterica, Neisseria meningitides (N. meningitides),H. influenzae, S. aureus, Helicobacter pylori (H. pylori), Streptococcus pneumoniae (S. pneumoniae),and P. aeruginosa [72]. Various studies suggested that hypermutations play an important role in acquisition of antibiotic resistance in pathogens [73].

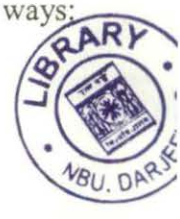
**Adaptive mutagenesis**

The mutation process has classically been studied in actively dividing bacteria. It was assumed that most mutations occur as the consequence of errors during the DNA replication process. But recent experimental data have clearly revealed that mutations occur also in non-dividing or slowly dividing cells. These mutations are called adaptive mutations. A large number of antibiotic resistant mutants may come from adaptive mutation process under bacterial normal conditions. E. coli exposed to antibiotic streptomycin exhibits a hypermutable phenotype. Some antibiotics (quinolones) are able to induce the SOS mutagenic response and increase the rate of emergence of resistance in E. coli [71, 74].

**Horizontal gene transfer**

A principal mechanism by which is a transfer of resistance genes from one bacterium to another is called a horizontal gene transfer. The key mechanisms of resistance gene transfer in a bacterium are plasmid transfer, transfer by viral delivery, and transfer of free DNA. Antibiotic resistance genes may be transferred by three different ways;





transduction (via bacteriophages), conjugation (via plasmids and conjugative transposons) and transformation (via incorporation into the chromosome of chromosomal DNA, plasmids, and other DNAs) [71, 75]. Resistance genes can be further incorporated into the recipient chromosome by recombination or transposition and may have single mutations or several changes in gene sequence. This type of genetic transfer not only happen between closely related bacteria but can also occur between phylogenetically distant bacterial genera, in particular between gram-positive and gram-negative bacteria. The horizontal genes transfer is more effective than chromosomal mutation [75]. Plasmid encoded genes that give antibiotic resistance to main classes of antibiotics such as aminoglycosides, cephalosporins and fluoroquinolones [76]. Horizontal transfer of resistance genes is a process for the propagation of multiple drug resistance because resistance genes can be found in clusters and transferred together to the recipient. This is possible by the existence of specific DNA structures, are called integrons. Integrons are DNA elements with the ability to capture genes that encoding antibiotic resistance by specific recombination mechanism. These elements are placed either on the bacterial chromosome or on broad host range plasmids [77, 78]. Integrons forming clusters of antibiotic resistance genes by the capture one or more gene cassettes within the same attachment site. Gene cassettes are the smallest mobile genetic materials that can carry resistance determinants. These can encode several types of resistance including to b-lactams, chloramphenicol, trimethoprim, aminoglycosides and quinolones [79]. Each of these antibiotic classes several distinct gene cassettes have been reported. Resistance gene cassettes have been found for the most classes of antibiotics. Horizontal transfer of the resistance genes can be done when an integron is incorporated into a broad host range plasmid. A plasmid with a pre-existing resistance gene cassette can get additional resistance gene cassettes from donor plasmids, thus spreading multi resistance [80].

**Biochemistry of antibiotic resistance**

The main types of biochemical mechanisms that bacteria used for resistance are generated by several mechanisms: (i) Antibiotic inactivation (ii) Target modification (iii) Efflux pumps and outer membrane (OM) permeability changes (iv) Target bypass. Each of these four categories also contains an amazing diversity of resistance mechanisms. Sometimes a single bacterial strain may have several types of resistance mechanisms. Each of the four main categories will be discussed below

**Antibiotic inactivation**

The mechanisms of antibiotic inactivation consist of the production of enzymes that degrade or modify the drug itself. Biochemical strategies include hydrolysis, group transfer, and redox mechanisms [71].

**(i)Antibiotic inactivation by hydrolysis**

Large number of antibiotics may contain hydrolytically susceptible chemical bonds such as esters and amides. Several enzymes have evolved to target and cleave these susceptible bonds. As a result, provide a means of destroying antibiotic activity. These enzymes can be emitted by the bacteria, inactivating antibiotics before they reach their target within the bacteria. The β-lactamase enzyme breaks the β-lactam ring that have ester and amide bond of the penicillin and cephalosporin antibiotics. Several Gram-negative and Gram-positive bacteria produce such enzymes and above 200 different β-lactamases have been identified. β-Lactamases are divided into four groups on the basis of functional characteristics, including preferred antibiotic substrate. Clinical isolates normally produce β-lactamases having different functional groups. Extended-spectrum β-lactamases

(ESBLs) mediate resistance take placed to all penicillins, third generation cephalosporins (e.g. ceftazidime, cefotaxime, ceftriaxone) and aztreonam [81-83].

**(ii)Antibiotic inactivation by group transfer**

The most diverse family of resistant enzymes inactivating aminoglycosides, chloramphenicol, streptogramin, macrolides, or rifampicin is called transferases. They can inactivate antibiotics by binding adenylyl, phosphoryl, or acetyl groups to the periphery of the antibiotic molecule. These modifications reduced the affinity of antibiotics to target enzymes. Chemical strategy includes transferof O-acetylation and N-acetylation, O-phosphorylation , O-nucleotidylation, O-ribosylation, O-glycosylation and thiol. Allthese modification strategies require a co-substrate for their activity (ATP, acetyl-CoA, $NAD^+$, UDP-glucose, or glutathione) and therefore these processes are restricted to the cytoplasm [71, 84, and 85].

**(iii)Antibiotic inactivation by redox process**

Oxidation and reduction reactions are used by pathogenic bacteria as a resistance mechanism against antibiotics []. For example, one is the oxidation of tetracycline antibiotics by the TetX enzyme and other is Streptomyces virginiae, producer of the type A streptogramin antibiotic virginiamycin M1, protects itself from its own antibiotic by reducing a critical ketone group to an alcohol at position [71,86].

**Target modification**

The second major resistance mechanism is the modification of antibiotic targets which makes the antibiotic unable to bind the targets properly. It is possible for mutational changes to occur in the target that reduce susceptibility to inhibition even as keeping cellular function [71, 87].

**Peptidoglycan structure alteration**

The peptidoglycan component of the bacterial cell wall gives an excellent selective target for the antibiotics. This is crucial for the growth and survival of most bacteria. Thus, enzymes involved in synthesis and assembly of the peptidoglycan component of the bacterial cell wall, which give excellent targets for selective inhibition. The presence of mutation in penicillin-binding proteins (PBPs) leads to a reduced affinity to β-lactam antibiotics and possibly causing β-lactam resistance in many bacteria strains, such as H. influenzae, N. gonorrhoeae, N. meningitidis, anaerobes, S. dysenteriae [71, 88].

**Protein Synthesis Interference**

A wide range of antibiotics (aminoglycosides, tetracyclines, macrolides, chloramphenicol, fusidic acid, mupirocin, streptogramins, oxazolidinones) can interfere with protein synthesis on different levels of protein metabolism. The resistance to antibiotics that obstruct the protein synthesis or transcription via RNA polymerase is achieved by modification of the specific target [89]. The macrolide, lincosamide and streptogramin Bgroup of antibiotics (MLS antibiotics) block protein synthesis in gram negative bacteria by binding to the 50S ribosomal subunit [90].Then the 50S subunit undergoes a posttranscriptional modification (methylation). RNA methyltransferase involves RNA that is close to the binding place of antibiotics. Mutations in 23S rRNA is same as nonmethylated rRNA, which are associated with MLS resistance. Nonmethylated 23S rRNA and 16S rRNA in Haloarcula marismortui cause resistance to kasugamycin and sparsomycin [91].

**DNA synthesis interference**

The mechanism of resistance is a modification of two enzymes, DNA and topoisomerase IV. Resistance is conferred by mutations in specific regions of the structural genes that sufficiently alter these enzymes preventing the binding of antibiotics. The most common mutations in this region cause resistance through decreased drug affinity for the altered gyrase–DNA complex [92].

**Efflux pumps and outer membrane (OM) permeability**

Efflux pumps are the membrane proteins that export the antibiotics from the cell and maintain their low intracellular concentrations. Reduced outer membrane (OM) permeability results in reduced uptake of antibiotics. The reduced uptake and active efflux induce low level resistance in many clinically important bacteria [71].

**Efflux pumps**

Efflux pumps affect all classes of antibiotics such as macrolides, tetracyclines, and fluoroquinolones. Because these antibiotics inhibit different parts of protein and DNA biosynthesis and therefore must be intracellular to exert their effect. Efflux pumps vary in both their specificity and mechanism [71, 93]. Although some efflux systems are drug-specific, many are multidrug transporters that are capable of expelling a wide spectrum of structurally unrelated drugs. Thus Efflux pumps are contributing significantly to bacterial multidrug resistance (MDR). Efflux transporters can be classified into single or multi component pumps. Single component efflux systems transfer their substrates across the cytoplasmic membrane e.g. tetracycline and macrolide transporters are single component efflux systems. Multi component pumps found in gram-negative bacteria and together with a periplasmic membrane synthesis protein e.g. resistance-nodulation- division

(RND) family members have broader substrates. Antibiotics of all classes are susceptible to the activation of efflux systems except polymyxins. Efflux pumps can be specific to antibiotics and are multidrug transporters that are capable to pump a wide range of distinct antibiotic (e.g. macrolides, tetracyclines, fluoroquinolones) and thus significantly contribute to multidrug resistance (MDR) [71, 94].

**Outer membrane (OM) permeability changes**

Outer membrane is an asymmetric bilayer. Gram-negative bacteria have an outer membrane consisting of an inner layer containing phospholipids and an outer layer containing the lipid a moiety of lipopolysaccharides (LPS). This outer membrane (OM) composition slows down drug penetration and transport across the OM is accomplished by porin proteins that form water-filled channels. Drug molecules can enter the OM employing one of the following methods: (i) by diffusion through porins, (ii) by diffusion through the bilayer, (iii) by self-promoted uptake. The mode of permeability employed by a drug molecule largely depends on its chemical composition. For example, hydrophilic antibiotics either enter the periplasm through porins (e.g. β-lactams) or self-promoted uptake (aminoglycosides). Antibiotics such as β-lactams, chloramphenicol and fluoroquinolones pass the Gram-negative outer membrane via porins [71].

**Target bypass**

This type mechanism of bacterial resistance to antibiotics is specific. Bacteria generate two types of targets: one is sensitive to antibiotics and the alternative one (usually an enzyme) that is resistant to inhibition of antibiotic. Bacteria produce an alternative target enzyme that is resistant to inhibition of antibiotic. The example is in bypassing inhibition of dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS) enzymes in tetrahydrofolate biosynthesis. They are inhibited by trimethoprim and sulfonamides,

respectively. In several trimethoprim and sulfonamide resistant strains, a second enzyme that has low affinity for the inhibitors is produced [71].

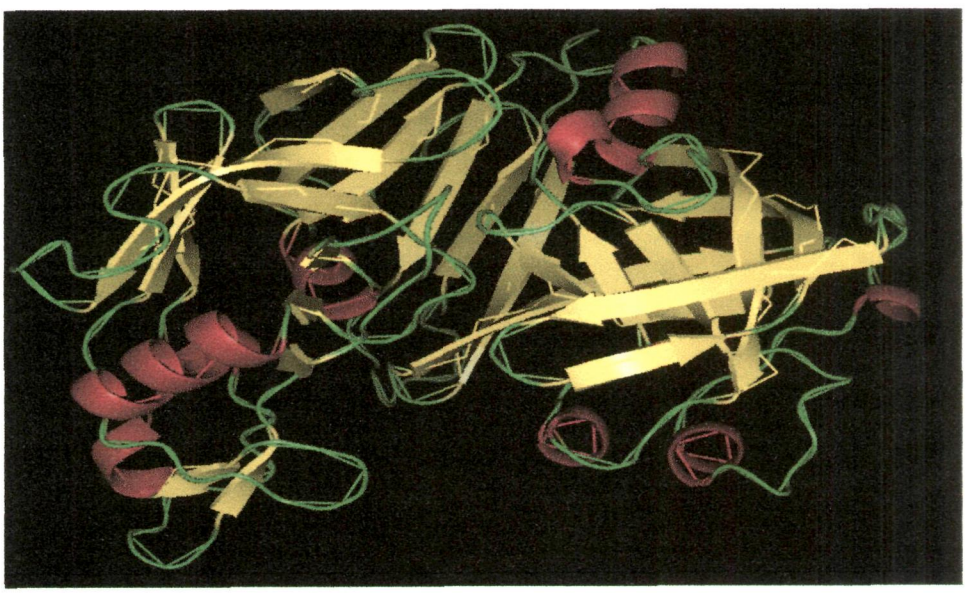
## Receptors

Receptors are generally protein molecule found embedded within the phospholipid bilayer of cell membranes that receives chemical signals from outside the cell. Most receptors are proteins, but also nucleic acids (particularly DNA) are receptors for several drugs. However sometimes in pharmacology, the term is also used for other proteins that are drug targets, such as enzymes, transporters and ion channels [95].

In our work we have dealt with receptor protein molecules, Human renin, Type IIa receptor protein tyrosine phosphatases sigma, Anthrax Lethal toxin, Human bifunctional glutamyl-prolyl-tRNAsynthetase, P38 MAP kinase.

## Human renin

The human renin contains 340 amino acids. Structurally renin consist two homologous lobes with an active site at interface [96, 97]. The catalytic activity of the active site is due to two aspartic acid residues (Asp38, Asp226), one located in each lobe of the renin molecule. The active site of renin can accommodate seven amino acid residues of angiotensinogen and cleaves the Leu10-val11 peptide bond within angiotensinogen to



generate angiotensin I.

**Figure 1: Ribbon diagrams of human renin (helixes are shown in red, sheets are in yellow, and loops in green).**

Renin was one of the first hormones to be discovered in 1898, the first data were presented indirectly suggesting the existence of a renally derived factor that increases blood pressure [98]. This hormone enzyme initiates the enzymatic cascade generating the angiotensin peptides that regulate blood pressure, cell growth, apoptosis and electrolyte balance, to mention only some of the foremost recognized functions. Renin is rate limiting in the production of angiotensin II (Ang II), a hormone that ultimately integrates cardiovascular and renal function in the control of blood pressure as well as salt and volume homeostasis. For instance, renin seems to be of vast importance for maintaining arterial blood pressure. Renin is an aspartyl protease involved in the regulation of blood pressure and electrolyte, homeostasis. As abnormalities in the level of circulating renin contribute to certain forms of hypertension, the enzyme has been a target in the development of antihypertensive drugs.

## Type IIa receptor protein tyrosine phosphatases sigma

Type IIa RPTPs contain variable number of extracellular immunoglobulin (Ig) domains and two to nine fibronectin type III (FNIII) domains and two cytoplasmic phosphatase domains Chicken RPTPσ contains 203 number of amino acid residues. Coles et al. determined crystal structures of the two N-terminal Ig domains (Ig1-2), which formed the minimal stable unit, for examples across family members and species (chicken and human RPTPσ, human RPTPδ and LAR, *Drosophila* DLAR). A V-shape arrangement of Ig1 and Ig2 is stabilized by conserved interactions. Hydrophobic residues (LEU124, LEU129, PRO130, and PHE133 ) of Side chains lie on the Ig1-Ig2 Pro-rich loop and also pack closely with ILE42, VAL44, and ALA 212 in a hydrophobic interdomain region, Two salt bridges ARG91 – GLU205 and GLU126 – ARG215 and a network of hydrogen bonds were formed by the hydroxyl groups of SER50 and TYR216, the side chain amide

of GLN41, the backbone carbonyls of ILE42 and PHE171, the backbone amide of VAL214 and two water molecules, hold Ig1 and Ig2 in a rigid arrangement. The inter domain interactions, including the two water molecules make the Ig1 and Ig2 in a rigid arrangement [99].

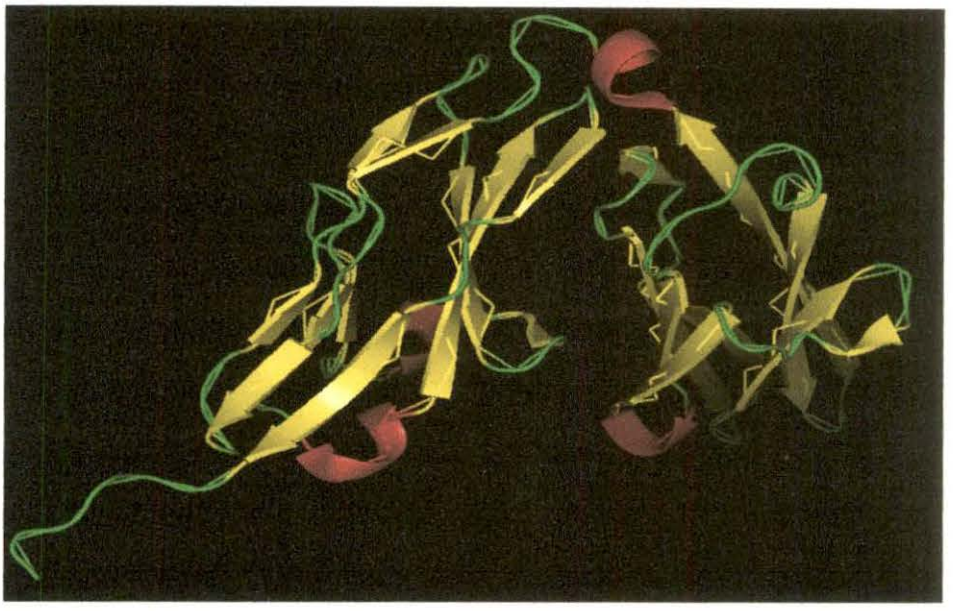


**Figure 2: Ribbon diagrams of Type IIa receptor protein tyrosine phosphatases sigma (helixes are shown in red, sheets are in yellow, and loops in green).**

The complex pattern of neural connectivity established during nervous system development depends on the ability of the axon's motile tip, the growth cone, to receive, transduce and integrate multiple environmental signals. Protein phosphorylation on tyrosine residues plays a key role in these processes. Cellular phosphotyrosine levels are controlled by two major families of enzymes, the protein tyrosine kinases and the protein tyrosine phosphatases (PTPs). Enzymes, tyrosine kinases and protein tyrosine phosphatases (PTPs) controls the extent of phosphorylation of tyrosine residues on cellular proteins. These enzymes are found in both cytoplasmic and transmembrane (receptor-like) forms, and the biochemical interactions between them lead to a diversity of cellular behaviors. Receptor protein tyrosine phosphatases (RPTPs) are a family of cell surface receptors important for growth and development of the nervous system from worms to humans. Twenty one human RPTPs were identified, and highly conserved

orthologues and homologues exist in vertebrates and invertebrates. Vertebrate family members (RPTPσ, LAR and RPTPδ) and invertebrate orthologues (Drosophila DLAR) have particularly well recognized roles in axonal growth cones, regulating neuronal growth and guidance and participating in excitatory synapse formation and maintenance. *RPTPσ* mice show neurological and neuroendocrine defects as well as increased nerve regeneration. Deficiency of RPTPδ mice show impaired learning and memory. RPTPσ and δ double mutant mice have a developmental loss of motor neurons leading to paralysis.

## Anthrax Lethal factor

Anthrax lethal factor have four domains: domain I is the N terminal attached with PA; domain II, III, IV closely associated with each other and also hold the N terminal tail of MAPKK2 [100].

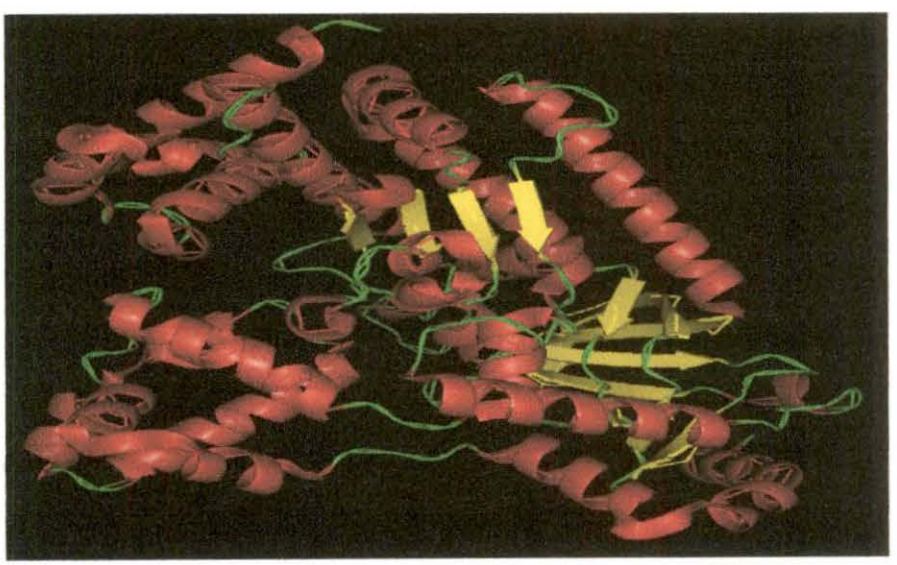


**Figure 3: Ribbon Diagrams of Anthrax Lethal factor (helixes are shown in red, sheets are in yellow, and loops in green).**

Anthrax is a disease most commonly noted among herbivores and people who work with or handle animal products. Anthrax is caused by the spore-forming organism Bacillus

anthracis. Many bacteria that colonize mammalian hosts have evolved mechanisms for introducing bacterial enzymes into the cytosolic compartment of host cells. These enzymes disrupt metabolism in various ways, disabling professional phagocytes or other cells of the host's immune system. Bacillus anthracis secretes three distinct proteins: protective antigen (PA, 83 kDa), oedema factor (EF, 89 kDa) and lethal factor (LF, 83 kDa), none of which is toxic if tested separately [101]. The diseases are marked by the production of protective antigen, the protein that allows the penetration of the cell membrane by either edema factor or lethal factor. Edema factor functions within a human cell as a powerful calmodulin and calcium dependent adenylate cyclase, resulting in edema of the tissues. LF is a zinc- metallopeptidase that can specifically cleave the mitogen-activated protein kinase (MAPK) of the cell. This activity thought to be responsible for the detrimental and life-threatening effects of the enzyme as immune cells, such as macrophages are particularly susceptible.

## Human bifunctional glutamyl-prolyl-tRNAsynthetase

Human EPRS contains three tandem-repeated peptides motifs (EPRS-R123) of 57 amino acids. EPRS-R123 contains 208 amino acids. EPRS-R1 consists of two helices (residues 679-699 and 702-721) arranged in a helix-turn-helix. Two helices (679-699, 702-721) are interacting each other with hydrophobic residues (Val692, Leu695, Ala670, Val705, and Ala708) proximal to the turn involved in helix-helix interactions. The C-terminal loop folds back to interact with the aromatic residues (Tyr682 and Tyr719) in the helices. The hydrophobic interactions between Tyr727 in the loop and Tyr682 and Tyr719 in the helices play an important role for the C-terminal loop formation. Also some hydrophobic residues are involved in helix-helix interactions, but others are exposed on the surface. These observations suggest that they may play a role in protein-protein or protein-nucleic acid interaction [102].

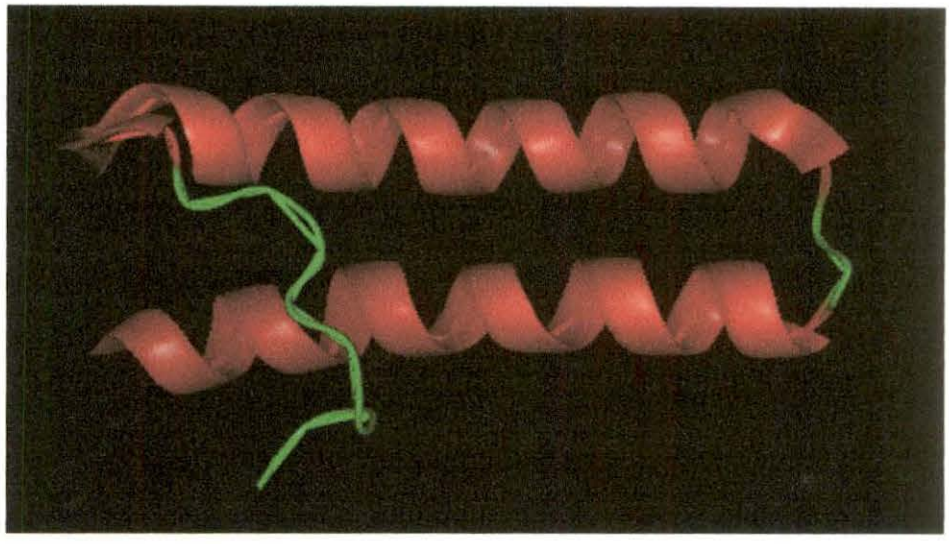


**Figure 4: Ribbon Diagrams of Human bifunctionalglutamyl-prolyl-tRNAsynthetase (helixes are shown in red, and loops are in green).**

Aminoacyl-tRNA synthetases catalyze the attachment of specific amino acid with the correct tRNA for subsequent use in protein biosynthesis. Moreover to this fundamental role in translation, these enzymes are extensively studied as models for the evolution of modular proteins and for insight into the mechanisms of RNA-protein recognition. These enzymes also have additional cellular functions, such **as** synthesis of signaling dinucleotides.

One of the most interesting and least understood properties of aminoacyl-tRNA synthetases, in higher eukaryotes several of these enzymes are exists in a high molecular mass multienzyme complex. As isolated from a variety of sources, this complex has nine individual aminoacyl-tRNA synthetases and one bifunctional polypeptide. These are the enzymes specific for arginine (dimer of 70-kDa polypeptides), aspartic acid (dimer of 56-kDa polypeptides), glutamine (95 kDa), isoleucine (140 kDa), leucine (132 kDA), lysine (dimer of 77-kDa polypeptides), methionine (105 kDA), and the glutamyl/prolyl-tRNA synthetase (160 kDa). Human bifunctional glutamyl-prolyl-tRNA synthetase (EPRS) has

three tandem repeats (EPRS-R1, EPRS-R2, EPRS-R3) of 57 amino acids linking the two catalytic domains. These repeated motifs have been shown to be involved in protein-protein and protein-nucleic acid interactions. The single copy of the homologous motifs has also been found in a number of different aminoacyl-tRNA synthetases [103].

## P38 MAP kinase

The crystal structure of p38 kinase shows a conserved catalytic core having a small N-terminal lobe and a large C-terminal lobe. The complex structure of kinase with ATP disclosed that the ATP nucleotide is bound in the deep cleft at the interface between the two lobes and phosphate groups are found near the opening of the cleft. C-terminal residues interact with the g-phosphate and initiate the kinase reaction [104].

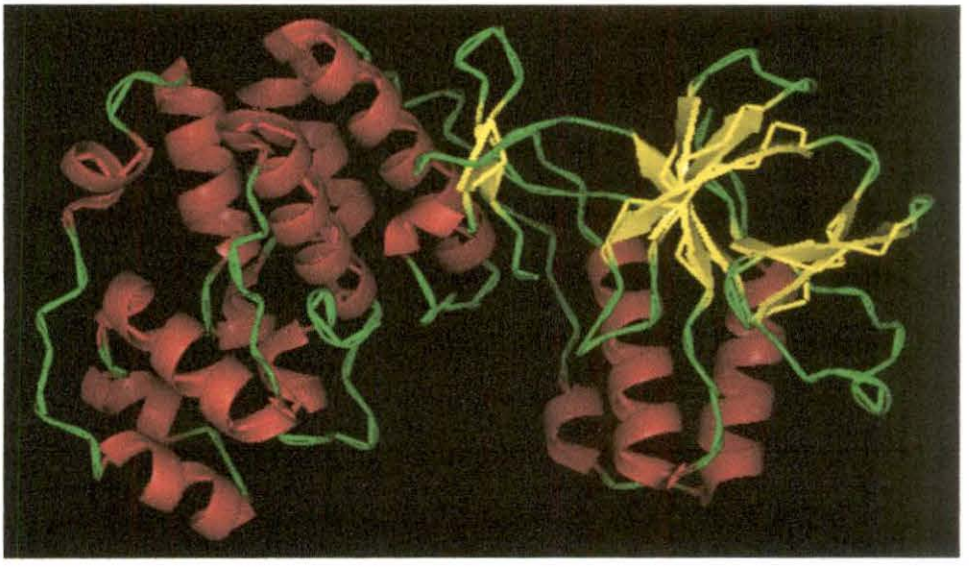


**Figure 5: Ribbon Diagrams of P38 MAP kinase (helixes are shown in red, sheets are in yellow, and loops in green.**

Mitogen activated protein kinases (MAPK) are a class of serine/threonine protein kinases that play a crucial role in signal transduction in many cellular process including growth, differentiation, cell death, and survival. The p38 MAP kinase plays an important role in regulating the production of proinflammatory cytokines, such as tumor necrosis factor-α and interleukin-1β [105]. The proinflammatory cytokines (tumor necrosis factor- α and

interleukin-1β) mediate the inflammatory response associated with the immunological recognition of infectious agents. Elevated levels of proinflammatory cytokines are associated with several diseases of autoimmunity, such as rheumatoid arthritis, osteoarthritis, toxic shock syndrome, diabetes and inflammatory bowel disease. Therefore blocking the p38 MAP kinase may offer an effective therapy for treating many inflammatory diseases. In humans cells six different groups of MAPK that have been identified, such as the extracellular signal- regulated protein kinases (ERK1, ERK2); c-Jun N-terminal kinases (JNK1, JNK2, JNK3) p38s (p38a, p38b, p38g, p38d); ERK5; ERK3s (ERK3, p97 MAPK, MAPK4); ERK7s (ERK7, ERK8). Each group of MAPKs can be simulated by a separate signal transduction pathway in response to different extracellular stimuli.

**1.4: Objectives of the research work**

I. Quantitative structure activity relationship (QSAR) between the structure and physiochemical properties of inhibitors (drug molecules) to their corresponding biological activity. This would help in search of new medicines. With the mathematical and statistical analysis we predict the drug activity and this lead to designing of new potent drug.

II. To improve the biological activities of drug molecules and also predict the biological activities of untested compounds.

III. The increasing resistance of harmful microorganisms to conventional antibiotics has created demand for new antimicrobial agents. Computational approaches such as cheminformatics and bioinformatics are accelerating the process of antimicrobial drug discovery and design of new potent antimicrobial drugs by a rational basis for the selection of chemical structure.

IV. Quantitative structure–activity relationships (QSARs), molecular simulation dynamics and, molecular docking have all benefited from the genomic and proteomic databases and have thus become standard tools in the developed novel products for treating infections.

V. Docking is frequently used to predict the binding orientation of ligand (drug molecule) to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role to predict the ligand conformation and orientation (or posing) within a protein binding site. Generally there are two aims of docking studies: accurate structural modeling and correct prediction of activity. Thus docking studies identified the molecular features that are responsible for specific biological recognition, or the prediction of compound modifications that improve potency of drug molecules.

VI. Molecular dynamics simulation of protein to understand the motional and structural properties. The trajectory files obtained during the whole simulation run was analyzed and was interpreted to understand the structure function relationship in contest of the biological activity of protein.

**1.5: References**

1. Cros AFA. Action de l'alcohol amylique sur l'organisme. Strasbourg, University of Strasbourg, Thesis, 1863.

2. Crum-Brown A, Fraser TR, Trans R. Soc. Edinburgh 1868–1869, 25, 151

3. Richet and Seancs CR (1893) Soc. Biol. Ses. Fil., 9, 775 (1893).

4. Meyer H (1899) Arch. Exp. *Pathol. Pharmakol.*,**42**, 109.

5. E. Overton (1901) Studien Uberdie Narkose, Fischer, Jena, Germany.

6. Hammett LP (1935) Some relations between reaction rates and equilibrium constants. *Chem Rev.*, **17**, 125-36.

7. Hammett LP (1937) The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J Am Chem Soc.*, **59**, 96-103.

8. Ferguson J (1939) The use of Chemical Potentials as Indices of Toxicity. *Proc. R. Soc. Lond. B.*, **127**, 387-404.

9. Bruice TC, Kharasch, N. and Winzler, R. J., Arch. Biochem. Biophys. **62**,305 - 317 (I 956)

10. Zahradnik R, and Chvapil M (1962) Experientia **16**, 511-512

11. Zahradnik, R., Arch. Int. Pharmacodyn. Ther. **135**, 311-329

12. Zahradnik, R., Experientia **18**, 534-536 (1962)

13. Bell PH, Roblin RO (1942) Studies in chemotherapy. vii. a theory of the relation of structure to activity of sulfanilamide type compounds1. *J. Am. Chem. Soc.*, **64**, 2905-2917.

14. Taft RW (1952) Polar and steric substituent constants for aliphatic and o-Benzoate groups from rates of esterification and hydrolysis of esters1. J. *Am. Chem. Soc.*,**74**, 3120-3128.

15. Hansen OR (1962) *Acta Chem. Scand.***16**, 1593- 1600.

16. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature*, **194**, 178-180.

17. Fujita T, Iwasa J, Hansch CJ (1964) *Am. Chem. Soc.*, **86**, 5175

18. Hansch C, Fujita TP (1964) Analysis. A Method for the correlation of biological activity and chemicalstructure. *J. Am. Chem. Soc.*, **86**, 1616-1626.

19. C. Hansch (1969) Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.*, **2**, 232.

20. Free, S.M., Jr.; Wilson, J.W. A Mathematical contribution to structure-activity studies. J. Med. Chem., 1964, 7, 395-399

21. Fujita T, Ban T (1971) Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *J. Med. Chem.*, **14**, 148-152.

22. Gordon M, Scantlebury GR (1964) Non-random polycondensation: Statistical theory of the substitution effect. *Trans Faraday Soc.*,**60**, 604–621.

23. Spialter L (1964) The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP). *J Chem Doc.*, **4**, 261–269.

24. Balaban AT, Harary F (1968) Chemical graphs. V. Enumeration and proposed nomenclature of benzenoid catacondensed polycyclic aromatic hydrocarbons. *Tetrahedron.***24**, 2505–2516.

25. Harary F (1969) Graph theory. Addison-Wesley, Reading MA.

26. Kubinyi H (1976) Quantitative structure-activity relationships. IV. Nonlinear dependence of biological activity on hydrophobic character: a new model. Arzneimittelforschung, **26**, 1991-1997.

27. KonemannH (1981) *Toxicology*, **19**, 209.

28. Veith GD, Call DJ, Brooke LT (1983) *Can. J. Fish Aquat. Sci.*, **40**, 743.

29. Z Simon (1974) *Angew. Chem. Int. Ed. Eng.*, **13**, 719.

30. Hall LH and Kier LB (1977)*J. Pharm. Sci.*, **66**,642

31. Hansch C, Gao H (1997) Comparative QSAR: Radical reactions of benzene derivatives in chemistry and biology. *Chem. Rev.*, **97**, 2995-3060.

32. W.J. DunnIII, S.Wold, U.Edlund, S. Hellberg,and J. Gasteeger, Quant. Struct.-Act. Relat., 1984, 3, 131.

33. Rose GD, Geselowitz AR, et al. (1985) *Science*, **229**, 834.

34. Schneider HJ, Angew. Chem. Int. Ed. Engl., 30, 1417 (1991)

35. Israelachvili JN and Wennerstrom H (1992) J.Phys. Chem., 96, 520.

36. Nusselder JJH and EngbertsLangmuir JBFN (1991)7, 2089.

37. Kauzmann W (1959) *Adv. Protein Chem.*, **14**, 1.

38. Wiener HJ (1947) *Am. Chem. Soc.*, **69**, 17-20.

39. Randic MJ (1975) *Am. Chem. Soc.* ,**97**, 6609-6615.

40. Balaban AT (1982) *Chem. Phys. Lett.*, **89**, 399-404.

41. Schultz HPJ (1989) *Chem. Inf. Comput. Sci.*, **29**, 227-222.

42. Kier LB (1986) Indexes of Molecular Shape From Chemical Graphs. *Acta Pharm. Jugosl.*, **36**, 171-188.

43. Kier LB, Hall LH (1976) Molecular Connectiuity in Chemistry and Drug Research; Academic Press: New York

44. Kier LB, Hall LH (1986) Molecular connectivity in Structure-Actiuity Analysis, Chemometrics Series, Research Studies Press Ltd.: Wiley: New York, Vol. 9.

45. Burden FJ (1989) Chem. Inf. Comput. Sci., 29, 225-227.

46. Pearlman RS, Smith K (1998) *Perspect. Drug Discov. Des.*, 9-11,339-353.

47. Shanon CE (1948) A mathematical theory of Communication. *Bell Syst. Tech.J.*, *27*: 379-423.

48. Basak SC, Roy AB, Ghosh JJ (1979) Study of Structure-Function Relationship of Pharmacological and Toxicological Agents Using Information Theory. In proceeding of the llnd International conference on mathematical modeling, University of Missouri at Rolla. 851-856.

49. Basak SC and Magnuson VR (1983) Molecular Topology and Narcosis. A Quantitative Structure-Activity Relationship (QSAR) Study of Alcohols Usins Complementary Information Content, Arzneimittle-Froschung, *Drug Research.* 501-503.

50. Kier LB, Hall LH (1990) An electrotopological-state index for atoms in molecules. *Pharm Res*,.7,801–807

51. Bottcher CJF, van Belle OC, Bordewijk P, RipA (1973) Theory of Electric Polarization. *Elsevier, Amsterdam (The Netherlands).***1**, 378.

52. Chenzhong C and Zhiliang L (1998) Molecular Polarizability. 1. Relationship to Water Solubility of Alkanes and Alcohols. *J. Chem. Inf Comput. Sci.*, **38**, 1-7.

53. Karelson M, Lobanov VS, and Katritzky AR (1996) *Chem. Rev.*, **96**, 1027.

54. P. S. Magee in ACS Symposium Series 37, American Chemical Society, Washington, DC, 1980.

55. Gupta SP (1991) *Chem. Rev.*, **91**, 1109.

56. D. Boyd in A. L. Parrill and M. Rami-Reddy, Eds., Rational Drug Design, ACS Symposium Series 719, American Chemical Society, Washington, DC, 1999, p. 346.

57. Strohl WR (1997) Biotechnology of Antibiotics, Marcel Dekker Inc., New York, USA.

58. Fred C. Tenover. Mechanisms of Antimicrobial Resistance in Bacteria. The American Journal of Medicine 2006, 119(6A), S3-S10.

59. Benton B, Breukink E, et al. (2007) Telavancin inhibits peptidoglycan biosynthesis through preferential targeting of transglycosylation: Evidence for a multivalent interaction between telavancin and lipid II, *Int. J. Antimicrob. Agents (Suppl.)*, **29**, 51–52.

60. Leach KL,Swaney SM, et al. (2007) The site of action of oxazolidinone antibiotics in living bacteria and in human mitochondria, *Mol. Cell*, **26** 393–402.

61. Tenover FC (2006) Mechanisms of antimicrobial resistance inbacteria, *Am. J. Med. (Suppl.)*, 119, 3–10.

62. Straus SK, Hancock REW (2006) Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: Comparison with cationic antimicrobial peptides and lipopeptides, *Biochim. Biophys. Acta*, **1758**, 1215–1223.

63. SAGA T, YAMAGUCHI T (2009) History of antimicrobial Agents and Resistant Bacteria, *JMAJ.*, 52(2).

64. Powers JH (2004) Antimicrobial drug development—the past, the present, and the future. *Clin Microbiol Infect.*,**10** (Suppl 4),23–31.

65. Jevons MP. (1961) "Celbenin"-resistant Staphylococci. *Br Med J.*, **1**, 124–125.

66. . R. Kraovec, I. Jerman, Bacterial multicellularity as a possible source of antibiotic resistance, Med. Hypotheses, 60 (2003) 484–488.

67. Hooper DC (1999) Mechanisms of fluoroquinolone resistance. *Drug Resist Updat*; **2**(1):38-55.

68. Nakamura S, Nakamura M, Kojima T, Yoshida, (1989) gyr A and gy rB mutations in quinolone-resistant strains of Escherichia coli, *Antimicrob. Agents Chemother.*, **33**, 254–255.

69. Rice LB, Sahm D, Binomo RA. Mechanisms of resistance to antibacterial agents. In: Murray PR, Baron EJ, Jorgensen JH, Phaller MA, Yolken RH, editors. Manual of clinical microbiology. Washington: ASM Press; **2003**. 1074-101.

70. Attacking the enemy: antimicrobial agents and chemotherapy. In: Mims C, Dockrell HM, Goering RV, Roitt I, Wakelin D, Zuckerman M, editors. Medical microbiology. Elsevier Mosby; 2004. p. 473-507.

71. Džidic S, Šuškovic J, Kos B (2008) Antibiotic resistance mechanisms in bacteria: biochemical and genetic aspects. *Food Technol Biotechnol.*, **46**, 11-21.

72. Denamur E, Bonacorsi S, et al. (2002) High frequency of mutator strains among human uropathogenic Escherichia coli isolates, *J. Bacteriol.*,**184**, 605–609.

73. Oliver A, Canton R, Campo P, et al. (2000) High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, **288** (5469), 1251-1254.

74. Macia MD, Blanquer D, et al (2005) Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections. *Antimicrob. Agents Chemother.*, **49**, 3382–3386.

75. Alekshun MN, Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell.*, **128**, 1037-50.

76. Nordmann P, Poirel L (2002) Emerging carbapenemases in Gram-negative aerobes. *Clin-MicrobiolInfect.*, **8** (6):321-331.

77. Rowe-Magnus DA, Mazel D (1999) Resistance gene capture,*Curr. Opin. Microbiol.***2**, 483–488.

78. Rowe-Magnus DA, Mazel D (1999) Resistance gene capture. *Curr. Opin. Microbiol.* **2**, 483–488.

79. Boucher Y, Labbate M, Koenig JE, Stokes HW (2007) Integrons: Mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15**, 301–309.

80. Spratt BG, Bowler LD, Zhang QY, Zhou J, Smith JM (1992)Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species, *J. Mol. Evol.* **34**, 115–125.

81. Bush K, Jacoby GA, Medeiros AA (1995) A functional classification scheme for b-lactamases and its correlation with molecular structure, *Antimicrob. Agents Chemother.***39**, 1211–1233.

82. Kotra LP, Mobashery S (1999) Mechanistic and clinical aspects of b-lactam antibiotics and b-lactamases, Arch. Immunol. Ther.Exp. (Warsaw), 47, 211–216.

83. Shah A, Hasan FS, Ahmed A (2004) Hameed, Extended spectrum b-lactamases (ESBLs): Characterization, epidemiology and detection, *Crit. Rev. Microbiol.* **30**, 25–32.

84. Mayer KH, Opal SM, Medeiros AA (1995) Mechanisms of antibiotic resistance. In: Mandell GL, Bennett JE, Dolin R, editors. Basic principles in the diagnosis and management of infectious diseases. Churchill Livingstone: An Imprint of Elsevier, 212-25.

85. Vetting MW, Magnet S, et al. (2004) A bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones, *Chem. Biol.*,**11**, 565–573.

86. Hawkey PM (1998) The origins and molecular basis of antibiotic resistance. *BMJ.*, **317**, 657-60.

87. Spratt BG (1994) Resistance to antibiotics mediated by target alterations, Science, 264, 388–393.

88. Dowson CG, Coffey TJ, Spratt BG (1994) Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to b-lactam antibiotics, *Trends Microbiol.*,**2**, 361–366.

89. Happi CT, Gbotosho GO, et al. (2005) Polymorphisms in Plasmodium falciparum dhfr and dhps genes and age related in vivo sulfaxine pyrimethamine resistance in malaria-infected patients from Nigeria. *Acta Tropica*, **95**, 183–193.

90. Weisblum B (1995) Erythromycin resistance by ribosome modification. *Antimicrob. Agents Chemother.*, 39, 577–585.

91. Wang G, Taylor DE (1998) Site-specific mutations in the 23SrRNA gene of Helicobacter pylori confer two types of resistance to macrolide–lincosamide–streptogramin B antibiotics, Antimicrob. *Agents Chemother.*, **42**, 1952–1958.

92. Willmott CJ, Maxwell A (1993) A single point mutation in theDNA gyrase A protein greatly reduces binding of fluoroquinolones to the gyrase–DNA complex. Antimicrob. *Agents Chemother.*, **37**, 126–127.

93. Webber MA, Piddoc kLJ (2003) Th e importance of efflux pumpsin bacterial antibiotic resistance, *J. Antimicrob. Chemother.*,**51**, 9–11.

94. Lambert PA (2002) Mechanisms of antibiotic resistance in Pseudomonas aeruginosa. *J R Soc Med.*, **95**, Suppl 41:22-6.

95. Stephenson RP (1956) A modification of receptor theory. Brit. J.Pharmacol., 11, 379.

96. Rahuel J, Rasetti V, Maibaum J, Rüeger H, Göschke R, et al. (2000) MG. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chemistry & biology*, 7: 493-504.

97. Gradman AH, Kad R (2008) Renin Inhibition in Hypertension. *J. Am. Coll. Cardiol.*, **51**: 519-528.

98. Laragh JH (1960) The role of aldosterone in man: Evidence for regulation of electrolyte balance and arterial pressure by renaladrenal system which may be involved in malignant hypertension. *J.A.M.A.*, 174: 293-295.

99. Coles CH, Shen Y, et al., (2011) Proteoglycan-Specific Molecular Switch for RPTPσ Clustering and Neuronal Extension. *Science*. 332: 484-488

100. Shoop WL, Xiong Y, et al. (2005) Anthrax lethal factor inhibition. *PANAS*, 102: 7958-7963.

101. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC (1997) Crystal structure of Anthrax toxin protective antigen. *Nature*.**385**: 833-838.

102. Rho SB, Lee JS, et al. (1998)A multifunctional repeated motif is present in human bifunctionaltRNAsynthetase.*J.Biol.Chem.*,**273:** 11267-11273.

103. R Fett, R Knippers (1991)The Primary Structure of Human Glutarninyl-tRNASynthetase.*J.Biol. Chem.*, **266**, 1448-1455.

104. Pargellis C, Gilmore L, Graham AG, Grob PM, Hickey ER (2002)Inhibition of P38 Map-Kinase by utilizing a Novel Allosteric Binding State. *J Regan Nat. Struct. Biol.*, **9**: 268-272.

105. Davis RJ (1998) Signal transduction by the c-Jun N-terminal kinase (JNK): from inflammation to development. *Curr. Opin. Cell Biol.***10**: 205–219.

# Chapter II

# Review of Literature

## 2.1: QSAR and its development

In 1868 Crum-Brown and Fraser formulated relationship between the physiological activity of alkaloids with its chemical structure [1]. In the same year Richardson also expressed the idea that the chemical structure of a substance as a function of solubility. In 1893, Richet suggested that their toxicities of organic compounds were inversely related to their water solubilities [3]. Overton and Meyer correlated partition coefficients of a group of organic compounds with their anesthetic potencies and concluded that narcotic activity is dependent on the lipophilicity (olive oil/water partition coefficients) of the molecules [4-5]. Louis Hammett correlated electronic properties of organic acids and bases with their dissociation constants and reactivity. He postulated electronic sigma-rho constants and established the linear free-energy relationship (LFER) principle [6, 7]. In 1939 Ferguson correlated depressant action with the relative saturation of volatile compounds in their vehicle and gave a thermodynamic generalization to the toxicity [8].

Bruce, Kharasch, and Winzler were constructed a group contribution to biological activity values in a series of thyroid hormone analogs. This study may be considered as a first Free Wilson-type analysis [9]. Zahradnik constructed an equation similar to the Hammett equation, which was used for three decades to describe the reactivity of organic compounds in a quantitative manner [10-12].

The extensive work of Bell and Roblin (1942) established the importance of ionization of a series of sulfanilamides with their antibacterial activities [13].

The theoretical QSAR approaches developed at the end of 1940s. In this approach biological activities and physico-chemical properties correlated to theoretical numerical indexes derived from the molecular structure. The first topological index based on the graph theory was developed by Wiener. The Wiener index [14] and the Platt number [15]

used as a topological descriptor in 1947 to constructed a QSAR model. The elementary work of Taft in 1950s was developed a relationships between physico-chemical properties and solute solvent interaction energies, based on steric, polar, and resonance parameters for substituent groups in congeneric compounds [16].

In 1962 Hansen constructed relationship between the toxicities of substituted benzoic acids and the electronic σ constants of their substituents, which is the first real Hammett type relationship [17].

The extensive work of Hansch the QSAR approach began to assume its modern look in 1960s. The definition of Hansch model led to the explosion of QSAR analysis and is known as Hansch analysis. Prof. Corwin Hansch proposed the use of linear multiple regressions in order to predict the biological response of compounds yet to be synthesized [18-21]. In this approach, each chemical structure is represented by several parameters which explain hydrophobicity, steric and electronic properties and is generally formalized into the following equation.

$$\log\frac{1}{C} = a(\text{hydrophobic parameter}) + b(\text{electronic parameter}) + c(\text{steric parameter}) + \text{constant} \quad (1)$$

Where C represents the molar concentration producing the biological effect; a, b and c are the regression coefficients. Hansch et al analyzed the anti adrenergic activities of meta and para substituted N, N-dimethyl-α-bromo-phenethylamines [22]. Hansch proposed equation 2.

$$\log\frac{1}{C} = 1.22\pi + 1.59\sigma + 7.82 \quad (2)$$

$$n = 22;\ r = 0.918;\ s = 0.238$$

Where $\pi$ is the hydrophobic substituent bonding constant [23], $\sigma$ represents the Hammett's substituent constant for electronic effects; n is the number of compounds; r is the correlation coefficient, and s is the standard error of the estimate. Hansch's methodology has been basically applied to examine different biological data such as affinity data, inhibition constant, pharmacokinetic parameters and applied in the frame work of diverse therapeutics areas such as antibacterial, anticancer and antimalarial drugs. At the same time, besides the Hansch approach, Free and Wilson developed a methodology of additive substituent contributions to biological activities. The Free-Wilson approach (1964) is a structure activity-based methodology because it incorporates the contributions made by various structural fragments to the overall biological activity [24]. It is represented by following equation.

$$BA = \Sigma a_i x_i + u \qquad (3)$$

BA is the biological activity, u is the average contribution of the parent molecule, and $a_i$ is the contribution of each structural feature; xi denotes the presence Xi = 1 or absence Xi = 0 of a particular structural fragment. The inhibitory activity of a series of heterocyclic compounds against K. pneumonia has been studied by Free-Wilson type analysis [25]. Other applications of the Free-Wilson approach also applied on the antimycobacterial activity of 4-alkyl- thiobenzanilides, the antibacterial activity of fluoronapthyridines, and the benzodiazepine receptor-binding ability of some non-benzodiapzepine compounds such as 3-X-imidazo- [1, 2-b] pyridazines, 2-phenylimidazo [1, 2-a] pyridines, 2-(alkoxycarbony) imidazo [2, 1 p] benzothiazoles and 2-arylquinolones [26,27].

In 1971, Fujita and Ban simplified the Free-Wilson equation estimating the activity for the series of non substituted compound and derived an equation that used the logarithm of

activity, which brought the activity parameter in line with other free energy related terms [28].

$$\text{Log BA} = \Sigma G_i X_i + u \qquad (4)$$

In equation 8, u is the calculated biological activity value of the unsubstituted Parent compound of a particular series. $G_i$ denotes the biological activity contribution of the substituents, whereas $X_i$ is the presence ($X_i$ = 1) or absence ($X_i$ = 0) of a particular structural fragment structure. This method has the advantage of being independent of the possible problems associated with the calculation of molecular descriptors.

In 1970s, the used of quantum-chemical descriptors in QSAR modeling [29], although they actually were conceived several years before to encode information about important properties of molecules in the framework of quantum-chemistry. The valuable work of Balaban [30], Randic [31], Kier and Hall [32] led to developed of the QSAR approaches based on topological indexes. Since the mid-1980s, leading to the development of the 3D-QSAR based on the geometrical aspects of molecular structures. Geometrical descriptors were obtained from the 3D spatial coordinates of a molecule, such as shadow indexes [33], charged partial surface area descriptors [34], WHIM descriptors [35], gravitational indexes [36], EVA descriptors [37], 3D-MoRSE descriptors [38], and GETAWAY descriptors [39].

In 1976, Kubinyi examined the transport of drugs via aqueous and lipoidal compartment systems and further refined the parabolic equation of Hansch to develop an advanced bilinear (non-linear) QSAR model [40].

$$\text{Log } 1/C = a \log P - b \log (\beta P+1) + k \qquad (5)$$

Connectivity indices based on hydrogen suppressed molecular structures are rich in information on branching, 3-atom fragments, the degree of substitution, proximity of substituents and length, and hetero atom of substituted rings [41-42].

The Variations on this activity-based approach was extended by Klopman et al. and Enslein et al. Topological methods were used to construct the relation between molecular structure and physical/biological activity, [43-44].

In the early 1980s, Hans Konemann and Gilman Veith was developed multi-class-based, hydrophobic dependent models for industrial organic chemicals that must share credit for the revival of QSAR [45-46].

In 1980, Dietrich et al. was constructed a quantitative structure-activity relationship for the inhibition of purified E. coli dihydrofolate reductase by 23 5-(substituted benzyl) -2,4-diaminopyrimidines and Comparison of the QSAR for E. coli enzyme inhibition with that previously obtained for bovine enzyme offers the first general explanation for greater selectivity of the important antibacterial agent trimethoprim [47].

A quantitative structure-activity relationship was formulated by Hansch for the binding of a set of substituted benzene sulfonamides to human carbonic anhydrase. Qualitative aspects of the QSAR are correlated with a color stereo molecular graphics model of the enzyme-inhibitor complex which was constructed from the X-ray crystallographic coordinates of the enzyme [48].

An analysis of the inhibition constants of pyrazoles, phenylacetamides, formylbenzylamines, and acetamides acting on liver alcohol dehydrogenase (ADH) yields quantitative structure-activity relationships (QSAR) having a linear dependency on octanol-water partition coefficients (log P). Authors are suggesting that complete

desolation of the substituents on binding to the enzyme. Study of a molecular graphics model of ADH constructed from the X-ray crystallographic coordinates shows that the substituents are surrounded in a long hydrophobic channel which is so narrow that water of solvation must be removed from them in the binding process [49].

Morgenstern et al. Study the hydrolysis of a set of 28 X-phenyl Hippocrates and derived a quantitative structure-activity relationship. Using the x-ray crystallographic coordinates for chymotrypsin and computer graphics, a model was constructed which is used to understand the quantitative structure activity relationship. They found that when polar substituents have the option of binding to hydrophobic space or remaining in the aqueous phase [50].

In 1979, Cramer and Milne was proposed a new approach to describe molecular properties by aligning molecule in space and by mapping their molecular field to a 3D grid [51]. Vectors were extracted from these fields by principal component analysis and correlated with the biological activities. Later this approach was further developed as the DYLOMMS (dynamic lattice-oriented molecular modeling system) method (Kubinyi et al.) [52]. Svante Wold (1986) was proposed the use of partial least squares (PLS) analysis to correlate the field values with the biological activities. This method is more suitable than principal component analysis [52-55].

Cramer in 1988, proposed a new approach in the field of QSAR of three-dimensional molecular parameters and the method was called comparative molecular field analysis (CoMFA), which was later developed as 3D-QSAR [56]. The method is still under active development and has found many successful applications. The conformers, stereo isomers or enantiomers of chemical compounds in 3D-QSAR models allowed the comparison molecular structures thereby setting up a representative structural group known as the

pharmacophore [57,58]. Other 3D-QSAR approaches have been developed, such as Comparative Molecular Similarity Indices Analysis (CoMSIA) [58, 59] or Self Organizing Molecular Field Analysis (SomFA) [60], some of which incorporate comparisons of different sets of molecular descriptors.

Thomas et. al. describe a three-dimensional molecular modeling program using comparative molecular field analysis to derive quantitative structure-activity relationship fitting pharmacological potencies and binding affinities of cannabinoids. The analysis has established to precisely fit the pharmacological activity of cannabinoid analogs, with cross-validated $r^2$ values of greater than 0.3 and final analysis $r^2$ values of greater than 0.88. Additionally, this study has further characterized the steric and electrostatic properties that account for the variations in their potency. This method can be utilized for designing cannabinoid agonists and it is capable of predicting the activity of unknown and thereby to facilitate rational drug design [61].

Waller et al. constructed a statistically significant QSAR model using lipophilic and dipole moment characteristics of the molecules as physical descriptor variables in the regression equation. Comparative molecular field analysis (CoMFA) was employed as a three-dimensional QSAR technique to explore changes in the steric and electrostatic fields of the molecules that can account for differences in biological activity values. A highly predictive model was achieved. These modeling techniques represent the evolutionary process by which structure-activity methods were employed for the development of more potent inhibitors of astrocytic chloride transport [62].

Hansch et al. developed comparative QSAR (C-QSAR), incorporated in the C-QSAR program (1997) [63].

Hologram QSAR (HQSAR) was developed by Heritage and Lowis in 1997. In this model the structures are converted into all possible fragments, which are assigned specific integers, and then hashed into a fingerprint to form the molecular hologram. The bin occupancies of these holograms are used as the QSAR descriptors, encoding the chemical and topological information of molecules [64-65]

Cho et al. developed Inverse QSAR, which try to find values for the molecular descriptors that possess a desired activity/property value. In other words, it consists of finding the optimum sets of descriptor values best matching a target activity and then generating a focused library of candidate structures from the solution set of descriptor values [66].

In 1999, Labute developed Binary QSAR to handle binary activity measurements from high-throughput screening, and molecular descriptor vectors as input. A probability distribution for actives and inactives is then determined based on Bayes' Theorem [67].

A three-dimensional quantitative structure-activity relationships (3D QSAR) method, Comparative Molecular Field Analysis (CoMFA), was applied to a set of 75 dipyridodiazepinone (nevirapine) derivatives active against wild-type (WT) and mutant-type (Y181C) HIV-1 reverse transcriptase. All dipyridodiazepinone derivatives, divided into a training set of 53 derivatives and a test set of 22 derivatives, were constructed. CoMFA models give the satisfactory predictive ability regarding WT and Y181C inhibitions, with r2 cv = 0.624 and 0.726, respectively. The results obtained provide information for a better understanding of the inhibitor-receptor interactions of dipyridodiazepinone analogs [68].

Zefirov et al. constructed a QSAR model using solvation index and obtained a very good one-parameter regression. They are generated the structures of the whole possible set of

small sulfides (C2-C6) and the statistics were recognized by real prediction using an external test set of sulfides. The variants of extended prediction with extrapolated data and QSAR using an expanded training set were also performed, and all these data also revealed the preference of the solvation index [69].

Golbraikhet. al. used two-dimensional (2D) molecular descriptors and k nearest neighbors (kNN) QSAR method for the analysis of several datasets. No correlation between the values of q2 for the training set and predictive ability for the test set was found for any of the datasets. They assumed that the high value of LOO q2 appears to be the necessary but not the sufficient condition for the model to have a high predictive power. They also argued that this is the general property of QSAR models developed using LOO cross-validation and formulate a set of criteria for evaluation of predictive ability of QSAR models [70].

A novel and effective method for drug design and screening was developed by Liu et al. They used to develop quantitation and classification models which can be used as a potential screening mechanism for a novel series of 5-diarylimidazoles inhibitors of COX-2 and calculated constitutional, topological, geometrical, electrostatic, and quantum-chemical features. Quantitative modelling results in a nonlinear, seven-descriptor model based on SVMs with root mean-square errors of 0.107 and 0.136 for training and prediction sets, respectively. The accuracy for training and test sets is 91.2% and 88.2%, respectively [71].

Basak et al. Used different classes of graph theoretic indices, e.g., topostructural indices, topochemical indices, geometrical (3D) indices and, quantum chemical descriptors, for the development of predictive models for vapor based on a structurally diverse set of 469 chemicals. Initially, a set of 379 molecular descriptors was calculated. Comparatively,

three linear regression methodologies were used to develop hierarchical QSAR (HiQSAR) models, namely ridge regression (RR), principal components regression (PCR), and partial least squares (PLS) regression. The results indicate that, in general, RR outperforms PCR and PLS, and that the easily calculated topological descriptors are sufficient for the prediction of vapor pressure based on this large, diverse set of chemicals [72].

A linear quantitative structure-activity relationship was developed for a series of para-substituted aromatic sulfonamides by using topological index methodologies [Melagraki et al.]. They were calculated large series of topological indices and the stepwise regression method was employed to derive the most significant model. A brilliant result was achieved using multi-parametric regressions and the work is quite useful for modeling carbonic anhydrase inhibition [73].

Kuz'minet. al. was constructed a QASR using the 50% cytotoxic concentration (CC50) in HeLa cells, the 50% inhibitory concentration (IC50) against human rhinovirus 2 (HRV-2), and the selectivity index (SI = CC50/IC50) of [(biphenyloxy) propyl] isoxazole derivatives. Statistic characteristics for partial least-squares models are quite satisfactory (R2 = 0.838 - 0.918; Q2 = 0.695 - 0.87) for prediction of CC50, IC50, and SI values. Models are permitting the virtual screening and molecular design of new compounds with strong anti-HRV-2 activity [74].

A quantitative structure activity relationship (QSAR) analysis has been performed on a data set of 42 aryl alkenyl amides/imines as bacterial efflux pump inhibitor using several types of descriptors including topological, spatial, thermodynamic, information content and E-state indices. Statistically significant model ($r^2$=0.87) was obtained with the descriptors like radius of gyration and heat of formation, A log P atom types and solvent

accessible charged surface area playing an important role in determining the activity of the compounds against bacterial efflux pump. The model was also tested successfully for external validation criteria. The model able to predict the activity of new compounds but also explained the important regions in the molecules in a quantitative manner [75].

A method in its embryonic stage of development uses both graph bond distances and Euclidean distances between atoms to calculate E-state values for each atom in a molecule that is sensitive to conformational structure. Recently, these electrotopological indices that encode significant structured information on the topological state of atoms and fragments as well as their valence electron content have been applied to biological and toxicity data [76].

## 2.2: Future Prospects

Quantitative structure activity relationship (QSAR) one of the most popular theoretical method for modeling the manner in which chemical structure of a compound correlate with the biological activity and physical, chemical and technological properties. Using this approach we can now estimate biological activities and physical or chemical properties of series of newly designed compounds before making the final conclusion on whether or not to synthesize. Such predictions are based on several structural features, or molecular descriptors of compounds. The variations in biological activity with molecular structures are essential concepts for development of QSAR models and also the importance of proper building and validation of models. QSAR is a promising method that facilitates the drug discovery Process. QSAR can also be used in industries apart from drug design such as food industry, [30–agro chemicals, fine chemical industry, and material design. The selection of descriptors in QSAR studies is one of the major challenges. Also biggest challenges in QSAR studies using three-dimensional descriptors are to generate the bioactive conformation of the molecules. Ligand-based methods, including traditional quantitative structure

activity relationships (QSARs) and modern 3D-QSAR techniques are based entirely on experimental structure activity relationships for receptor ligands.

Because docking used in 3D-QSAR gives a static snapshot of the dynamic ligand binding process and several studies have suggested using molecular dynamics simulations. Molecular dynamics simulations of substrate molecules the motional properties and also understand the structural variations of such multibody complexes under physiological conditions. Therefore MD simulations were carried out with explicit solvent of the binding poses of the identified biological hits. They are helped in evaluating the stability of the binding site interactions of protein and ligand (drug molecule) and also identify perturbations in the interaction profiles that would not be possible through docking studies.

The increasing resistance of harmful microorganisms to conventional antibiotics has created demand for new antimicrobial agents. Computational approaches such as chemoinformatics and bioinformatics are accelerating the process of antimicrobial drug discovery and design of new potent antimicrobial drugs by a rational basis for the selection of chemical structure.

Now Genomics, molecular simulation and dynamics, molecular docking, structural/functional class prediction, and quantitative structure–activity relationships (QSARs) have all benefited from the genomic and proteomic databases and have thus become standard tools in the develop novel products for treating infections. The major parts of the research has been focused on the design of new antimicrobial drugs, motional properties of protein, binding cavity of drug and interactions between drug and its substrate (protein). Therefore these lead to the development of new potent drugs.

## 2.3: References


1. Cros AFA Action de l'alcohol amyliquesur l'organisme. Strasbourg, University of Strasbourg, Thesis, 1863.
2. A. Crum-Brown, T.R. Fraser, Trans. R. Soc. Edinburgh, 1868–1869, 25, 151
3. Richet and C. R. Seancs, Soc. Biol. Ses. Fil., 9, 775 (1893).
4. H. Meyer, Arch. Exp. Pathol. Pharmakol., 42, 109 (1899).
5. E. Overton, Studien Uberdie Narkose, Fischer, Jena, Germany, 1901.
6. Hammett LP. Some relations between reaction rates and equilibrium constants. Chem Rev 1935; 17:125-36.
7. Hammett LP (1937) The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J Am Chem.*, Soc., 59, 96-103.
8. Ferguson J (1939) The Use of Chemical Potentials as Indices of Toxicity. *Proc. R. Soc. Lond. B*, **127**, 387-404.
9. Bruice TC, Kharasch N and Winzler RJ (1956) *Arch. Biochem. Biophys.* **62**,305 - 317.
10. Zahradnik R, Chvapil M (1960) *Experientia*, **16**, 511-512.
11. Zahradnik R (1962), *Arch. Int. Pharmacodyn. Ther.* **135**, 311-329.
12. Zahradnik R (1962), Experientia, 18, 534-536.
13. Bell PH, Roblin RO (1942) Studies in chemotherapy. vii. a theory of the relation of structure to activity of sulfanilamide type compounds1. *J. Am. Chem. Soc.*,**64**, 2905-2917.

14. Wiener H (1947) Influence of interatomic forces on paraffin properties. *J ChemPhys***15**, 766

15. Platt JR (1947) Influence of neighbor bonds on additive bond properties in paraffins. *J ChemPhys*, **15**, 419–420

16. Taft RW (1952) Polar and steric substituent constants for aliphatic and o-Benzoate groups from rates of esterification and hydrolysis of esters1. *J. Am. Chem. Soc.*, **74**, 3120-3128.

17. Hansen OR (1962) *Acta Chem. Scand.*,**16**, 1593- 1600

18. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature*, **194**, 178-180.

19. Hansch C, Fujita TP (1964) Analysis. A Method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, **86**, 1616-1626.

20. Hansch C, Steward AR (1964) The use of substituent constants in the analysis of the structure activity relationship in penicillin derivatives. *J Med Chem.*, **7**, 691–694

21. Hansch C (1969) Quantitative approach to biochemical structure-activity relationships. **Acc. Chem. Res., 2**, 232.

22. Hansch C, Lien EJ (1968) An analysis of the structure–activity relationship in the adrenergic blocking activity of the βhaloalkylamines. *BiochemPharmacol.*, **17**, 709–720.

23. Hansch C, Leo A, Nikaitani D (1972) Additive-constitutive character of partition coefficients. *J. Org Chem.*, **37**, 3090–3092.

24. Free, S.M., Jr.; Wilson, J.W. A Mathematical contribution to structure-activity studies. J. Med. Chem., 1964, 7, 395-399

25. Yalcin E, Sener SE, Owren I, and Temiz O in E. Sanz, J. Giraldo, and F. Manaut, Eds., QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, Prous Science, Barcelona/Philadelphia, 1995, p. 147.

26. Kunes J, Jachym J, Tirasko P, Odlerova Z, and Waisser K, Collect. Czech. (1997) *Chem. Commun.*, **62**, 1503.

27. Gupta SP and Paleti A (1998) *Bioorg. Med. Chem.*, **6**, 2213.

28. Fujita T, Ban T (1971) Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *J. Med. Chem.*, **14**, 148-152.

29. Kier LB (1971) Molecular orbital theory in drug research. Academic Press, New York

30. Balaban AT, Harary F (1971) The characteristic polynomial does not uniquely determine the topology of a molecule. *J Chem Doc.*, **11**, 258–259.

31. Randi´ c M (1974) On the recognition of identical graphs representing molecular topology. *J Chem Phys.*, **60**, 3920–3928.

32. Kier LB, Hall LH, Murray WJ et al. (1975) Molecular connectivity I: Relationship to nonspecificlocal anesthesia. *J Pharm Sci.*, **64**, 1971–1974.

33. Rohrbaugh RH, Jurs PC (1987) Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal ChimActa.*, **199,** 99–109.

34. Stanton DT, Jurs PC (1990) Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Anal Chem.*, **62,** 2323–2329.

35. Todeschini R, Lasagni M, Marengo E (1994) New molecular descriptors for 2D- and 3D-structuresTheory. *J Chemom.*, **8,** 263–273.

36. Katritzky AR, Mu L, Lobanov VS et al. (1996) Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J Phys Chem.*, **100,** 10400–10407.

37. Ferguson AM, Heritage TW, Jonathon P et al. (1997) EVA: A new theoretically based molecular descriptor for use in QSAR\QSPR analysis. *J Comput Aid MolDes.*, **11,** 143–152.

38. Schuur J, Selzer P, Gasteiger J (1996) The coding of the three-dimensional structure of moleculesby molecular transforms and its application to structure-spectra correlations and studies of biological activity. *J Chem Inf Comput Sci.*, **36,** 334–344.

39. Consonni V, Todeschini R, Pavan M (2002) Structure/response correlations and similarity/diversity nalysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors. *J Chem Inf ComputSci*, **42,** 682–692.

40. Kubinyi H (1976) Quantitative structure-activity relationships. IV. Nonlinear dependence of biological activity on hydrophobic character: a new model. Arzneimittelforschung, **26**, 1991-1997.

41. KierLB and HallL H (1976) Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York/London.

42. Hall LH and Kier LB J. Pharm. Sci., 64, 1978 (1975).

43. Klopman G (1984) *J. Am. Chem. Soc.*, **106**, 7315.

44. Blake BW, Enslein K, Gombar VK, and Borgstedt HH (1990) *Mutat. Res.*, **241**, 261.

45. Konemann H (1981) Toxicology, 19, 209.

46. Veith GD, Call DJ, Brooke LT (1983) *Can. J. Fish Aquat. Sci.* **40**, 743.

47. Dietrich SW, Blaney JM, Reynolds MA, Jow PY, Hansch C (1980) Quantitative structure-selectivity relationships. Comparison of the inhibition of Escherichia coli and bovine liver dihydrofolate reductase by 5-(substituted-benzyl)-2,4-diaminopyrimidines. *J Med Chem.*, **23**(11): 1205-12.

48. Hansch C, McClarin J, Klein T, Langridge R (1985) A quantitative structure-activity relationship and molecular graphics study of carbonic anhydrase inhibitors. *MolPharmacol.*,**27**(5): 493-8.

49. Hansch C, Klein T, McClarin J, Langridge R, Cornell NW (1986) A quantitative structure-activity relationship and molecular graphics analysis of hydrophobic effects in the interactions of inhibitors with alcohol dehydrogenase. J. *Med. Chem.*, **29** (5), 615-20.

50. Lewis Morgenstern S, Maurizio Recanatini SB, Teri E. Kleinll, Wayne Steinmetz S, Chun-ZhengYangll, Robert Langridgell, and Corwin Ha. Chymotrypsin Hydrolysis of X-Phenyl Hippurates A quantitative structure-activity relationship and molecular graphics. J. Bid. Chern. 1987, 262, 10767- 10772.

51. Cramer RD, III and Milne M (1979) Abstracts of Papers of the Am. Chem. Soc. M., April, Computer Chemistry Section, no. 44.

52. Kubinyi H, (ed.), '3D QSAR in Drug Design. Theory, Methods and Applications', ESCOM, Leiden, 1993.

53. KubinyiH (1993) 'QSAR: Hansch Analysis and Related Approaches', VCH, Weinheim.

54. Cramer RD, III, Persp. Drug Discov. Design, 1993,1, 269 278.

55. H. van de Waterbeemd, (ed.), 'Advanced Computer-Assisted Techniques in Drug Discovery', VCH, Weinheim, 1995.

56. Cramer RD, III. Patterson DE and Bunce JD (1988) *J. Am. Chem. Soc.*, **110**, 5959 5967.

57. Kubinyi H, ed. (1993) 3D QSAR in Drug Design. Theory, Metbods and Applications, ESCOM Science Publishers.

58. Kubinyi H, Folkers G and Martin YC eds (1997) 3D QSAR in Dpv¢g Design Ligand-Protein Interactions and Molecular Similan O; (Vol. 2) and Recent Aduances (Vol. 3), Kluwer Academic Publishers.

59. Klebe G, Abraham U, and Mietzner T (1994) J. Med. Chem., 37, 4130 4146.

60. Robinson DD, Winn PJ, Lyne PD, Richards WG (1999) Self-organizing molecular field analysis: a tool for structure-activity studies. J. Med. Chem., 42, 573-583.

61. Thomas BF, Compton DR, Martin BR, Semus SF (1991) Modeling the cannabinoid receptor: a three-dimensional quantitative structure-activity analysis. Mol Pharmacol. 40(5):656-65.

62. Waller CL, Wyrick SD, Kemp WE, Park HM, Smith FT (1994) Conformational analysis, molecular modeling, and quantitative structure-activity relationship studies of agents for the inhibition of astrocytic chloride transport. *Pharm Res.*,**11**(1), 47-53.

63. Hansch C, Gao H (1997) Comparative QSAR: Radical reactions of benzene derivatives in chemistry and biology. *Chem. Rev.*, **97**, 2995-3060.

64. Hurst T, Heritage T (1997) HQSAR - A Highly Predictive QSAR Technique Based on Molecular Holograms. In: 213th ACS Natl. Meeting, San Francisco, CA.

65. Lowis DR (1997) HQSAR: A New, Highly Predictive QSAR Technique. In: Tripos Technical Notes; Tripos Inc.: USA, Vol. 1.

66. Cho SJ, Zheng W, Tropsha A (1998) Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. J. Chem. Inf. Comput. Sci., 38, 259-268.

67. Labute P (1999) Binary QSAR: a new method for the determination of quantitative structure activity relationships. Pac. Symp. Biocomput., 444-455.

68. Pungpo P, Hannongbua S (2000) Three-dimensional quantitative structure-activity relationships study on HIV-1 reverse transcriptase inhibitors in the class of dipyridodiazepinone derivatives, using comparative molecular field analysis. J Mol Graph Model.18 (6), 581-90, 601.

69. Zefirov NS, Palyulin VA (2001) QSAR for boiling points of "small" sulfides. Are the "high-quality structure-Property-activity regressions" the real high quality QSAR models? *J Chem Inf Comput Sci.*, **41**(4), 1022-7.

70. Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model. **20** (4), 269-76.

71. Liu HX, Zhang RS, Yao XJ, Liu MC, Hu ZD, Fan BT (2004) QSAR and classification models of a novel series of COX-2 selective inhibitors: 1,5-diarylimidazoles based on support vector machines. *J Comput Aided MolDes.*, **18**(6), 389-99.

72. Basak SC, Mills D (2005) Development of quantitative structure-activity relationship models for vapor pressure estimation using computed molecular descriptors. *ARKIVOC*, (x), 308-320.

73. Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou O, Supuran CT (2006) QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. Bioorg Med Chem., 14(4),1108-1114.

74. Kuz'min VE, Artemenko AG, Muratov EN, Volineckaya IL, Makarov VA, Riabova OB, Wutzler P, Schmidtke M (2007) Quantitative structure-activity

relationship studies of [(biphenyloxy) propyl] isoxazole derivatives. Inhibitors of human rhinovirus 2 replication. *J Med Chem.*, **50**(17), 4205-13.

75. Nargotra A, Koul S, Sharma S, Khan IA, Kumar A, Thota N, Koul JL, SC Taneja, G.N. Quantitative structure activity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors. European Journal of Medicinal Chemistry 44 (2009) 229-238

76. KierLB and HallLH (1999) Molecular Structure Description. The Electrotopological State, Academic Press, San Diego, CA.

# Chapter III

# Materials and Methods

In this chapter we intended to describe the descriptors and methods which we employed in our work.

### 3.1: Calculation of molecular descriptors

#### Graph theoretical indices

A graph is the application of set itself i.e. a collection of elements of the set, and of binary relations between these elements. In case of a chemist the geometrical realization of graph is more appealing, namely a collection of points i.e. elements of the set. The name 'Graph' originates from this geometrical realization.

As the shape or length of the lines, or angles between lines has no specification so the graphs are topological rather than geometrical objects. Its most important features the vicinity relationship between points.

There are two types of correspondence between graphs and chemical categories in chemistry.

i) Structural or constitutional graphs—In this type molecules are presented by and covalent bonds are presented by lines.

ii) Reaction graphs—In this type chemical species are presented by points and the conversion between these species are presented by lines.

In our study we mainly deal with the structural or constitutional graphs only. The Graph theory was independently discovered on several occasions [1] and three names deserve special mention – Euler, Kirchoff and Cayley [2-4].

Euler published the first known paper on graph theory. It deals with Konigsberg bridge problem.

Kirchoff discovered graph theory while solving the calculation of currents in electrical networks [3]. Organic chemistry becomes the third breeding ground for graph theory. The early organic chemists Couper, Butlerov and Kekule founded the structure theory. They present a covalent bond between two atoms as line joining two points. In this way every structural formula is a graph. Cayleyput forward his concept of tree to enumeration of chemical isomers [4]. It was a challenging mathematical problem which was solved by him in 1874.

The graph theory has various applications in modern science like cryptography, networking etc.

The chemical structure of any chemical compound may be represented by graph and is termed as chemical graph. Characteristic invariants of graphs are related with structural property of molecules. These invariants are termed as topological indices.

Topological Indices:- The topological indices (TIs) are numerical invariants that quantitatively characterize molecular structure. A graph G=(V,E) is an ordered pair of two sets V and E. V represents a nonempty set and E represents unordered pair of elements of set V. When V represents the atoms of a molecule and elements of V symbolize covalent bonds between pairs of atoms, then G becomes a molecular graph. This type of graph also termed as constitutional graph, because there is no stereo-chemical information. A numerical graph invariant that characterizes the molecular structure is called a topological index. In this study we have calculated various topological indices like Information content, Structural Information content, etc. Some of the indices based on the nature of atom of its adjacent and some are depends on the bonds. Most of this can derive from the various matrices corresponding to a molecular graph.

Information theoretic topological indices are calculated by the application of information theory on chemical graphs. An appropriate set of n-elements is derived from a molecular graph G depending upon certain structural characteristics. On the basis of a equivalence relation defined on A1, the set A is partitioned into h disjoint subsets Ai of order (I=1, 2,... $h_i \sum_{i=1}^{n} n_i = n$). A probability distribution is then assigned to the set of equivalence classes.

A1, A2,A3,.....,Ah

p1, p2 .....ph

where $p_i = \frac{n_i}{n}$ is the probability that a randomly selected element of A will occur in the Ith subset.

The mean information content of an element of A is defined by Shanon's relation [11].

$$IC = -\sum_{i=1}^{n} p_i \log_2 p_i$$

The logarithm is taken at base 2 for measuring the information content in bits. The total information content of the set A is then n*IC.

In this method chemical species are symbolized by weighted linear graphs. Two vertices u0 and v0 of a molecular graph are said to be equivalent with respect to the rth order neighborhood if and only if ,corresponding to each path $u_1,u_2,....u_r$ of length r , there is a distinct path $v_1,v_2,.....v_r$ of the same length ,such that the paths have similar edge weights ,and both $u_0$ and $v_0$ are connected to the same number and type of atoms up to the rth order bonded neighbors.

Once partitioning of the vertex set for a particular order of neighborhood is completed, ICr is calculated from equation. Basak , Roy and Ghosh defined another information theoretic measure , Structural information content (SICr ) [12] which is calculated as

$$SIC_r = \frac{IC_r}{\log_2 n}$$

Where ICr is calculated from equation and n is the total number of vertices of the graph.

Another information-theoretic invariant, Complementary information content (CICr) [13], is defined as

$$CIC_r = \log_2 n - IC$$

CICr represents the difference between the maximum possible complexity of a graph (where each vertex belongs to a separate equivalence class) and the realized topological information of chemical structures [14].

**Dipole moment**

The polarity of a molecule is represented by the dipole moment. Dipole moment can be defined as the product of magnitude of charge and the distance of separation between the charges. An electric dipole consists of a pair of charges of equal magnitude and opposite signs separate by a distance if the positive and negative charges in a molecule do not overlap, the molecule possesses a permanent dipole moment (μ) (polar molecule). Generally molecular dipole moment is calculated using the following formula.

$$\mu = \sum q_i r_i$$

Where $r_i$ is the radius-vector of an atom i from the origin of the coordinate system (centre of charge or centre of mass) and $q_i$ is the partial charge of atom i

The magnitude of one or more of the vector's components along the x, y and z Cartesian axes can also be used. Dipole moment is the measure of net molecular polarity of the molecule. The attraction between two polar molecules is called dipole-dipole interaction. The dipole moment produced electrostatic interactions with biological macromolecules. The energy of dipole-dipole interactions can be described by the following equation.

$$E = \frac{2\,\mu_1\mu_2 cos\theta_1 cos\theta_2}{Dr^3}$$

Where m is the dipole moment, u is the angle between the two poles of the dipole, D is the dielectric constant of the medium and r is the distance between the charges involved in the dipole.

The magnitude of the dipole moment is used as a descriptor in the QSAR analysis and the use of dipole moments in QSAR studies was proposed by Lien et al [15-17].

**Polarizability**

The polarizability of a molecule (α) is one of the most significant electrical properties, which characterizes the ability of the electron cloud of an atom or a molecule to be distorted from its normal shape by the external electric field. Due to this distortion of electronic system an induced electric dipole moment appears. It is defined by the coefficient of proportionality between the strength of an applied electric field (E) and the magnitude of the induced dipole moment ($\mu_{ind}$) using the following equation.

$$\mu_{ind} = \alpha E$$

If a molecule have a small number of electrons, its polarizability is lower than that of the molecule containing atoms with a larger number of electrons and a more diffuse electron

distribution. Experimentally, polarizability is calculated by the Lorentz-Lorenz relation [18].

$$MR_D = \left(\frac{n_D^2 - 1}{n_D^2 + 2}\right)\frac{M}{\rho} = \frac{4}{3}\pi N_0 \alpha$$

Where, $N_0$ is the Avogadro constant, $n_D$ is the refractive index, $\rho$ is the dencity and M is the molecular mass.

The van der Waals (or London) forces are the universal attractive force between atoms that hold nonpolar molecules together in the liquid phase. This attractive force are base on polarizability and the fluctuating dipoles or shifts in electron clouds of the atoms tend to induce opposite dipoles in adjacent molecules, resulting in a net overall attraction. The energy of this interaction inversely proportion to $1/r^6$, where r is the distance separating the two molecules. The van der Waals force operates at a distance between 0.4–0.6 nm and exerts an attraction force less than 0.5 kcal/mol. although individual van der Waals forces make a low energy contribution to an event but they become significant when summed up over a large area with close surface contact of the atoms. The polarizability at the surface of both the drug and its binding site (receptor) contribute to the interaction energy. Abraham and coworkers introduced polarizability as parameters in QSAR [19, 20].

**Refractivity**

Refractivity (MR) one of the most important chemico-physical properties used as descriptor in QSAR studies. It has been shown to be related to lipophilicity, molar volume and steric bulk. The importance of splitting the MR into its atomic components for QSAR studies oriented to three-dimensional molecules was demonstrated by Crippen et al [21].

**Partition coefficient (logP)**

Partition coefficient (P) is defined as the ratio of concentration of the solute in the organic phase to the non-ionised solute concentration in the water phase, at equilibrium.

$$P = \frac{C_{org}}{C_{water}}$$

Where $C_{org}$is the equilibrium concentration of the solute in the oil phase and $C_{water}$ is the equilibrium concentration of the solute in water [22]. P denote the distribution of a compound between two phases organic and water. Partition coefficient used in its logarithmic form (logP). The zero logP value indicates that the solute is equally soluble in the two phases, a negative logP imply that the solute is more soluble in water, and a positive value suggests that a greater solubility in the oil phase. Examples of organic phases are octanol, cyclohexane and chloroform etc. LogP is probably the most commonly used descriptor of lipophilicity or hydrophobicity and it is generally understand the ability of the solute to cross lipid membranes. Lipophilicity involves many stages of drug action. Prior to reaching a pharmacological target, lipophilicity determines solubility, reactivity and degradation of drugs, as well as formulation of pharmaceuticals. Moreover, compound lipophilicity is of principal importance for biological activity as the affinity for a lipophilic environment facilitates the transport of chemicals through membranes in a biological system and the formation of complexes between compounds and receptor binding site. Cell membrane, a selectively- permeable barrier, mostly consists of a phospholipid bilayer with embedded proteins. Amphiphilic phospholipids composed of fatty acid chains at one end and hydrophilic ionized head regions at the other arrange spontaneously in the lipid bilayer. The drug interaction with lipid structures present in the organism is strongly related to its lipophilicity. Molecules with lower logP values cannot easily enter the lipid phase of the membranes, whereas molecules

**Partition coefficient (logP)**

Partition coefficient (P) is defined as the ratio of concentration of the solute in the organic phase to the non-ionised solute concentration in the water phase, at equilibrium.

$$P = \frac{C_{org}}{C_{water}}$$

Where $C_{org}$ is the equilibrium concentration of the solute in the oil phase and $C_{water}$ is the equilibrium concentration of the solute in water [22]. P denote the distribution of a compound between two phases organic and water. Partition coefficient used in its logarithmic form (logP). The zero logP value indicates that the solute is equally soluble in the two phases, a negative logP imply that the solute is more soluble in water, and a positive value suggests that a greater solubility in the oil phase. Examples of organic phases are octanol, cyclohexane and chloroform etc. LogP is probably the most commonly used descriptor of lipophilicity or hydrophobicity and it is generally understand the ability of the solute to cross lipid membranes. Lipophilicity involves many stages of drug action. Prior to reaching a pharmacological target, lipophilicity determines solubility, reactivity and degradation of drugs, as well as formulation of pharmaceuticals. Moreover, compound lipophilicity is of principal importance for biological activity as the affinity for a lipophilic environment facilitates the transport of chemicals through membranes in a biological system and the formation of complexes between compounds and receptor binding site. Cell membrane, a selectively- permeable barrier, mostly consists of a phospholipid bilayer with embedded proteins. Amphiphilic phospholipids composed of fatty acid chains at one end and hydrophilic ionized head regions at the other arrange spontaneously in the lipid bilayer. The drug interaction with lipid structures present in the organism is strongly related to its lipophilicity. Molecules with lower logP values cannot easily enter the lipid phase of the membranes, whereas molecules

with high logP values are trapped in the membrane. Therefore molecules have intermediate logP values (e.g. between about 0 and4) can readily cross cell membranes [23, 24].

LogP can be calculated by a number of experimental methods. However, various computational methods and software are available for calculation of logP for the octanol-water system.

**van der Waals surfaces area (VSA) and solvent-accessible surface area (SASA)**

Molecular surface area is the area of the outer surface of the volume from which solvent molecules are excluded due to the presence of the solute molecule in a solution. It is based on the Van der waals molecular surface (defined by the Van der Waals radii of the atoms (represented as spheres) in the molecule), however, Van der waals molecular surface contains small gaps and crevices, which are inaccessible to other atoms and molecules (for example solvent molecules). The molecular surface area is defined by excluding these gaps and crevices. Thus, the molecular surface consists of the Van der waals surfaces of the atoms where they can enter in a contact with the solvent molecules, and additionally, of the surfaces of the solvent molecules, placed in contacts with the Van der waals surfaces of two or more atoms of the investigated molecule [25].

Typically water is used as a solvent to perform calculations of the molecular surface area. For practical reasons the shape of the water molecule is considered as a sphere with a radius of 1.4 to 1.7 Å, which is the average distance from the centre of the oxygen atom to the Van der aals surface of the water molecule. A solvent-accessible molecular surface area is defined by the centre of a probe sphere (solvent molecule, typically water), when it is rolled over the molecular surface.

**Energies of the frontier molecular orbital's HOMO and LUMO**

Partial atomic charges appear due to the different ability of atoms to withdraw electron The energies of the frontier orbitals HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) are very popular quantum chemical descriptors in QSAR. These orbitals play an important role in many chemical reactions and determining electronic band gaps in solids [26]. These orbitals are also responsible for the formation of many charge transfer complexes [27, 28]. According to the frontier molecular orbital theory of chemical reaction, transition state formation involves the interaction between the frontier orbitals (HOMO and LUMO) of reacting species. Thus, the treatment of the frontier molecular orbitals separately from the other orbitals is based on the nature of chemical reactions [29]. The HOMO energy is directly related to the ionization potential and characterizes the susceptibility of the molecule toward attack by electrophiles, where as the energy of the LUMO is directly related to the electron affinity and characterizes the susceptibility of the molecule toward attack by nucleophiles. A higher HOMO energy imply higher affinity of a molecule to react as a nucleophile and a lower LUMO energy suggests stronger electrophilic nature of a molecule. The difference in energy between the HOMO and the LUMO is an important for the stability. A large HOMO-LUMO gap implies high stability for the molecule i.e. the lower reactivity in chemical reactions The HOMO and LUMO energies are calculated with the methods of the quantum mechanics [30,31].

Commonly used software for calculating Molecular Descriptors are HyperChem (HyperChem 5.1, Hypercube Inc., Gainesville, Florida, USA), ACD/LogP KOWWIN, Pallas, TOPKAT, Dragon (TALETE srl), GAMESS, MOPAC, MervinlogP calculator,

**3.2 Docking simulation methods**

With the rapid increase in computational power, in silico methods became widely used in the fields of structural molecular biology and structure-based drug design. Molecular docking is one of these computational techniques [32, 33]. Docking is a method involves the prediction of the preferred orientation of one molecule to second which bound to each other to form a stable complex. Preferred orientation of the molecule may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions. Docking is generally devised as a multi-step process in which each step introduces one or more additional degrees of complexity. The process begins with the application of docking algorithms that fit the small molecules in the active site of substrate.

Examples of some popular protein-ligand docking systems include AutoDock [34], GOLD [35], DOCK [36], GLIDE [37], ICM [38], FlexX [39] and SITUS [40]. Of these, molecular docking simulation was carried out using the autodock 4.2. AutoDock has verified software capable of quickly and accurately predicting bound conformations and binding energies of ligands with protein.

AutoDock employed a grid-based method, which permit rapid evaluation of the binding energy of trial conformations .Grid maps are calculated by AutoGrid. AutoDock need pre-calculated grid maps. Grid maps of each atom type present in the ligand being docked. Grids maps help to make the docking calculations extremely fast. A grid map consists of a three dimensional lattice of regularly spaced points, surrounding and centered on some region of interest of the macromolecule under study. Typical grid point spacing varies from 0.2 Å to 1.0 Å (default value is 0.375 Å). Each point within the grid map stores the potential energy of a probe atom or fictional group that is due to all the atoms in the macromolecule.

Morris et al. consider three search methods, the Lamarckian genetic algorithm, Monte Carlo simulated annealing, anda traditional genetic algorithm. The primary method for conformational searching is a Lamarckian genetic algorithm [41]. A population of trial conformations is created, and then in successive generations these individuals mutate, exchange conformational parameters, and compete in a manner analogous to biological evolution, ultimately selecting individuals with lowest binding energy. The "Lamarckian" aspect is an added feature that allows individual conformations to search their local conformational space, finding local minima, and then pass this information to later generations. The Lamarckian Genetic Algorithm provides the most efficient search for general applications and in most cases will be the technique used. It is typically effective for systems with about 10 rotatable bonds in the ligand. A Lamarckian genetic algorithm combined with a scoring function based on the AMBER force field [42].

MonteCarlo (MC) methods are amongthe most established and widely used stochastic optimization techniques [43]. The combination of atomistic potential energy models with stochastic search techniques has produced some of the most powerful methods for both structure optimization and prediction.

The docking simulation is carried out using the Metropolis method, also known as Monte Carlo simulated annealing. With the protein static throughout the simulation, the substrate molecule performs a random walk in the space around the protein. At each step in the simulation, a small random displacement is applied to each of the degrees of freedom of the substrate: translation of its center of gravity; orientation; and rotation around each of its flexible internal dihedral angles.

This displacement results in a new configuration, whose energy is evaluated using the grid interpolation procedure described above. This new energy is compared to the energy

of the preceding step. If the new energy is lower, the new configuration is immediately accepted. If the new energy is higher, then the configuration is accepted or rejected based upon a probability expression dependent on a user defined temperature, T. The probability of acceptance is given by

$$P(\Delta) = e^{(-\frac{\Delta E}{k_B T})}$$

Where ΔE is the difference in energy from the previous step, and kB is the Boltzmann constant. At high enough temperatures, almost all steps are accepted. At lower temperatures, fewer high energy structures are accepted.

The simulation proceeds as a series of cycles, each at a specified temperature. Each cycle contains a large number of individual steps, accepting or rejecting the steps based upon the current temperature. After a specified number of acceptances or rejections, the next cycle begins with a temperature lowered by a specified schedule such as

$$T_i = gT_{i-1}$$

Where Ti is the temperature at cycle i, and g is a constant between 0 and 1.

Simulated annealing allows an efficient exploration of the complex configurational space with multiple minima that is typical of a docking problem. The separation of the calculation of the molecular affinity grids from the docking simulation provides modularity to the procedure, allowing the exploration of a range of representations of molecular interactions, from constant dielectrics to finite difference methods and from standard 12-6 potential functions to distributions based on observed binding sites.

The Genetic Algorithm may also be run without the local search, but this is typically less efficient than the Lamarckian genetic algorithm combination.

AutoDock4 uses a semiempirical free energy force field to predict binding free energies of small molecules to macromolecular targets. The semiempirical free energy force field estimates the energetic of the process of binding of two (or more) molecules in a water environment using pair-wise terms to evaluate the interaction between the two molecules and an empirical method to estimate the contribution of the surrounding water. This differs from a traditional molecular mechanics force field, which also relies on pair-wise atomic terms, but typically uses explicit water molecules to evaluate solvation contributions. The goal of the empirical free energy force field is to capture the complex enthalpic and entropic contributions in a limited number of easily evaluated terms.

The protein and ligand start in an unbound conformation. The force field evaluates the free energy of binding in two steps. The free energy of binding is estimated to be equal to the difference between (1) the energy of the ligand and the protein in a separated unbound state and (2) the energy of the ligand and protein in their bound conformation (complex). Evaluate the intramolecular energetics of the transition from the unbound state to the bound conformation for each of the molecules separately and then evaluate the intermolecular energetic of bringing the two molecules together into the bound complex.

The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding ($\Delta S_{conf}$):

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

In equation 1, L refers to the "ligand" and P refers to the "protein" in a protein–ligand complex; this approach is equally applicable for any types of molecules in a complex. The first two terms are intramolecular energies for the bound and unbound states of the ligand, and the following two terms are intramolecular energies for the bound and unbound states of the protein. The third parentheses represent the change in

intermolecular energy between the bound and unbound states of protein and ligand. It is assumed that the two molecules are sufficiently distant from one another in the unbound state that $V_{unbound}^{P-L}$ is zero. The bound state of the protein is identical with the protein unbound state, and the difference in their intramolecular energy is zero.

The pair-wise atomic terms include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

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

$$W_{sol} \sum_{i,j}(S_i V_j + S_j V_i) e^{-\frac{r_{ij}^2}{2\sigma^2}}$$

The weighting constants W are the ones that are optimized to calibrate the empirical free energy based on a set of experimentally determining binding constant. The first term is a typical 6/12 potential for dispersion/repulsion interactions. The Parameters A and B were taken from the Amber force field [44]. The second term is a directional H-bond term based on a 10/12 potential [45]. The parameters C and D are assigned to give a maximal well depth of 5 kcal/mol at 1.9 A° for hydrogen bonds with oxygen and nitrogen and a depth of 1 kcal/ mol at 2.5 A° for hydrogen bonds with sulpher [46,47]. Directionality of the hydrogen bond interaction represented by the function E(t), is dependent on the angle t away from ideal bonding geometry [48]. Directionality is further enhanced by limiting the number of hydrogen bonds available to each point in the grid to the actual number of hydrogen bonds that could be formed. The third term is a screened Coulomb potential for electrostatic interactions [49]. The final term is a desolvation potential based on the volume (V) of the atoms surrounding a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance. The distance weighting factor σ is set to 3.5 A° [50].

This force field is standardized for a united atom model, which explicitly includes heavy atoms and polar hydrogen atoms. Intramolecular energies are calculated for all pairs of atoms within the ligand (or protein, if it has free torsional degrees of freedom), excluding 1–2, 1–3, and 1–4 interactions.

The term for the loss of torsional entropy upon binding ($\Delta S_{conf}$) is directly proportional to the number of rotatable bonds in the molecule ($N_{tors}$):

$$\Delta S_{conf} = W_{conf} N_{tors}$$

The number of rotatable bonds includes all torsional degrees of freedom, including rotation of polar hydrogen atoms on hydroxyl groups and the like.

**Desolvation**

In the development of an empirical free energy function for AUTODOCK, the desolvation term was most challenging, because AUTODOCK uses a grid based method for energy evaluation, and most published solvation methods are based on surface area calculations.

The desolvation term is calculated using the general approach of Wesson and Eisenberg [51]. For the Calculate of desolvation energy, two information are needed (1) an atomic solvation parameter for each atom type, which is an estimate of the energy needed to transfer the atom between a fully hydrated state and a fully buried state and (2) an estimate of the amount of desolvation when the ligand is docked. The amount of desolvation is calculated by the volume-summing method, which is similar to the Stouten et al. method [50]. Huey et al. introduced a modified approach for the atomic solvation parameters based on the chemical type and the atomic charge of the atom. This approach employed in AutoDock and other docking methods. Incorporation of the atomic charge

into the salvation parameter removes the need to use two discrete charged and uncharged atom types for oxygen and nitrogen.

The solvation parameter ($S_i$) for a given atom is calculated as:

$$S_i = (ASP_i + QASP \times |q_i|)$$

Where $q_i$ refers to the atomic charge and ASP and QASP are the atomic solvation parameters derived here. The ASP is calibrated using six atom types such as aliphatic carbons (C), aromatic carbons (A), nitrogen, oxygen, sulfur, and hydrogen. A single QASP is calibrated over the set of charges on all atom types.

For each atom in the protein, the volume term in the free energy equation was evaluated:

$$\Delta V_i = \sum_{k \neq i} V_k \times e^{-\frac{k}{2\sigma^2}}$$

Where k is all atoms in the protein and all atoms in the same amino acid residue represented as i. The maximal value of $\Delta V$ for each amino acid type over the entire set of proteins was then determined. These values were used to perform a least-squares fit of the model to a set of experimental vacuum-to-water transfer energies to determine values for the atomic solvation parameters ASP and QASP [52].

We used a simple approximation for incorporation of additional atom types in the desolvation model. The ASP is assigned to the average of the values from the six atom types used in the calibration and the same QASP is applied.

**Unbound States**

In order to estimate a free energy of binding, AutoDock needs to estimate energy for the unbound state of the ligand and protein. Morris et al. investigated three approaches to the

unbound state. The first approach is a fully extended conformation, which models a fully solvated conformation with few internal contacts. A short optimization was performed on the ligand in isolation using a uniform potential inversely proportional to the distance between each pair of atoms. This moves all atoms as far away from one another as possible. The second approach is a minimized conformation that has substantial internal contacts, modeling a folded state for the unbound ligand. A short Lamarckian genetic algorithm conformational search was performed, using an empty affinity grid. As expected, these conformations tend to bury hydrophobic portions inside and form internal hydrogen bond interactions.

The third approach the assumption used in AutoDock3 and many other docking methods. In this, it is assumed that the conformation of the unbound state is identical to the conformation of the bound state.

**Coordinate Sets**

The force field was standardized and tested using a large collection of protein complexes for which experimental information on binding strength is available. The force field was calibrated on a set of 188 complexes. Binding data were collected from the Ligand–Protein Database (http://lpdb.scripps.edu), and coordinates were obtained from the Protein Data Bank (http://www.pdb.org). These complexes were checked and corrected if necessary. Hydrogen atoms were added automatically using Babel, atomic charges were added using the Gasteiger PEOE method, and then nonpolar hydrogen atoms were merged [53-54]. The Gasteiger method was selected for its fast and easy operation and ready availability as part of Babel. Ligands were processed in ADT (http://autodock.scripps.edu/resources/adt) to assign atom types and torsion degrees of freedom. Finally, a short optimization of the ligand was performed.

**Redocking**

Redocking experiments were carried out with AutoDock4 and the new empirical free energy force field. For each complex, 50 docking experiments were performed using the Lamarckian genetic algorithm conformational search with the default parameters from AutoDock3. A maximum of 25 million energy evaluations was applied for each experiment. The results were clustered using a tolerance of 2.0 A°.

## 3.3: Molecular dynamics simulation methods

Molecular dynamics is an important tool to investigate the microscopic behaviors by integrating the motions of particles or particle clusters, based on Newtonian dynamics. Theoretically it has been applied to study the dynamics of a macromolecular system. Now molecular dynamics simulations have been used most widely for studying protein motions.

At the commencement of a dynamic simulation, an initial set of atomic coordinates and velocities are needed. Generally, these are obtained from the X-ray crystallographic or NMR structure data, or by model-building (based on the structure of a homologous protein, for example). Given a set of coordinates, a preliminary calculation serves to equilibrate the system.

The structures are first refined using an energy minimization algorithm to reduce local stresses due to the nonbonded atomic overlaps, distortions of bond length, etc.

Then the protein atoms are assigned velocities (v) taken at random from a Maxwellian distribution corresponding to a low temperature, and a simulation is performed for a period of a few picoseconds.

This is done bythe laws of classical mechanics, and most notably the Newton's 2nd law:

$$F_i = m_i a_i$$

Where $F_i$ the force acting upon ith particle at the time, $m_i$ is the atomic mass of the particle, $a_i$ is the acceleration ($a_i = d^2r_i/dt^2$) and introducing it into the equation for the position $r_i$ at time $t+\Delta t$, given $r_i$ at time t:

$$r_i(t+\Delta t) = r_i(t)+v_i\Delta t+\frac{1}{2}a_i(\Delta t)^2$$

The equilibration is continued by alternating new velocity assignments, chosen from Maxwellian distributions corresponding to successively increased temperatures with intervals of dynamical relaxation. The temperature T is measured in terms of the mean kinetic energy for the system composed of N atoms:

$$\frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3}{2} N k_B T$$

Where $v^2_i$ *is the* average velocity squared of the ith atom and $k_B$ is the Boltzmann constant. The equilibration period is considered completed when the temperature is stable for longer than about 10 ps, the atomic momenta obey maxwellian distribution and different regions of the protein have the same average temperature.

Integration of the equations of motion after equilibration generates the coordinates and velocities of the atoms as a function of time. Many numerical algorithms are used to solve the equation of motion such as Predictor-corrector algorithm, Verlet algorithm and Leap-frog algorithm etc.

**Predictor corrector algorithm**

Predictor-corrector algorithms compose commonly used class of methods to integrate the equations of motion and more frequently used in molecular dynamics due to Gear, and consists of three steps, namely :Predictor, Force evaluation, Corrector [55].

If the classical trajectory is continuous then an estimate of the positions, velocities etc. at time t+$\partial$t may be obtained by Taylor expansion about time t:

$$r^p(t+\partial t) = r(t) + v(t)\partial t + \frac{1}{2}a(t)\partial t^2 + \ldots\ldots$$

$$v^p(t+\partial t) = v(t) + a(t)\partial t + \frac{1}{2}b(t)\partial t^2 + \ldots\ldots$$

$$a^p(t+\partial t) = a(t) + b(t)\partial t + \ldots\ldots$$

Where r represents the position, $v$ is the velocity (the first derivative with respect to time), a is the acceleration (the second derivative with respect to time), etc. But the above equation will not generate correct trajectories as time advances and not introduced the equation of motion. These enter through the correction step. Calculated the new position $r^p$, the forces at time t+$\partial$t and hence the correct accelerations $a^c$ (t+$\partial$t). These can be compared with the predicted acceleration from the above equation.

To estimate the size of the error in the prediction step:

$$\Delta a(t+\partial t) = a^C(t+\partial t) - a^P(t+\partial t)$$

This error and the results of predictor step are fed into the corrector which gives;

$$r^C(t+\partial t) = r^P(t+\partial t) + c_o \Delta a(t+\partial t)$$

$$V^C(t+\partial t) = v^P(t+\partial t) + c_1 \Delta a(t+\partial t)$$

$$a^c(t+\partial t) = a^P(t+\partial t) + c_2 \Delta a(t+\partial t)$$

The idea is that $r^c(t+\partial t)$ etc are now better approximation to the true positions, velocities etc. The general scheme of a stepwise MD simulation based on a predictor-corrector algorithm, which may be summarized as follow:

(a) Predict the positions velocities accelerations at time $(t+\partial t)$ using the correct values of these equation.

(b) Evaluate the forces and hence accelerations $a_i = f_i/m_i$, from the new position.

(c) Correct the predicted positions velocities accelerations using the new acceleration.

(d) Calculate any variables of interest such as energy, order parameters before returning to a for the next step.

**Verlet algorithm**

One of the most simplest and common method of integrating the equation of motion is called Verlet algorithm [56-57]. The method is a direct solution of the second order equations. The Verlet algorithm uses positions and accelerations at time $t$ and the positions from time $t-\partial t$ to calculate new positions at time $t+\partial t$. The Verlet algorithm uses no explicit velocities.

$$r(t + \partial t) = r(t) + v(t)\partial t + \tfrac{1}{2}a(t)\partial t^2$$

$$r(t - \partial t) = r(t) - v(t) + \tfrac{1}{2}a(t)\partial t^2$$

Summing these two equations, one obtains

$$r(t + \partial t) = 2r(t) - r(t - \partial t) + a(t)\partial t^2$$

The velocities do not explicitly appear in Verlet algorithm but they are useful for calculating the kinetic energy and hence the total energy. They may be obtained from the formula,

$$V(t) = \{ r(t+\partial t) - r(t-\partial t)\}/2\partial t$$

The velocities are not required to compute the trajectories, but they are useful for calculating observables like the kinetic energy. Success of the Verlet algorithm is straightforward and also storage requirements are modest, comprising two sets of positions (r(t) and r(t-δt)) and accelerations(a(t)).

**Leap-frog algorithm**

Leap-frog integration is equivalent to calculating positions and velocities at interleaved time points, interleaved in such a way that they "leapfrog" over each other. In this algorithm, the velocities are first calculated at time *t+ 1/2dt;* these are used to calculate the positions, *r,* at time *t+dt.* In this way, the velocities leap over the positions, and then the positions *leap* over the velocities [58-59].

$$r(t + \partial t) = r(t) + v(t + \tfrac{1}{2}\partial t)\partial t$$

$$v(t + \tfrac{1}{2}\partial t) = v(t - \tfrac{1}{2}\partial t) + a(t)\partial t$$

The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time $t$ can be approximated by the relationship:

$$v(t) = 1/2[v\{t-½\partial t) + v(t+½\partial t)]$$

In our study MD simulation is performed using GROMACS software.

Different types of software are used in molecular dynamics simulation. Some common and widely used software are given in appendix III.

**3.4: References**


1. Hrary F (1969) Graph theory. Chapter 1, Addision Wesley, Reading mass
2. Euler. L. (1736) Comment. Acad. Sci. I. Petropolitanae, 8, 128, translated in Sci. Amer, 189, 66 (1953)
3. Kirchoff. G. (1842) Ann. Phys. Chem. 72,497.
4. Cayley A. (1874). Philos. Mag. 67, 444.
5. M. Randic, P.J. Hansen and P.C. Jurs, J. Chem. Inf. Comput. Sci. 28 (1988) 60.
6. R. Todeschini and V. Consonni, Handbook of Molecular Descriptors (Wiley-VCH, Weinheim, 2000).
7. M. Randic, On characterization of molecular branching, J. Amer. Chem. Soc. 97 (1975) 6609.
8. Wiener HJ (1947) *Am. Chem. Soc.*, **69**, 17-20
9. Harary F (1969) Graph theory. Addison-Wesley, Reading MA.

10. Balaban AT, Harary F (1968) Chemical graphs. V. Enumeration and proposed nomenclature of benzenoidcata condensed polycyclic aromatic hydrocarbons. *Tetrahedron*. **24**, 2505–2516.

11. Shanon CE (1948) A mathematical theory of Communication. *Bell Syst. Tech.J.*, *27*: 379-423.

12. Basak SC, Roy AB, Ghosh JJ (1979) Study of Structure-Function Relationship of Pharmacological and Toxicological Agents Using Information Theory. In proceeding of the llnd International conference on mathematical modeling, University of Missouri at *Rolla*. 851-856.

13. Basak SC and Magnuson VR (1983) Molecular Topology and Narcosis. A Quantitative Structure-Activity Relationship (QSAR) Study of Alcohols Usins Complementary Information Content, Arzneimittle-Froschung, *Drug Research*. 501-503.

14. Lien EJ, Liao RCH, Shinouda HGJ (1979) *Pharm. Sci.*, 68: 463-465.

15. Lien EJ, Guo ZR, Li RL, Su CTJ (1982)*Pharm. Sci.* **71**: 641-655.

16. Li WY, Guo ZR, Lien EJJ (1984) *Pharm. Sci.* **73**: 553-558.

17. Kurt E. Oughstun and Natalie A (2003) Cartwright. On the Lorentz-Lorenz formula and the Lorentz model of dielectric dispersion. OPTICS EXPRESS. **11:** 1541.

18. Abraham MH (2004)The factors that influence permeation across the blood-brain barrier. *European Journal of Medicinal Chemistry*, **39**:235–240.

19. Gaudio AC, Korolkovas A,Takahata Y (1994) Quantitative structure activity relationships for 1,4-Dihydropyridine calcium channel antagonists (nifedipine analogs) - a quantum chemical classical approach. *Journal of pharmaceutical sciences*, **83**: 1110–1115.

20. Ghose AK, Crippen GM (1987) Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure-Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem.ln$Comput.Sci.*,**27**: 21-35.

21. Smith RN, Hansch C, Ames MM (1975)*J. Pharm. Sci.*, **64**: 599.

22. Tomlinson E, David SS, Parr GD, James M, Farraj N, Kinkel JFM, Gaisser D, Wynn HJ in W. J. Dunn III, J. H. Block, and R. S. Pearlman, Eds., Partition Coefficient, Determination and Estimation, Pergamon, Oxford, UK, 1986, p. 83.

23. Hansch C, Kerley R (1970) Role of the Benzyl Moiety in Biochemical and Pharmacological Processes. *J. Med. Chem.*,**13**: 957-964.

24. Iwamura H (1980) Structure-Taste Relationehip of Perillartrine and Nitro and Cyanoaniline Derivatives. *J. Med. Chem.*, **23**: 308-312. D. E. Clark. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. Journal of Pharmaceutical Sciences, 88(8):815–821, 1999.

25. Zhou Z, Parr RGJ (1990) Am. Chem. Soc., 112, 5720.

26. Franke R (1984) Theoretical Drug Design Methods; Elsevier: Amsterdam. 115-123.

27. Osmialowski K, Halkiewicz J, Radecki A, Kaliszan RJ (1985).*Chromatogr.*, 346, 53.

28. Fukui K (1975) Theory of Orientation and Stereoselection; Springer Verlag: New York. 34-39.

29. Sklenar H, Ja¨ger J (1979)*Int. J. Quantum Chem.*, **16**: 467.

30. Fleming I (1976) Frontier Orbitals and Organic Chemical Reactions; John Wiley & Sons: New York.

31. Lengauer T,Rarey M, (1996) Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.*, **6**:402–406.

32. Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. *Journal of Molecular Recognition*, **9**:175–186.

33. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK,Olson AJ(1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, **19**: 1639–1662.

34. Jones G, Willett P, Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, **245**:43–53.

35. Shoichet BK, Kuntz ID, Bodian DL (1992) Molecular docking using shape descriptors. *Journal of Computational Chemistry*, 13:380–397.

36. Repasky MP, Shelley M, Friesner RA (2007) Flexible ligand docking with Glide. *Current Protocols in Bioinformatics*, 8.

37. Abagyan R, Totrov M,Kuznetsov D (1994) ICM - a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, **15**: 488–506.

38. Rarey M, Kramer B, Lengauer T,Klebe G () A fast flexible docking method using an incremental construction algorithm, *J Mol Biol.*,**261**: 470--489,.

39. Chacon P, Wriggers W (2002) Multi-Resolution Contour-Based Fitting of Macromolecular Structures. *Journal of Molecular Biology*, **317**: 375–384.

40. Morris GM, Goodsell DS, Halliday RS, HueyR, Hart WE, Belew RK, Olson AJJ (1998) *Comput Chem.*, **19**: 1639.

41. Cornell WD, Cieplak P,Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW,Kollman PA (1995)A SecondGeneration Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, **117**, 5179–5197.

42. Hart TN, Read RJ (1994) Multiple-start Monte Carlo docking of flexible ligands. Birkhaeuser, Boston.

43. Weiner SJ, Kollman PA,Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner PJ (1984)*Am Chem Soc.*, **106**: 765.

44. Morris GM, Goodsel DS, Halliday R, Huey R, Hart WE, Belew RK, Olson AJJ (1998) *Comput Chem.*, **19**, 1639.

45. Boobbyer DNA,Goodford PJ, McWhinnie PM, Wade RCJ (1989)*Med Chem.*, *32*: 1083.

46. Huey R,Goodsell DS, Morris GM, Olson AJ (2004)*Lett Drug Des Discov.*, **1**: 178.

47. Mehler EL, Solmajer T (1991) *Protein Eng.*,**4**, 903.

48. Stouten PFW, Frommel C, Nakamura H, Sander C (1993)MolSimul.,**10**, 97.

49. Wesson L, Eisenberg D (1992)*Protein Sci.*, **1**, 227.

50. Sharp KA, Nicholls A, Friedman R, Honig (1991)*Biochemistry*, **30**: 9686.

51. Tetko IV, Gasteiger J, Todeschini R., Mauri A, LivingstoneDJ, et al. (2005)*Comput Aided Mol Des*, **19**: 453.

52. Gasteiger J, Marsili M (1980)*Tetrahedron*, **36**: 3219.

53. C Gear (1971) Numerical Initial Value Problems in Ordinary Differential Equation, Prentice-Hall, Englewood Cliffs, NJ.

54. Allen M, Tildesley (1987) Computer Simulation of Liquids, Clarendon Press, New, York.

55. Frenkel D, SmitB, (2002) "Understanding molecular simulation: from algorithms to applications," 2nd ed., Academic Press, San Diego.

56. Gunsteren v, WF, Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. Mol. Sim. 1:173–185, 1988.

57. Hockney RW, Goel SP, Eastwood J (1974) Quiet High Resolution Computer Models of Plasma. J. Comp. Phys. 14:148–158.

# Chapter IV

# Molecular Docking and MD simulation of human renin: implication on its binding site

## 4.1: Introduction

Renin is a hormone enzyme produced from the inactive protein prorenin [1]. Renin initiates renin-angiotensin system (RAS) producing the angiotensin peptides that control blood pressure, cell growth, apoptosis and electrolyte balanced [2-3]. Renin is highly specific aspartic proteinases and mainly produced by Juxtaglomerular cell in the kidney. The Secretion of renin from Juxtaglomrular cell is controlled by several mechanisms, including the sympathetic nervous system, salt and fluid balance, and blood pressure [4-7]. It cleaves angiotensinogen to form the decapeptide angiotensin I. Then inactive decapeptide converted to active octapeptide angiotensin II by the angiotensin converting enzyme (ACE). Next the angiotensin II binds to the type 1 angiotensin II receptors (AT1) [8]. Stimulation of type 1 angiotensin II receptor increases arterial tone and also the secretion of aldosterone. Therefore angiotensin II plays a key role in blood pressure, fluid and electrolyte homeostasis [9].

Inhibiting the renin-angiotensin system reduced stimulation of the AT1 receptor and thereby gives therapeutic benefits for the treatment of hypertension and congestive heart failure. Beta blockers are the original renin-angiotensin system inhibitor that inhibits the renin release from the kidney. Reduced renin secretion leads to decreased concentrations of angiotensinI and II. This is the benefits of beta blockade for protecting the heart from the heart attack [10]. Renin inhibitors are more efficient than ACE inhibitors and AT1 receptor antagonists. Renin inhibitors also have fewer side effects than the ACE inhibitors and AT1 receptor antagonists [11]. Renin is highly specific for only angiotensinogen. Therefore renin identified as an ideal target for antihypertensive drugs. The potent inhibitors of renin could give a new alternative way to treat high blood pressure without inhibiting other biological substances and thereby no side effects [12].

The design and development of renin inhibitors were started with peptidomemtic and it is specific for the angiotensinogen. The synthetic peptide inhibitors were not successful in treating hypertension as renin inhibitors due to the unfavorable pharmacokinetic behavior [13-16]. Therefore there was need new classes of nonpeptidic renin inhibitors that fulfill the criteria for becoming successful drugs. Now modern research has lead to the discovery of several classes of non-peptidic renin inhibitors. Aliskiren is the first marketed non-peptidic renin inhibitor [17, 18].

The human renin contains 340 amino acids. Structurally renin consist two homologous lobes with an active site at interface [13, 18]. The catalytic activity of the active site is due to two aspartic acid residues (Asp38, Asp226), one located in each lobe of the renin molecule. The active site of renin can accommodate seven amino acid residues of angiotensinogen and cleaves the Leu10-val11 peptide bond within angiotensinogen to generate angiotensin I [19].

In this study MD simulation was performed to consider the flexibility of protein. Molecular docking studies were employed to determine the binding mode of renin and its inhibitor.

## 4.2: Materials and Methods

### Computational details

### Preparation of protein

In the present article crystallographic structure of Human renin in complex with ligand 72X was obtained from Protein Data Bank (PDB code 3GW5), which have 669 residues with two chains (A and B) [20]. Missing atoms were repaired by the SPDBV software package [21]. We have removed all ligands from protein to get the free protein. For

docking simulation using Autodock all polar hydrogen was added with the GROMACS modeling package. The resulting structure was optimized 54 steps of conjugate gradient minimization, employing the GROMACS force field. During minimization, the heavy atoms were kept fixed at their initial crystal coordinates by restraing. Minimization was carried out under a vacuum medium. Electrostatic interactions were calculated using the cut-off method. Finally, solvation parameters were added using the ADDSOL utility of AutoDock 4.2. Default values of atomic solvation parameters were used throughout the calculations. The grid maps of the protein were used in the docking experiments was calculated using the AutoGrid utility program.

**Preparation of inhibitor**

Piperidine derivative 72X of renin inhibitor was collected from Protein Data Bank. Structure of renin inhibitor 72X is represented in Figure 1. For docking experiments with AutoDock 4.2, ligand molecule was optimized, and saved as in pdb format with the aid of Arguslab 4.2. Next ligand is loaded in AutoDock Tool (ADT). Gasteiger charges are added and 43 non polar hydrogens are merged. ADT also detected 15 rotatable bonds and 13 numbers of torsional degrees of freedom (TORSDOF). ADT selected a root with the minimum number of rotatable branches. Root was detected. Next ligand was saved in PDBQT format. Then prepared ligand was used in docking simulation in the next step.

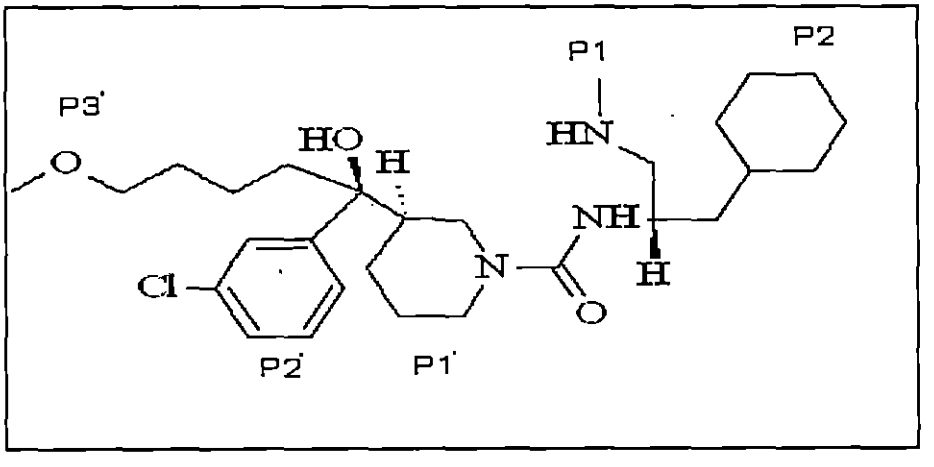


Fig 1: Structural representation of renin inhibitor 72X.

The piperidine and Teoc derivatives attached by amide linkage (at the P1-P1' position) in Renin inhibitor 72X [20]. We have divided the molecule in several moieties. P1' residue of 72X is a Piperidine moiety. The P2' residue of the renin inhibitor is a Chlorobenzene moiety. Similarly P3', P1, P2 residues are Methoxy butyl, (methyl amine) methyl, cyclohexyl moieties respectively.

We have listed the binding site residues of 72X from pdb and cross checked it by taking the residues within 3.5 Å from the inhibitor 72X. We selected 18 residues whose RMSD and Docking simulation have been performed. Each of these residues selected as a flexible residue in Autodock to evaluate the flexible docking.

**MD Simulation**

The MD simulation was carried out using GROMACS [24]. The 2.00 A° resolution x-ray structure of Human renin (PDB code 3GW5) was used as a starting structure. We have carried out MD simulation of free protein not the complex. The protein was solvated with SPC water molecules in a cubic box, having an edge length of 3.5 A°. The LINCS algorithm was employed to constrain all bond lengths (25). The simulation was conducted at a constant temperature (300K) and the Berendsen coupling method was used for coupling each component separately to a temperature bath [26]. To calculate longer-rang electrostatic contribution on a grid with spacing and a cutoff of 1.0 nm for Coulomb interaction we employed Lennard Jones interaction and the particle mesh Ewald method. MD simulation was performed for 6 ns. Before running simulation, an energy minimization was performed by steepest descent (sd) method. After that the positional

restraints were released and simulation is performed for 6ns with time step 2 fs. Finally the end of the simulation the respective trajectory files were examined with different tools of GROMACS.

**Docking**

Molecular docking studies were performed with AutoDock 4.2 using a Lamarckian genetic algorithm (LGA) to evaluate the inhibitor–enzyme interactions [27]. A grid map with 62 x 66 x 62 Å points with a grid spacing of 0.375 Å was generated using AutoGrid 4.2 and the grid was centered at x, y, z coordinates of 95.848, 108.694, 107.689. The distance dependent function of the dielectric constant was used for the calculation of the energy map. At the end of the docking experiment we get 10 dock conformers with different free energy of binding. The best conformer was chosen based on the lowest free energy of binding. The protein with the best conformer is saved as complex and analyzed using Molegro Molecular Viewer [28].

**4.3: Results and Discussion**

The overall structural stability of the free protein during the simulation has been monitored using several parameters likes the radius of gyration (Rg), RMSF and RMSD of individual residues were calculated over the course of the simulation.

The variation of radius of gyration (Rg) as a function of time is presented in Figure 2 and from this figure it is clear that the initial Rg value is 2.68172 and then Rg value decreases up to 437pswith Rg value 2.6126, after that Rg slightly increases up to 5261ps(2.69299). The overall plot of Rg during the simulation shows a periodical nature.

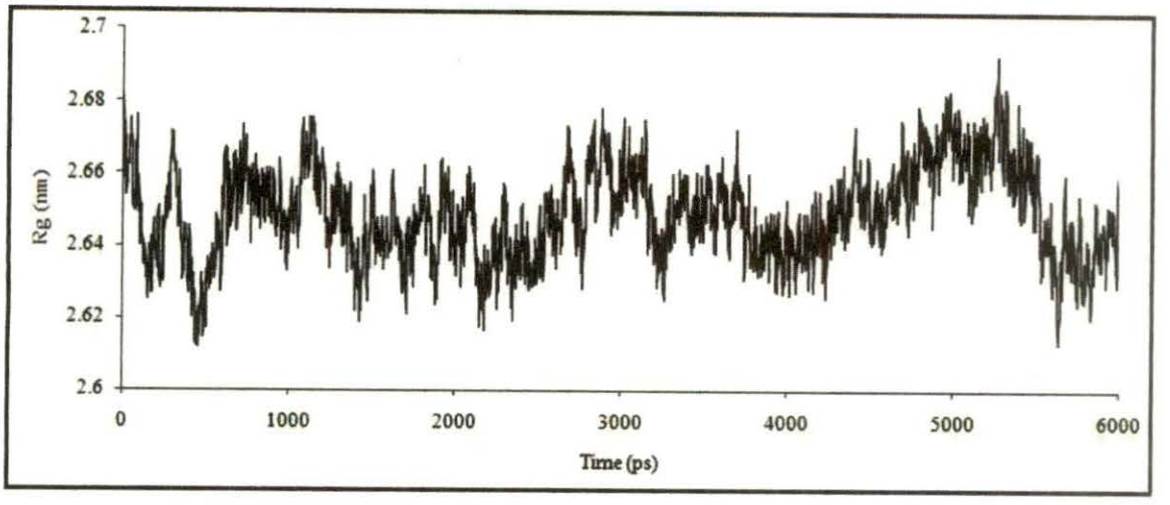
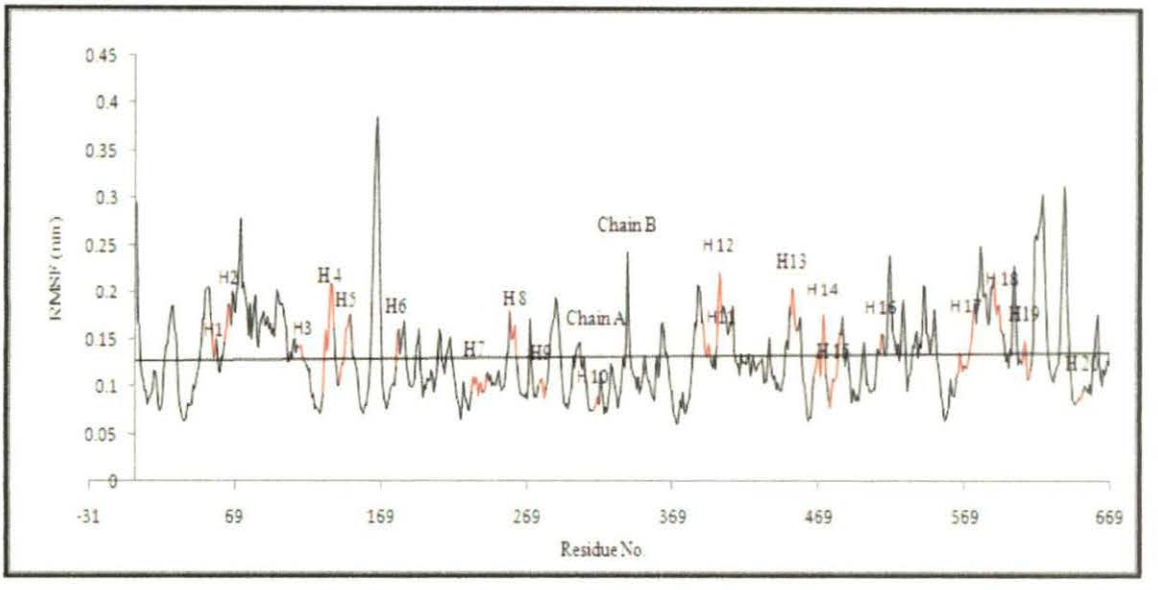



**Fig. (2). Radius of gyration (Rg) as a function of time with respect to starting structure during the MD simulations is shown for human renin.**

The flexibility of different segments of the protein is also revealed by looking at the root mean-square fluctuation (RMSF) of each residue from its time-averaged position is presented in Figure3.




| Helix(H) of Chain A | Residue No. in pdb | Residue No. after MD |
|---|---|---|
| H1 | 55-59 | 52-56 |
| H2 | 65-69 | 62-66 |

| Helix(H) of Chain B | Residue No. pdb | Residue No. after MD |
|---|---|---|
| H11 | 56-61 | 390-395 |
| H12 | 65-69 | 399-403 |

| Helix(H) of Chain A | Residue No. in pdb | Residue No. after MD |
|---|---|---|
| H3 | 11120 | 112-117 |
| H4 | 132-140 | 129-137 |
| H5 | 142-151 | 139-148 |
| H6 | 182-184 | 179-181 |
| H7 | 234-247 | 231-244 |
| H8 | 260-265 | 257-262 |
| H9 | 280-285 | 277-282 |
| H10 | 316-321 | 313-318 |

| Helix(H) of Chain B | Residue No. pdb | Residue No. after MD |
|---|---|---|
| H13 | 115-120 | 449-454 |
| H14 | 132-140 | 466-474 |
| H15 | 142-150 | 476-484 |
| H16 | 182-184 | 511-513 |
| H17 | 235-247 | 564-576 |
| H18 | 260-265 | 589-584 |
| H19 | 280-285 | 609-614 |
| H20 | 316-321 | 645-650 |

**Fig. (3). Root mean square fluctuations (RMSF) during the MD simulations are shown for human renin.**

Among the secondary structure beta strand has higher fluctuation than alphahelix. There is ten Helix in both chains (A&B) within the protein in which Helix, H2 in chain B has highest fluctuations and Helix H10 in chain A has lowest fluctuations. Heilix H2, H4, H6 and H8, in chain A and H13, H14, H16, H17and H18 in chain B shows considerable fluctuations.

To understand the fluctuation of binding site residues of protein, RMSD of each residue was calculated. The average RMSD and their corresponding standard deviation (SD) for the residues are represented in a histogram (Figure 4).

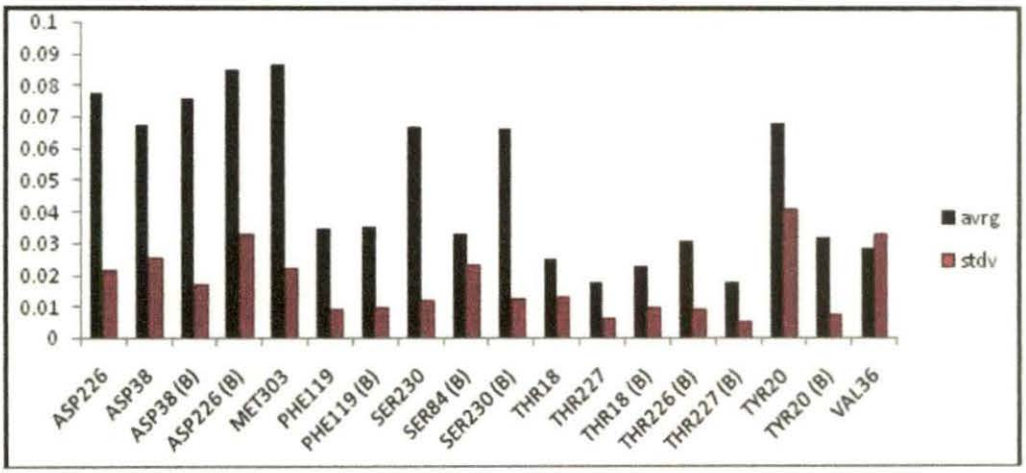



**Fig. (4). Average RMSD and their corresponding standard deviation (SD amino acid residues**

From the histogram it is examined that Thr227 shows the lowest fluctuation. Also Thr 227 (B), Thr18 and Val36 have low fluctuation. It is observed that Ser230 has low fluctuation than Met303.

To get insight about the binding mechanism, the residues of binding site were critically analyzed. The residues THR18, VAL36, THR227, SER230 and ASP38 (B) are very important for binding processes. From the RMSD values during the simulation time it is evident that the residues Thr227, Thr18 and Ser230 have very low fluctuation, as shown in Figure 5.

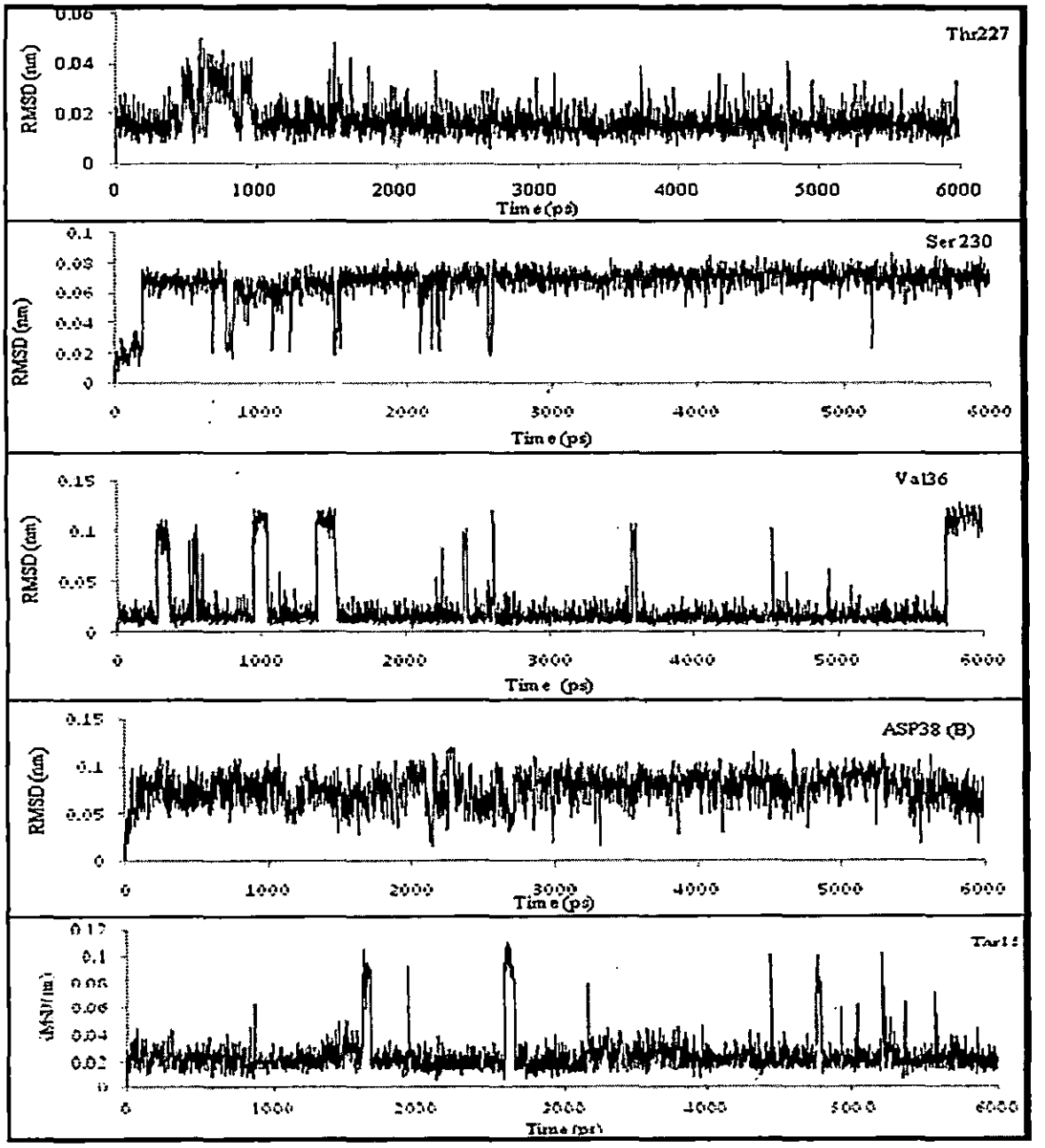



**Fig. (5). RMSD: Root mean square deviations (RMSD) of the Thr227, Ser 230, Val36, Asp38 (B), Thr18 as a function of time with respect to starting structure during the MD simulations.**

The residues Val36, Asp38 (B) have high fluctuation, and the fluctuation was moderate for the residues during the whole simulation time.

Now we have performed the docking simulation using those binding residues whose RMSD values have been already calculated. The residues having different fluctuation are

taken as flexible residue to see the effect of fluctuation in docking study. Each one of these residue selected as a flexible part to evaluate the flexible docking. Therefore we get 18 dock results and their corresponding binding energies are listed in Table 1.

**Table 1.The free energies of binding of 72X with different residues selected as a flexible residue calculated by AutoDock 4.2**

| Docking No | Residue | Binding Energy (kcal/mol ) |
|---|---|---|
| 12 | THR227 | -9.66 |
| 8 | SER230 | -9.53 |
| 18 | VAL36 | -9.34 |
| 3 | ASP38 (B) | -9.05 |
| 11 | THR18 | -8.96 |
| 5 | MET303 | -8.82 |
| 16 | TYR20 | -8.8 |
| 9 | SER84 (B) | -8.71 |
| 6 | PHE119 | -8.48 |
| 1 | ASP226 | -8.07 |
| 2 | ASP38 | -8.07 |
| 15 | THR227 (B) | -7.87 |
| 10 | SER230 (B) | -7.8 |
| 13 | THR18 (B) | -7.39 |
| 7 | PHE119 (B) | -7.07 |
| 4 | ASP226 (B) | -6.59 |
| 14 | THR85 (B) | -6.43 |
| 17 | TYR20 (B) | -5.49 |

From the table1 it is clearly shown that the docking with chain A residues as a flexible part have low binding energy. Low binding energy (-ve sign) indicates favorable binding (only a negative $\Delta G$ value is energetically favorable). Lowest binding energy value is obtained for Thr227 and. But docking with the chain B residues has moderately high

binding energy except Asp 38 (B) and Ser 84 (B). It is revealed from the RMSD binding energy data that low fluctuated residues have low binding energy. From table1 we choose five residues such as Thr227, Ser227, Val36, Asp38 (B) and Thr18. Each of them is favorably bound with inhibitor and RMSD is found lowest for Thr224 among the five residues.

Inhibitor 72X was successfully docked into the active site of human renin and calculated free energy of binding is shown in table1. From the table1 it is clear that the lowest binding energy value is obtained for Thr227 and docking with chain A residues as flexible part have low docking energy than the chain B residues except Asp 38 (B) and Ser 84 (B).

Docking structures are shown in Figure 6. The hydroxyl oxygen atom of the renin inhibitor was placed to make the hydrogen bond to an amide group of Ser230 (Figure 6A). The same hydroxyl oxygen atom makes a hydrogen bond with a side chain hydroxyl group of Ser230. The amino hydrogen atom of P1 moiety forms a hydrogen bond with the main chain carbonyl oxygen atom of Asp38. The hydroxyl hydrogen atom of the inhibitor is hydrogen bonded to an oxygen atom of the amide group of Gly228. The oxygen atom of P3' moiety forms a hydrogen bond with an amide group of Tyr20. A care full inspection of the binding pocket indicated that 72X adopt a position surrounded by hydrophobic groups Met303, Leu252, Thr85, Thr227, Phe124, Tyr231, Thr18, and Tyr20.

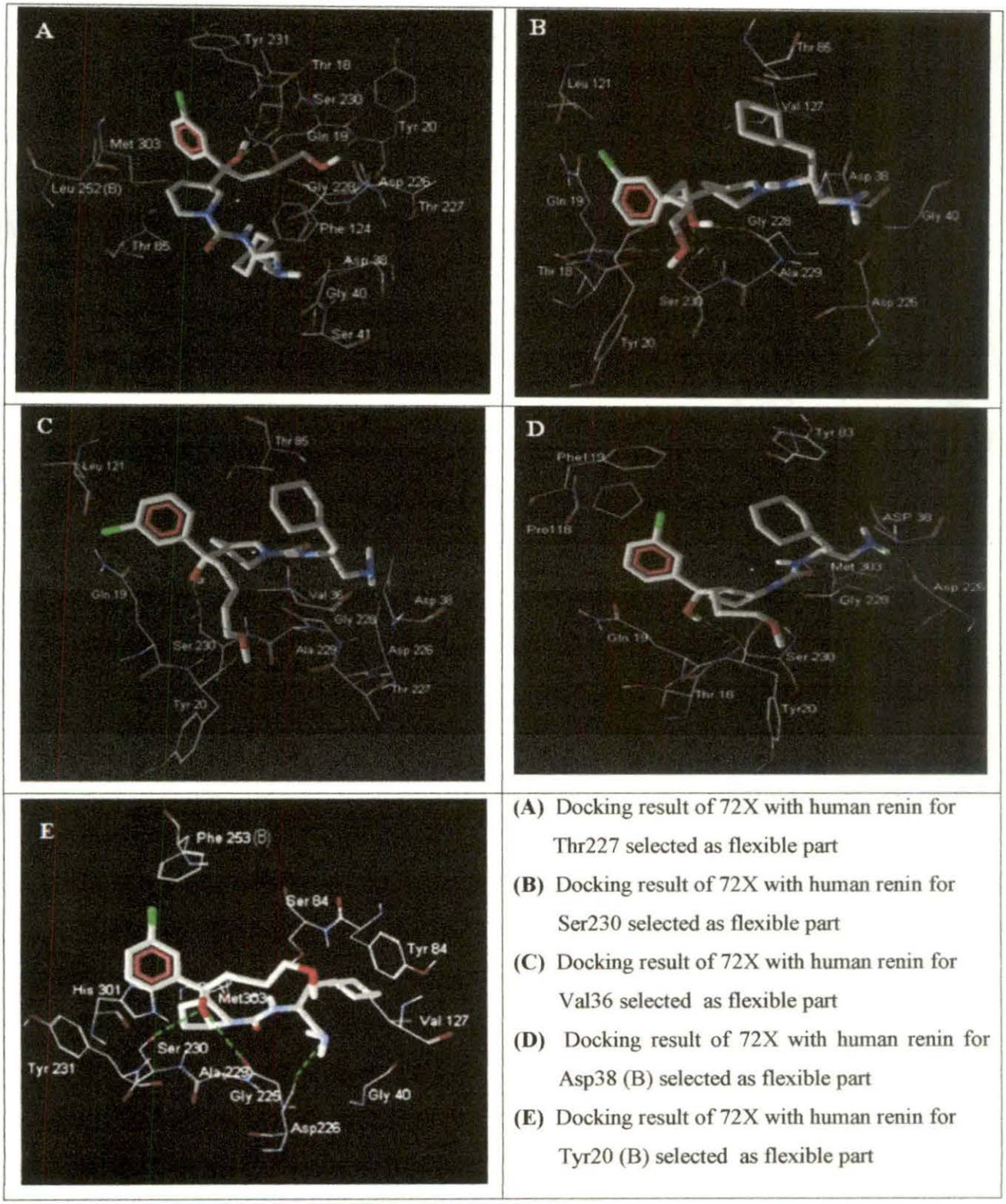
Figure 6B shows the four hydrogen bond interaction observed between inhibitor and human renin. The hydroxyl oxygen atom of inhibitor makes a hydrogen bond with the side chain hydroxyl group of Ser230. The same oxygen atom forms two hydrogen bonds with amide hydrogen of Ser230 and the carbonyl oxygen of Gly228. The fourth hydrogen bonding interaction observed between the oxygen atom of P3'moiety and amide

group of Tyr20. Inhibitor occupied a position in a hydrophobic cage surrounded by Leu121, Val127, Thr85, Thr18, Tyr20 and Ala229.

In Figure 6c, we examined three hydrogen bonds between 72X and Asp226, Gly228, Ser230. The amino hydrogen atom of P1 moiety makes a hydrogen bond with a side chain oxygen atom of Asp226. The hydrogen atom of the amide group of inhibitor forms a hydrogen bond with the amide oxygen of Gly228. Third hydrogen bonding involves between hydroxyl group of an inhibitor and side chain oxygen atom of Ser230. 72X make a position surrounded by hydrophobic residues Tyr20, Val36, Thr85, Leu221, Thr227 and Ala229.

Figure 6D shows three hydrogen bonds are formed between 72X and Asp226, Gly228, Ser230. It is observed that the hydrogen bonding interactions similar to the previous docking experiment (Figure 6C) 72X was surrounded by hydrophobic residues, such as Tyr20, Thr18, Met303, Tyr83 and Pro115.

The entire Figure 6A-Figure 6D imply that the P1 and P3' moieties hold large number of residues and their binding energies are low i.e. favorable binding. But the docking result in Figure 6E shows P1 moiety to have two nearest amino acid residues and P3' moiety have only one. Therefore docking for residue Tyr20 (B) as a flexible part in chain B binding energy is high. The hydroxyl hydrogen atom of inhibitor forms a hydrogen bond with a side chain oxygen atom of carboxylic acid group of Ser230. Second hydrogen bond forms between hydroxyl group of 72X and main chain carbonyl oxygen atom of Gly228. Another hydrogen bond forms between methyl amino group and main chain carbonyl group of Asp226. 72X surrounded by hydrophobic residues Phe 20 (B), Met303, Tyr83, Val 127, Ala226, Ala229 and Tyr23.




**(A)** Docking result of 72X with human renin for Thr227 selected as flexible part

**(B)** Docking result of 72X with human renin for Ser230 selected as flexible part

**(C)** Docking result of 72X with human renin for Val36 selected as flexible part

**(D)** Docking result of 72X with human renin for Asp38 (B) selected as flexible part

**(E)** Docking result of 72X with human renin for Tyr20 (B) selected as flexible part

**Fig. (6).** Docking results of 72X with a human renin for the Thr227, Ser230, Val (36), Asp38 (B), Tyr20 (B) selected as flexible part. The inhibitor 72x is presented by stick model and the residues in the active site of human renin are presented by wire frame.

Table 2 shows the free energies of binding of 72X with different residues selected as a flexible residue and their binding site residues

**Table 2.The free energies of binding of 72X with different residues selected as flexible residue calculated by AutoDock 4.2**

| Residue | Binding energy (kcal/mol) | Binding site residues |
|---|---|---|
| Thr227 | -9.66 | Asp226, Asp38, Gln19, Gly228, Gly40, Leu252(B), Met303, Phe124, Ser230, Ser41, Thr18, Thr227, Thr85, Tyr 231, Tyr20 |
| Ser230 | -9.53 | Ala229, Asp226, Asp38, Gln19, Gly228, Gly40, Leu121, Ser230, Thr18, Thr85, Tyr20, Val127, |
| Val36 | -9.34 | Ala229, Asp226, Asp38, Gly19, Gly228, Leu121, Ser230, Thr227, Thr85, Tyr20 Val36, |
| Asp38 (B) | -9.05 | Asp226, Asp38, Gln19, Gly228, Met303, Phe119, Pro118, Ser230, Thr18, Tyr20, Tyr83 |
| Tyr20 (B) | -5.49 | Ala229, Asp226, Gly228, Gly40, His301, Met303, Phe253 (B), Ser230, Ser84, Tyr231, Tyr83, Val127 |

Renin inhibitor 72X possesses a hydrophobic environment surrounded with P1, P2, P1', P2' and P3' residues. LogP value (4.59) suggests hydrophobicity of the inhibitor is quite high [29]. Thus the hydrophobic interaction is important for stabilizing the complex. Interactions of inhibitor moieties with the active site residues of human renin at 3.5 Å are shown in table 3.

**Table 3. Interactions of inhibitor moieties with the active site residues of human renin at 3.5 Å.**

| Residue | P1 | P2 | P1' | P2' | P3' |
|---|---|---|---|---|---|
| Thr 227 | Asp38, Asp226 Gly 40, Ser 41 | Phe 124 | Thr85, Leu252(B), Met303 | Tyr231 | Gln19, Gly228, Thr 18, Tyr20, Thr227 |
| Ser 230 | Asp38, Asp226, Ala 229, Gln40 | Val127, | Thr85 | Leu121 Gln19 | Thr18, Tyr20, Gly228 |
| Val36 | Asp38, Asp226, | Thr85 | | Leu121, Gln19 | Tyr 20, Thr227, Ala229, Gly228 |
| Asp38 (B) | Asp226 | Tyr83, Asp38 | | Pro118, Phe119 | Gly228, Ser230, Tyr20, Gln19, Thr18 |
| Tyr20 (B) | Asp226, Gly40 | Ser84, Tyr83 | His301, Ala229, Tyr231, Ser230 | Phe253(B) | Val127, |

Figure 6A shows highly hydrophobic P2 moiety of 72X is closest to the hydrophobic phenyl ring of Phe124 at about 3.5 Å. Piperidine ring of inhibitor accommodates hydrophobic pocket formed by the residues Met303, Leu252 (B), and Thr85. Also P3' moiety forms hydrophobic contacts with the Thr 18, Tyr 20, Thr227, Gln19 and Gly228. Within 3.5 Å, P1 moiety accommodates feebly hydrophobic region. P2' moiety forms hydrophobic contacts with phenyl ring of the Tyr 231. From the Figure 6B it is seen that P2-cyclohexyl residue closer to hydrophobic residue Val127. P1' residue close to Thr85. P3' moiety accommodates the hydrophobic residues Thr18, Tyr20 and Gly228. P2'

moiety is closer to Leu121, Gln19. Figure 6C shows, Thr85 and cyclohexane ring of the inhibitor close to each other. P2' moiety of the inhibitor exists in close proximity to Leu121 and Gln19. Strong hydrophobic interactions observed between P3' moiety and each of the Tyr20, Thr224, Ala229, Gly228. In Figure 6D, P3'-metoxy butane moiety placed in the hydrophobic pocket surrounded by hydrophobic residues, such asTyr20, Thr18, Gly228, Ser230 and Gln19. Figure 6E shows Val 127 closer to methoxy butane moiety. Fewer number of hydrophobic amino acid residues are surrounded the hydrophobic part of the inhibitor. All the docking result shows two aspartic acid residues Asp38 and Asp226 placed at the P1 moeity. The chlorobenzene, cyclohexyl and methoxy butyl moieties of 72X fragments in a hydrophobic pocket composed amino acid residues Phe119, Phe124, Val127 and Tyr83. Each and every one docking result shows that inhibitor in hydrophobic cage surrounded by largely hydrophobic residues of renin. From Figure 6A-Figure 6D it is seen that hydroxyl group, P3', P1 moiety lie in approximately same side in 72X but Figure 6E shows P3' moiety is distorted. Maximum numbers of residues interact with 72X at that region. Opposite side has less number of residues that interact with the inhibitor. We have obtained different lowest free energy conformers from the docking experiment and it is observed that they have large conformational change. These conformational changes may differentiate the binding affinity of the amino acid residues around the inhibitor. Figure 6A-Figure 6B shows maximum number residues in chain A are involved in the binding process whereas only two residues Leu252 (B) and Phe253 (B) of chain B are involved in the binding process.

Docked conformer with lowest binding energy value is superimposed with the ligand from crystallographic structure of the complex, using the Accelrys Discovery Studio 2.5 and RMSD was obtained 2.12 Å. It is presented in Figure 7.

**Fig. (7). Superimposition of the ligand from crystallographic structure of the complex with docked conformer of 72X. Ligand from crystallographic structure is shown in green and docked conformer in red.**

**4.4: References**


1. Danser AHJ, Denium J (2005) Renin, pro-renin and the putative (Pro) renin receptor. *Hypertension*, **46**: 1069–1076.

2. Laragh JH (1960)The role of aldosterone in man: Evidence for regulation of electrolyte balance and arterial pressure by renaladrenal system which may be involved in malignant hypertension. *J.A.M.A.*, **174**: 293-295.

3. Ames RP, Borkowski AJ, Sicinski AM, Laragh JH (1965) Prolonged infusions of angiotensin II and norepinephrine and blood pressure, electrolyte balance,

aldosterone and cortisol secretion in normal man and in cirrhosis with asrites. *J. Clin. Invest.*,**44**: 1171-118.

4. Laragh JH, Baer L, Brunner HR, Buhler FR, Vaughan JE (1972) Renin, angiotensin and aldosterone system in pathogenesis and management of hypertensive vascular disease. *Am. J. Med.*, **52**: 633-652.

5. Chen KS, Tang J (1972) Amino acid sequence around the epoxide-reactive residues in pepsin. *J. Biol. Chem.*, **247**: 2566-2574.

6. Hartsuck JA, Tang J (1972)The carboxylate ion in the active centre of pepsin. *J. Biol. Chem.*, **247**: 2575-2580.

7. Reudelhuber, T.L.; Ramla, D.; Chiu, L.; Mercure, C.; Seidah, N.G. Proteolytic processing of human pro-renin in renal and non-renal tissues. *Kidney Int.*, **1994**, 46, 1522–1524.

8. Kaschina E, UngerT (2003) Angiotensin AT1/AT2 receptors: regulation, signalling and function. *Blood Press.*,**12**, 70–88.

9. Kim, S.; Iwao, H. Molecular and cellular mechanisms of angiotensin II-mediated cardiovascular and renal diseases. *Pharmacological reviews*, **2000**, 52 (1), 11-34.

10. Campbell DJ, Aggarwal A, Esler M, Kaye D (2001) Beta-blockers, angiotensin II, and ACE inhibitors in patients with heart failure. *Lancet.* **358:** 1609-10.

11. Stanton A (2003) Therapeutic potential of renin inhibitors in the management of cardiovascular disorders. *American Journal of cardiovascular drugs*, **3**: 389-394.

12. Norman KH (*2005*) Therapeutic opportunities for renin inhibitors, *J.R.A.A.S.*, **1**, 107–109.

13. Rahuel J, Rasetti V, , Maibaum J, Rüeger H, Göschke R, et al. (2000) MG. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chemistry & biology*, **7**: 493-504.

14. Fisher ND, Hollenberg NK (2001) Is there a future for renin inhibitors?.*Expert. Opin. Investig. Drugs*, **10**: 417–426.

15. Wood JM, Maibaum J, Rahuel J, Grutter MG, Cohen NC, et al. (2003) Structure-based design of aliskiren, a novel orally effective renin inhibitor. *Biochem. Biophys. Res. Commun.*, **308**: 698-705.

16. Rongen GA, Lenders JW, Smits P, Thien T (1995) Clinical pharmacokinetics and efficacy of renin inhibitors. *Clinical Pharmacokinetics*,**29**: 6–14.

17. Cohen NC (2007) Structure based Drug design and the discovery of aliskarin (Tekturna): Perseverance and creativity to overcome a R&D pipeline challenged. *Chemical Biology & Drug Design*, **70**: 557-565.

18. Gradman AH, Kad R (2008) Renin Inhibition in Hypertension. *J. Am. Coll. Cardiol.*, **51**: 519-528.

19. SieleckiAR,Hayakawa K, FujinagaM, MurphyME, Fraser M, et al. (1989)Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5Å resolution. *Science*, 243: 1346-1351.

20. Ticea CM, Xua Z, Yuana J, Simpsona RD, Cacatiana ST,et al. (2009) Design and optimization of renin inhibitors: Orally bioavailable alkyl amines. *Bioorganic & Medicinal Chemistry Letters*. **19**: 3541–3545.

21. Peitsch MC (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans.*,**24**: 274–279.

22. http//www.seanet.com/ mthompson/Arguslab.

23. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. (2009) Goodsell, D.S.; Olson, A.J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, **162**: 785–791.

24. Lindahl E, Hess B, Spoel D (2001) GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.*, 7: 306-307.

25. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.*, **18**: 1463-1472.

26. Berendsen HJC, Postma JPM, van Gunsteren WF,Dinola A, Haak JR (1984) Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **81**: 3684-3690.

27. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.*, **19**: 1639-1662.

28. Molegro Molecular Viewer (MMV), **2008**. 1.0, Molegro, Denmark.

29. ChemAxon. MarvinSketch 4.1.8. **2007**. ChemAxonKft. Budapest, Hungary.

30. Discovery Studio 2.5 (2009) Accelrys Inc.: San Diego, CA, USA,

# Chapter V

# *Molecular dynamics simulation of chick Type IIa receptor protein tyrosine phosphatases sigma*

## 5.1: Introduction

Type IIa receptor protein tyrosine phosphatases (RPTPs), such as RPTPσ, LAR and RPTPδ, are cell surface receptors which play an important role in neuronal development, function and repair [1, 3]. Type IIa RPTPs contain variable number of extracellular immunoglobulin (Ig) domains and two to nine fibronectin type III (FNIII) domains and two cytoplasmic phosphatase domains. These extracellular domains are similar to cell adhesion molecules (CAMs), which imply that they play roles in cell–cell and cell–matrix interactions [4, 5]. In vertebrates and invertebrate, types IIa RPTPs are present in axons and growth cones, which regulating neuronal growth and guidance and participating in excitatory synapse formation and maintenance [6-11]. RNA interference-mediated knock-down indicate that they contribute to motor axon regulation in chick [12]. The role of leukocyte common antigen-related (LAR) subfamily members in axon guidance during mammalian development has not been reported. Lower level of LAR in the Central nervous systems (CNS) are observed during development, and deficiency of LAR causes minor defects in cholinergic innervations of the hippocampal dentate gyrus [13, 14]. RPTPσ and RPTPδ both develops mammalian nervous system [15]. Abnormal growth of the pituitary and neurological defects that include spastic movements and abnormal limb flexion observed in mice due to the deficiency of RPTPσ [16, 17]. Further, the absence of RPTPσ results in increased axon regeneration in both the peripheral and central nervous systems [18].

Extracellular regions, of Type IIa RPTP interact with Heparan sulfate proteoglycans (HSPGs) and chondroitin sulfate proteoglycans (CSPGs) [19-22]. These proteoglycans adjust neuronal growth, guidance and connectivity with CSPGs inhibiting and HSPGs promoting axon extension [23-25]. Increase in a cellular response of CSPGs in glial damage tissue after neural injury is an important factor limiting CNS axon and

regeneration [2]. CSPG reduces development of dorsal root ganglion neurons, whereas in RPTPσ neurons this inhibitory effect is decreased [22]. In developing chick retinal ganglion cell axons, RPTPσ promotes growth in response to basal lamina which contains HSPG ligands [20, 26]. CSPG-RPTPσ and HSPG-RPTPσ interactions are opposing effects on cell function, but how these opposing effects are mediated at the molecular level has been unknown.

Coles et al. show that Neurocan, a CSPG might reduce outgrowth of dorsal root ganglion neurons, whereas this inhibitory effect is decreased in RPTPσ neurons. Conversely, glypican-2, a HSPG, strongly promotes outgrowth of wild-type, but not RPTPσ neurons. They further show that the glycosaminoglycan (GAG) chains of neurocan and glypican-2 must be involved, and that their opposing effects are mediated through a common receptor, RPTPσ. Mutagenesis studies suggested a common GAG-binding site in the N-terminal Ig domain of RPTPσ. To examine the structural basis of proteoglycan recognition in RPTPσ and type IIa RPTPs, Coles et al. performed crystallographic studies [27]. The crystal structures of the two N-terminal Ig domains (Ig1-2) of different members of the RPTP family across different species formed a V-shaped arrangement of Ig1 and Ig2, which is stabilized by conserved interactions. Residues of RPTPσ previously shown to mediate GAG binding lie on loops between Ig1 β-strands, forming an extended positively charged surface. This region is highly conserved across family members and species, suggesting a common GAG-binding mode.

Coles et al. solved the Crystal structure of cheek RPTP σ with the 1.65 A° resolution. Cheek RPTP σ is a monomer, with 203 amino acids residues. In Ig1-Ig2 pro-rich loop the side chains have hydrophobic residues LEU124, LEU129, PRO130, and PHE133 and also pack closely with ILE42, VAL44, and ALA 212 in a hydrophobic interdomain region. Two Salt bridges were found between ARG91 – GLU205 (SB1) and GLU126 – ARG215

(SB2). These salt bridges are formed by hydrogen bond. Salt bridge SB1 contains two hydrogen bonds (HB1, HB2) between amide nitrogen of ARG91 and side chain oxygen of GLU205 while SB2 contain three hydrogen bonds (HB3, HB4, HB5) between amide nitrogen of ARG215 and side chain oxygen of GLU126. Hydrogen bonds formed between the hydroxyl groups of SER50 and TYR216 (HB6). Another hydrogen bond formed between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB7). Also the side chain amide of GLN41 and backbone carbonyls of PHE 171 forming hydrogen bond with the water molecules. These electrostatic interactions make the Ig1 and Ig2 in a rigid arrangement.

In this protein two salt bridges were found. A salt bridge is a non covalent interaction between the opposite charged residues of a protein [28, 29].

In this work, molecular dynamics simulations and principal component analysis of chick Type IIa receptor protein tyrosine phosphatases have been done which will help to understand its function and mode of action.

**5.2: Materials and Methods**

The simulation was performed using GROMACS [30]. The 1.65 A° resolution x-ray structure of chick Type IIa receptor protein tyrosine phosphates (Protein Data Bank code 2YD4) was used as a starting structure. A single monomer was solvated with SPC water molecules in a cubic box, having an edge length of 3.5 Å. The LINCS algorithm was applied to constrain all bond lengths [31]. The molecular dynamics simulation was performed at a constant temperature (300K) and using the Berendsen coupling method for coupling each component separately to a temperature bath [32]. A cutoff of 0.9 nm was used for Lennard Jones interaction and the particle mesh Ewald method was employed to calculate longer-rang electrostatic contribution on a grid with spacing and a cutoff of 1.0

nm for Coulomb interaction. The time step was 2 fs, with coordinates stored after every 2 ps. MD simulation was performed for 12 ns. Before running simulation, an energy minimization was performed in steepest descent method (converged at 876 steps) [33, 34]; and this was followed by 4.0 ns of simulation imposing positional restraints on the non-H atoms. The positional restraints were then released and 12 ns production run were obtained and analyzed. Analysis programs from GROMACS were used and PCA was performed with the MD trajectory.

### 5.3: Results and Discussion

During the MD simulation the overall structural stability of the protein is explored by several parameters like RMSD, radius of gyration (Rg), RMSF etc. These parameters give an idea of the folding and structural behavior of cheek RPTP σ.

The time evolution of RMSD is computed taking the constrained structure of the whole protein as initial structure and presented in Figure 1.

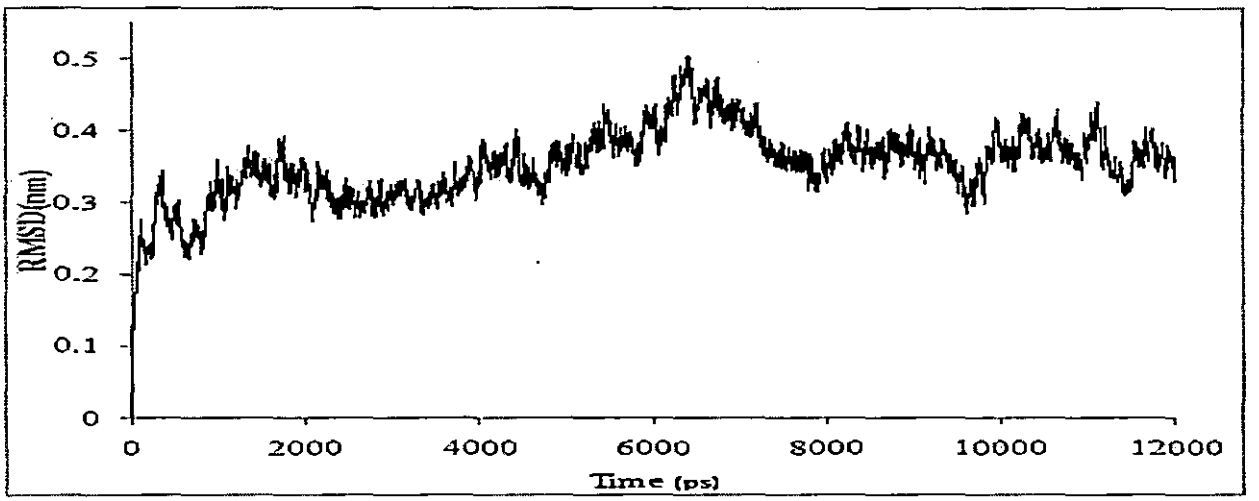



**Figure-1: RMSD: Root mean square deviations (RMSD) of the protein as a function of time with respect to starting structure during the MD simulations are shown for cheek RPTPσ.**

It is evident from figure 1 that RMSD slightly increases up to 1700ps, and then slightly decreases and again increases up to 6390ps and after that almost become stable. The time evolutions of RMSD of the whole protein of chick Type IIa receptor protein tyrosine phosphatases during the full simulation time shows an initial drift in RMSD which may be due to the difference of crystal structure with solution structure.

The variation of radius of gyration (Rg) as function of time is presented in Figure 2 and from this figure it is clear that the initial Rg value is 1.90673 and then Rg values decreases upto 6986ps with Rg value 1.7723, after that Rg slightly increases up to 7990ps (1.85288). Finally Rg values almost become stable.

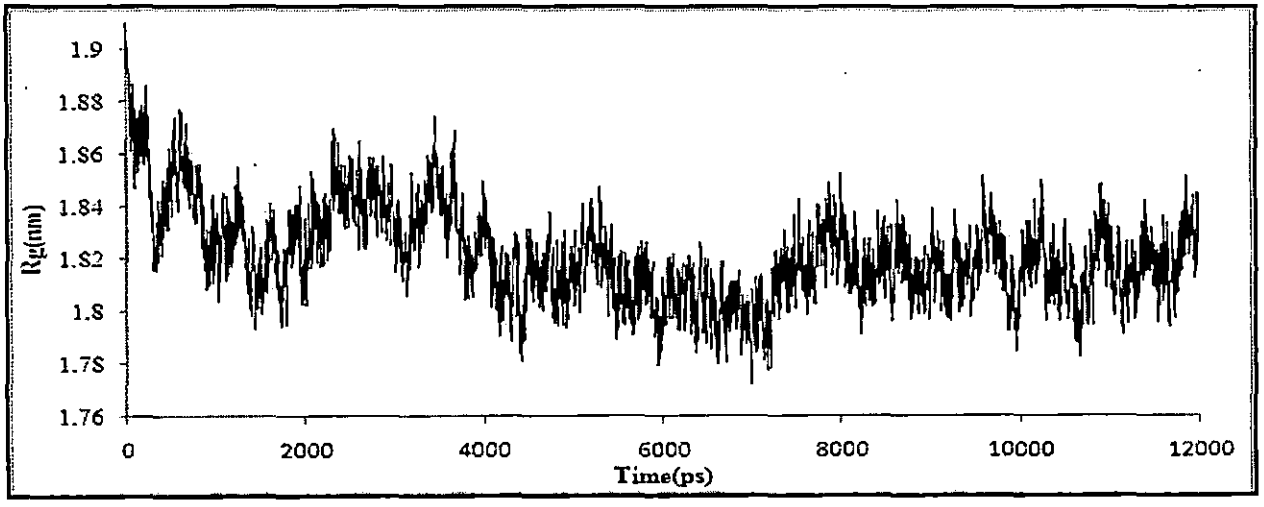



**Figure 2: Radius of gyration: Radius of gyration (Rg) as a function of time with respect to starting structure during the MD simulations is shown for cheek RPTPσ.**

The RMSD and Rg calculations of chick Type IIa receptor protein tyrosine phosphatases suggest that the protein is slightly flexible in nature.

The flexibility of different parts of the protein is revealed by the root mean square fluctuation (RMSF) of each residue from its time-averaged position (Figure 3).

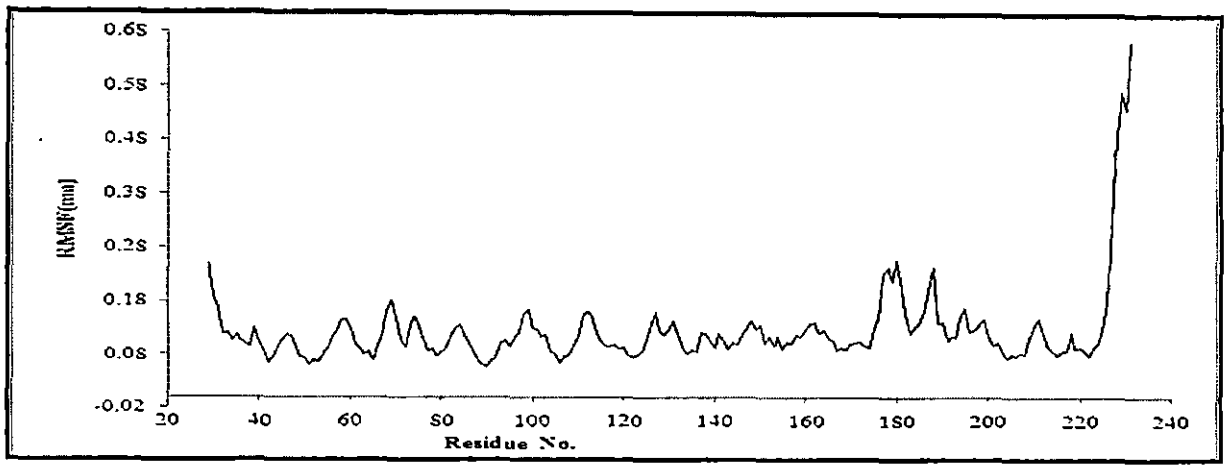



**Figure 3: Fluctuation: Root mean square fluctuations (RMSF) during the MD simulations are shown for cheek RPTPσ.**

From RMSF, it is evident that pronounced fluctuations are observed along some amino acid stretches (67-77, 95-105, 110-118,125-133 and 175-190), which indicate the flexibility of the protein in that region (Figure 3). But last portion of the protein (225-231), RMSF value increases rapidly and reached to highest value (0.6662) which indicates N terminal of the protein is free to move.

Principal component analysis was performed to investigate the nature of the fluctuations. PCA of the cheek RPTPσ simulation reveals that the first 10 eigenvectors account for 74.905% of the global motion and that the first eigenvector corresponds to 25.49% of the total motion, the second to 15.85% and the third to a further 9.37%n (Figure 4).

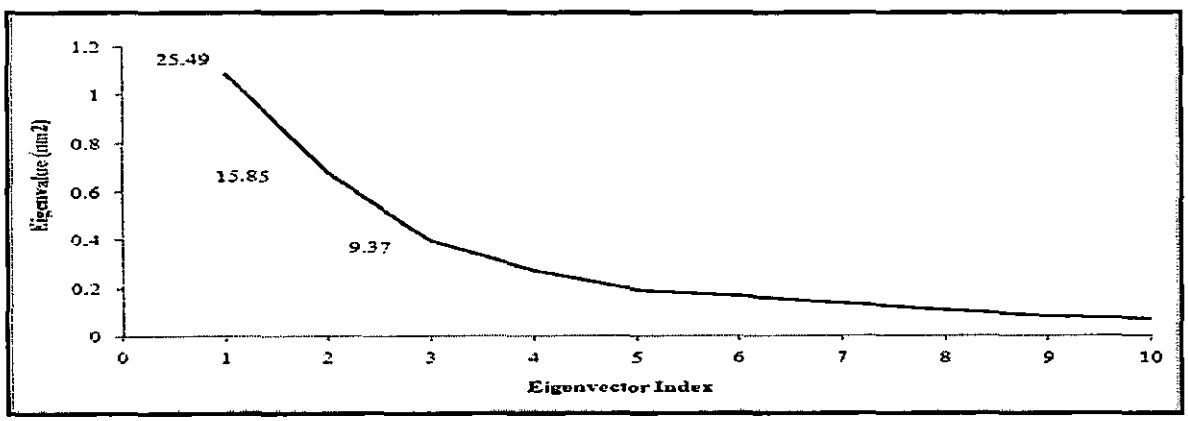



**Figure 4: Principal components analysis: Plot of Eigenvalues with Eigenvector Indices**

From Principal components analysis of the cheek RPTP-σ, it is evident that motion is very much distributed through all the principal components.

Further we investigate the nature of the fluctuations; principal component analyses (PCA) were carried out. The first four eigenvectors with largest eigenvalues were selected as the four principal components. Time evolution of principal component 1 (PC1), principal component 2 (PC2), principal component 3 (PC3) and principal component 4 (PC4) in water is represented in Figure 5.

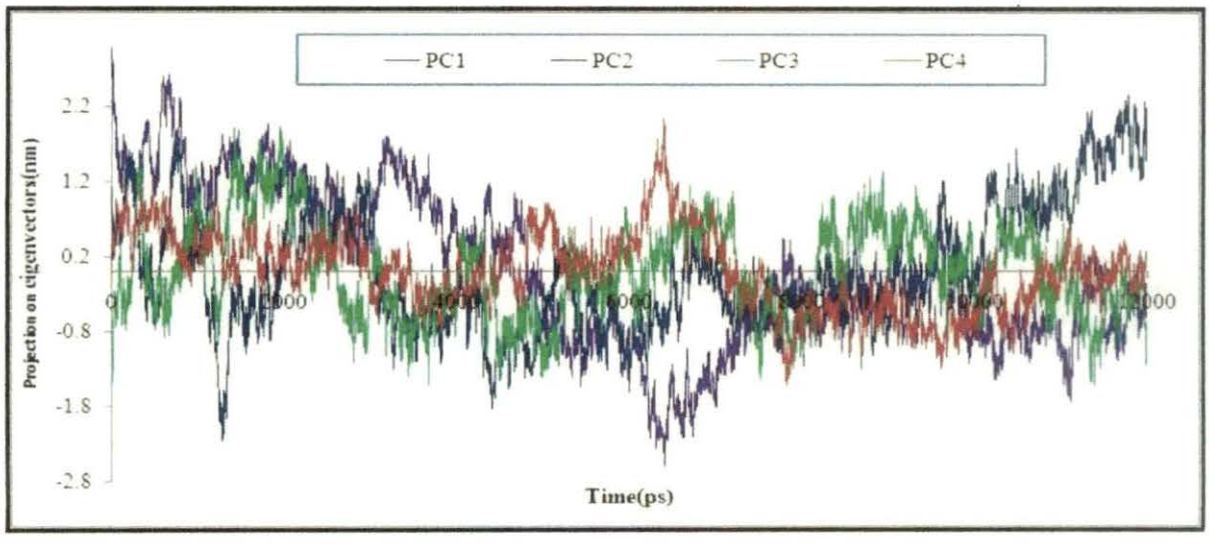



**Figure 5: Time evolution of principle components: Variation of four principle components with simulation time for cheek RPTPσ**

It is seen from Figure 5 that time evolution of principal component 1 (PC1), principal component 2 (PC2), fluctuates remarkably in comparison to principal component 3 (PC3) and principal component 4 (PC4).

The RMSF of Cα atoms calculated after projecting trajectories along their respective four principal components is presented in Figure 6.

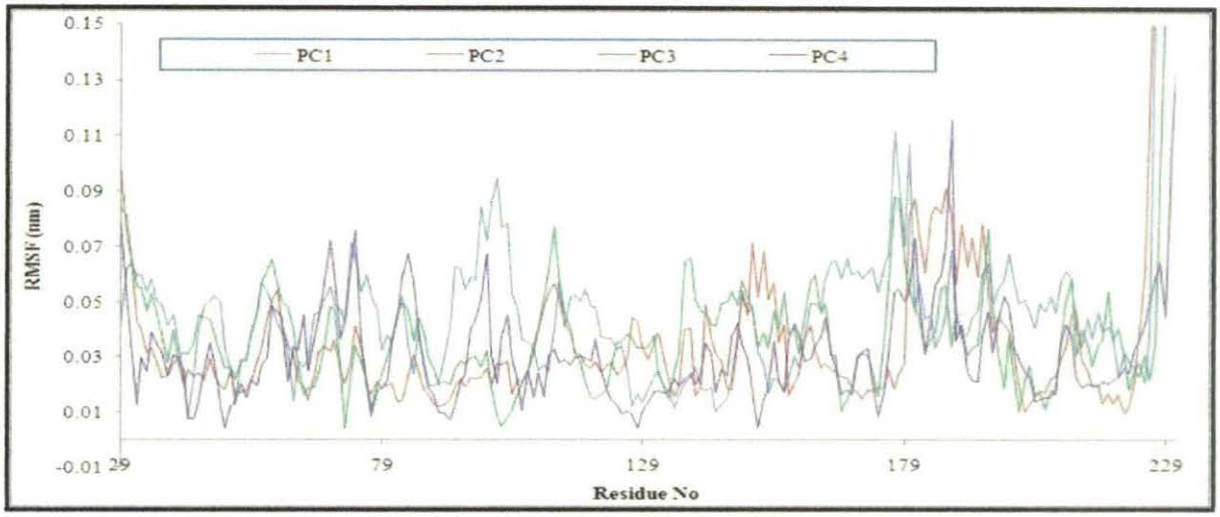



**Figure 6: Projections on principle components: The RMSF of Ca atoms calculated after projecting trajectories along their respective PC1 directions**

It is evident from the RMSF of Cα atoms calculated after projecting trajectories along their respective four principal components analysis for chick Type IIa receptor protein tyrosine phosphatases fluctuation is highest in the projection on PC1 which indicates that PC1will provide more information regarding the collective motion of the protein.

The probability of sampling the phase space determined by first two principal modes (PC1, PC2) during the simulations presented in Figure 7.

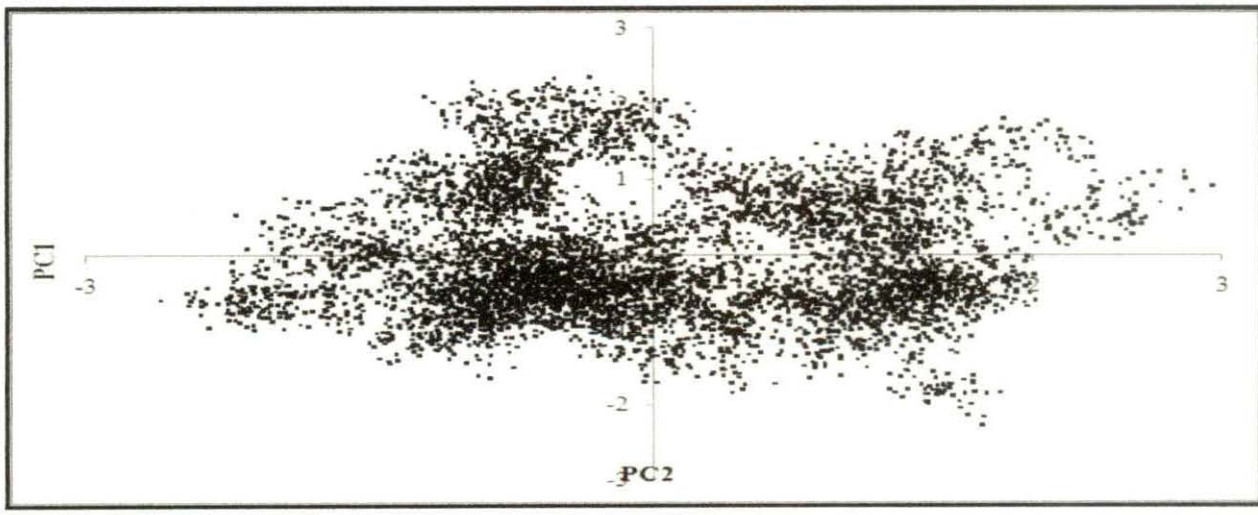



**Figure 7: Conformational Sampling: The probability of sampling the phase space determined by principal modes 1 and 2 during the simulations.**

It is clear from the probability of sampling the phase space determined by first two principal modes during the simulations the projection of the dynamics trajectory onto the first two PC that the protein is widely spread around the origin and high conformational space around the origin (Figure 7).

It is found that the number of hydrogen bonds ranged from 104 to 155 (Figure 8).

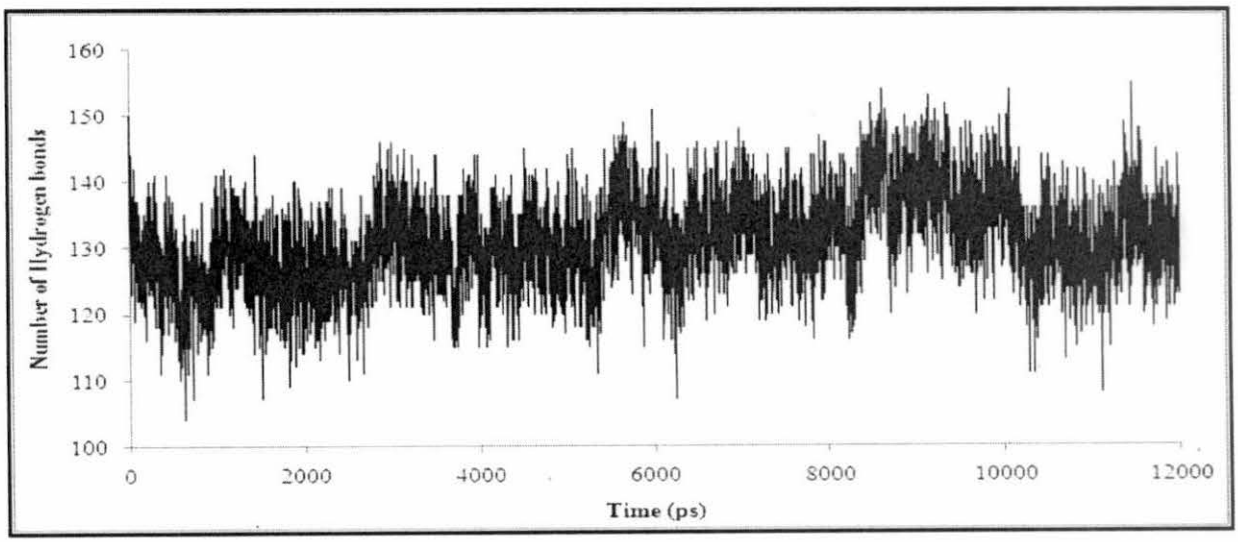



**Figure 8: Hydrogen bonds: Number of hydrogen bonds during the whole simulation time**

The number of hydrogen bonds ranges from 104 to 155 (Figure 8) indicate that during the simulation, several hydrogen bonds broke and formed.

Two salt bridges were found between ARG91 & GLU205 (SB1) and GLU126 & ARG215 (SB2) in RPTP σ. Among the four residues GLU126 has highest RMSD value where as ARG91 has lowest RMSD value. The residue GLU126 is more flexible than the other residue in salt bridges. Again RMSD analysis performed both salt bridges (ARG91 & GLU205 and GLU126 & ARG215). Mean RMSD values of SB2 (GLU126 & ARG215) is 0.180975, which is almost double than SB1 (ARG91 & GLU205). SB2 has high fluctuation comparison to SB1.

Hydrophobic residues in Ig1-Ig2 pro-rich loop contains seven residues (LEU124, PRO130, PHE133, VAL44, ALA212 ) of which VAL44 has highest RMSD value and ALA212 possess lowest RMSD value indicating conserved nature.

Lys loop in RPTP σ contain eleven residues (67-77). All the residues have high RMSD value, except LYS67, GLY69, and SER74 which indicates there mobile nature. But residues (ARG96, ARG99) in Arg loop have low RMSD value indicating low mobility.

A series of hydrogen bonds were observed in the salt bridge and in Ig1-Ig2 pro-rich loop. We have monitored the distance of several hydrogen bonds during the simulation time (Figure 9).

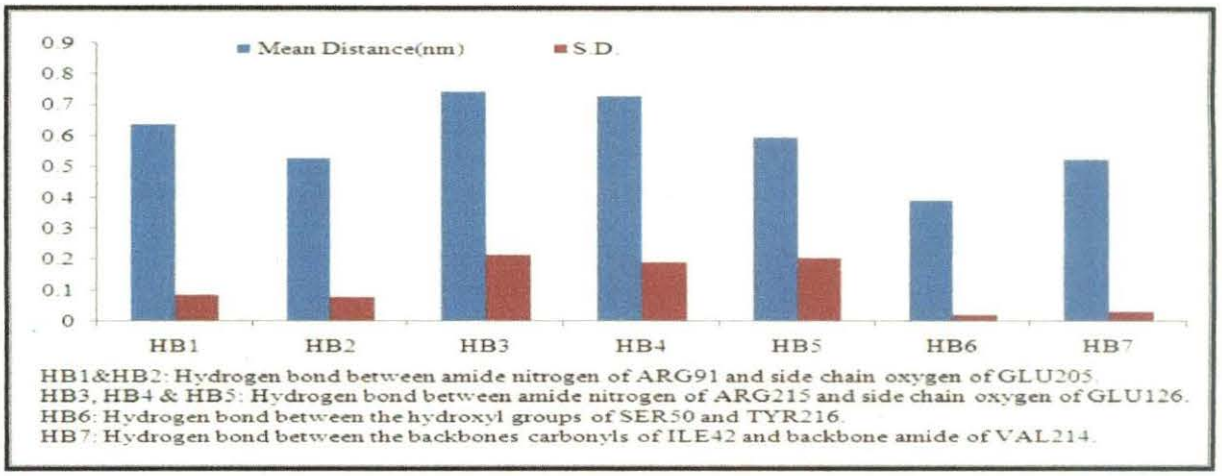



**Figure 9: Hydrogen bond of cheek RPTPσ: Hydrogen bonds within salt bridge and Ig1-Ig2 pro-rich loop during the simulation time.**

It was found that hydrogen bond between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB6) remain intact during the whole simulation time. Hydrogen bonds (HB2) slightly cross the limit of Hydrogen bonds distance. But other hydrogen bonds (HB1, HB3, HB4, HB5, and HB7) were broken during the simulation (Figure 9).

We have taken some selected snapshots from the dynamics trajectory considering time evolution of RMSD as a guideline. The snapshots were taken when the RMSD from initial structure was high, and they are represented in Figure 10.

**Figure 10: Snapshots: Snapshots of the cheek RPTPσ at different time.**

From the snapshots, it is also clear that there was slight change in the protein conformation.

We further examine the probability of accessing regions of the phase space determined PC1, PC2 & PC3 (Figure 11)

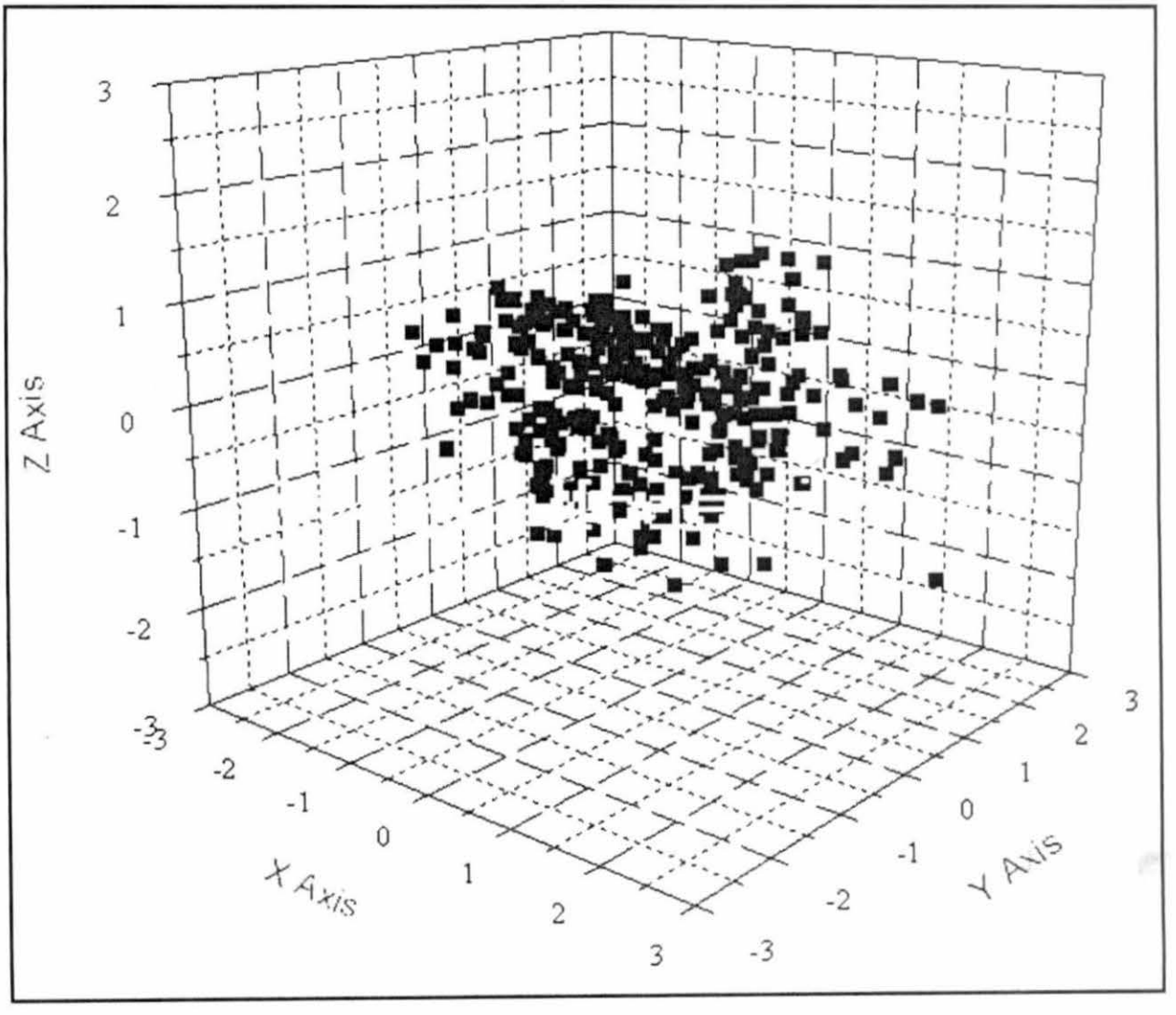



**Figure11: Space determined PC1, PC2 & PC3 representation: Probability of accessing regions of the phase space determined PC1, PC2 & PC3 representation for cheek RPTPσ**

From Figure 11, it is clear that protein show less arrangement in the XY plane and the protein highly arrange in the XZ plane.

## 5.4: References


1. Johnson KG and Vactor DV (2003) Receptor protein tyrosine phosphatases in nervous system development. *Physiol. Rev.*, **83**: 1-24.

2. Duan Y and GigerRJ (2010) A new role for RPTPsigma in spinal cord injury: signaling chondroitin sulfate proteoglycan inhibition. *Sci. Signal.*3: pe6.

3. Tonks N K (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* 7(11): 833-846.

4. Brady-Kalnay SM and Tonks NK (1995) Protein tyrosine phosphatases as adhesion receptors. *CurrOpin Cell Biol.***7**: 650–657.

5. Beltran PJ and Bixby JL (2003) Receptor protein tyrosine phosphatases as mediators of cellular adhesion. *Front Biosci.*,**8**: 87– 99.

6. Tian SS, Tsoulfas P, et al. (1991) Three receptor-linked protein-tyrosinephosphatases are selectively expressed on central nervous system axons in the Drosophila embryo. *Cell.* **67**: 675– 680.

7. Desai CJ, Popova E, et al. (1994) A Drosophila receptor tyrosine phosphatase expressed in the embryonic CNS and larval optic lobes is a member ofthe set of proteins bearing the "HRP" carbohydrate epitope. *J Neurosci.*,**14**: 7272–7283.

8. Stoker AW, Gehrig B, et al. (1995) Axonal localisation of the CAM-like tyrosine phosphatase CRYP alpha: a signalling molecule of embryonic growth cones. *Development.* **121**: 1833–1844.

9. Sommer L, Rao M, et al. (1997) RPTP delta and the novel proteintyrosine phosphatase RPTP psi are expressed in restricted regions of thedeveloping central nervous system. *DevDyn.*, **208**: 48–61.

10. Schaapveld RQ, Schepens JT, et al. (1998) Developmental expression of the cell adhesion molecule-like protein tyrosine phosphatases LAR, RPTPdelta and RPTPsigma in the mouse. *Mech Dev.*, **77**: 59–62.

11. Thompson KM, Uetani N, et al. (2003) Receptor protein tyrosine phosphatase sigma inhibits axonal regenerationand the rate of axon extension. *Mol Cell Neurosci.*,**23**: 681– 692.

12. Stepanek L, Stoker AW, et al, (2005) Receptor tyrosine phosphatases guide vertebrate motor axons during development. *J Neurosci.* **25**: 3813–3823.

13. Yeo TT, Yang T, et al. (1997) Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. *J Neurosci Res.*, **47**: 348 –360.

14. Van Lieshout EM, Van der Heijden, et al. (2001) A decrease in size and number of basal forebrain cholinergic neurons is paralleled by diminished hippocampal cholinergic innervation in mice lacking leukocyte common antigen-related protein tyrosine phosphatase activity. *Neuroscience.* **102**: 833– 841.

15. Yan H, Grossman A, et al. (1993) A novel receptor tyrosine phosphatase-sigma that is highly expressed in the nervous system. *J Biol Chem.*, **268**: 24880 –24886

16. Elchebly M, Wagner J, et al. (1999) Neuroendocrine dysplasia in mice lacking protein tyrosine phosphatase sigma. *Nat Genet.*,**21**:330 –333.

17. Wallace MJ, Batt J, et al. (1999) Neuronal defects and posterior pituitary hypoplasia in mice lacking the receptor tyrosine phosphatase PTPsigma. *Nat Genet.*, **21**:334 –338.

18. Saphieha PS, Duplan L, et al. (2005) Receptor protein tyrosine phosphatase sigma inhibits axon regrowth in the adult injured CNS. *Mol Cell Neurosci.*,**28**: 625– 635.

19. Aricescu AR, McKinnell IW, et al. (2002) Heparan sulfate proteoglycans are ligands for receptor protein tyrosine phosphatase sigma.*Mol. Cell. Biol.*,**22**: 1881-1892.

20. Fox AN and Zinn K (2005) Theheparan sulfate proteoglycan syndecan is an in vivo ligand for the Drosophila LAR receptor tyrosine phosphatase. *Curr. Biol.*,**15**(19): 1701-1711.

21. Johnson KG, Tenney AP, et al. (2006) The HSPGs Syndecan and Dallylike bind thereceptor phosphatase LAR and exert distinct effects on synaptic development.*Neuron.* **49:** 517–531.

22. Shen Y et al., (2009)PTPσ Is a Receptor for Chondroitin Sulfate Proteoglycan, an Inhibitor of Neural Regeneration. *Science.***326**: 592-596.

23. Bandtlow CE and Zimmermann DR (2000) Proteoglycans in the developing brain: new conceptual insights for old proteins. *Physiol. Rev.*,**80**: 1267-1290.

24. Silver J, Miller JH, (2004) Regeneration beyond the glial scar. Nat. Rev. *Neurosci.*, 5: 146-152.

25. Matsumoto Y, Irie F, et al. (2007) Netrin-1/DCC Signaling in Commissural Axon Guidance Requires Cell-Autonomous Expression of Heparan Sulfate. *Neurosci.*,**27**: 4342.

26. Ledig M and Haj F (1999) The receptor tyrosine phosphatase CRYPalpha promotes intraretinal axon growth. *J. Cell Biol.* **147**: 375-388.

27. Coles CH, Shen Y, et al., (2011) Proteoglycan-Specific Molecular Switch for RPTPσ Clustering and Neuronal Extension. *Science.* **332**: 484-488

28. Barlow D J and Thornton J M (1983) Ion-pairs in proteins. *J. Mol. Biol.* **168**: 867-885.

29. Xu D, Tsai CJ, et al. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.*, **10**: 999-1012.

30. Lindahl E, Hess B, et al. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**: 306-317.

31. Hess B, Bekker H, et al. (1997) "LINCS: A Linear Constraint Solver for Molecular Simulations", *J. Comput. Chem.*,**18**(12): 1463-1472.

32. Berendsen HJC, Postma JPM *et al.* (1984) Molecular dynamics with coupling to an external bath.*J. Chem. Phys***81:** 3684–3690.

33. Essmann U, Perera L, et al. (1995) A smooth particle mesh Ewald method. J. Chem. Phys., 103(19): 8577–8593.

34. Bothra AK, Roy S, et al. (1998) Molecular dynamics simulation of colchicinoids. *J. B. S. D.*,**15**: 999-1008.

# Chapter VI

# QSAR studies of Anthrax Lethal inhibitors through Quantum Chemical Indices

## 6.1: Introduction

Bacillus Anthracis is a gram positive, rod shaped spore- forming bacterium that causes infectious disease Anthrax [**1**]. The bacterium secreted three proteins protective antigen (PA), which is a pore forming sub unit; edema factor (EF), which is an adnylatecyclase that increases c AMP levels in cells and lethal factor (LF), which is a protease that cleaves mitogen activated protein kinase - kinase family [2-5]. These three proteins are individually non toxic but assemble to form two toxic complex, lethal toxin, and edema toxin [**6**]. Initially PA binds to host cell surface and cleaved by a cell associated furine type protease to form PA63 [7-10]. Then PA63 oligomerizes in to a heptameric pore-forming complex to which LF and EF bind [11]. The resulting pore-forming complex is internalized by receptor-mediated endocytosis and is trafficked in to a low pH endosome . Then complex undergoes a conformational change that allows the EF and LF to translocated in to the host cell cystol and exert their toxic action [12-14]. Inside the cystol, LF is then able to cleave mitogen activated protein kinases kinases (MAPKK) and disrupting interaction with mitogen activated protein kinase, which in turn results in inhibition of the signaling path way [15-17]. Hydroxamate analogues are used as inhibitors of the anthrax lethal toxin [18].

In this paper, we developed six QSAR models by stepwise regression analysis with respect to their experimental value of lnIc50. The method is very useful in elucidating the mechanism of chemical and biological interaction in various biomolecules. The QSAR approach employs thermodynamically derived and computational based descriptors to correlate biological activity (lnIc50) in isolated receptors. Three standard quantum chemical descriptors routinely used in QSAR analysis are quantum, electronic, surface and steric. For screening of chemical database or virtual libraries before their synthesis,

the use of QSAR models appears equally attractive to chemical manufacturers, pharmaceutical companies.

### 6.2: Materials and Methods

Structural and biological data of hydroxamate analogue was collected from site Binding db (www.bindingdb.org) as shown in Table 1 [19-21]. Structure of Hydroxamate analogue is represented in Figure 1.

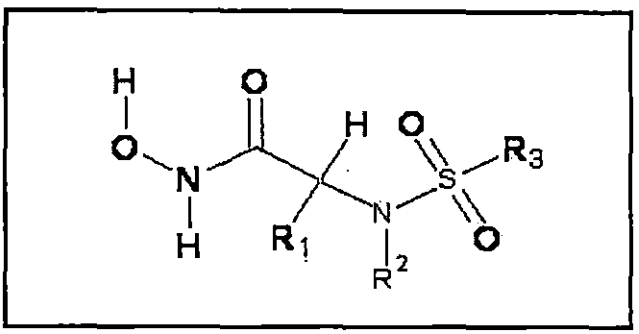



**Figure 1: Structural representation of Hydroxamate analogue**

**Table 1: Chemical structure of Hydroxamate analogue by substituting R1, R2, R3 of figure 1.**

| Compd No. | R1 | R2 | R3 |
|---|---|---|---|
| 1 | $-CH(CH_3)_2$ | -H | $4-C_6H_4-OCH_3$ |
| 2 | $-CH(CH_3)_2$ | $4-CH_2$-Pyrdinyl | $4-C_6H_4-OCH_3$ |
| 3 | $-CH(CH_3)_2$ | 3-$CH_2CH_2$Thiophenyl | $4-C_6H_4-OCH_3$ |
| 4 | $-CH(CH_3)_2$ | 1-$CH_2$-$CH_2$-Imidazoly | $4-C_6H_4-OCH_3$ |
| 5 | $-CH(CH_3)_2$ | -H | $C_6H_5$ |
| 6 | $-CH(CH_3)_2$ | -H | $2-F-C_6H_5$ |
| 7 | $-CH(CH_3)_2$ | -H | $3-Cl-C_6H_5$ |
| 8 | $-CH(CH_3)_2$ | -H | $4-F-C_6H_5$ |
| 9 | $-CH(CH_3)_2$ | -H | $4-CH_3-C_6H_5$ |
| 10 | $-CH(CH_3)_2$ | -H | $3-CH_3-C_6H_5$ |

| Compd No. | R1 | R2 | R3 |
|---|---|---|---|
| 11 | $-CH(CH_3)_2$ | -H | 4-F-3-$CH_3$-$C_6H_5$ |
| 12 | $-CH(CH_3)_2$ | -H | 3-F-4-$CH_3$-$C_6H_5$ |
| 13 | $-CH(CH_3)_2$ | -H | $-CH_2-CH_2-C_6H_5$ |
| 14 | $-CH(CH_3)_2$ | -H | 3-Chloro- benzyl |
| 15 | $-CH(CH_3)_2$ | -H | 4-Hydroxy-3-methyl- $C_6H_5$ |
| 16 | $-CH(CH_3)_2$ | -H | 4-Fluoro-3-hydroxymethyl-$C_6H_5$ |
| 17 | $-CH(CH_3)_2$ | -H | 4-Fluoro-2-hydroxymethyl-$C_6H_5$ |
| 18 | $-CH(CH_3)_2$ | -H | 2-Chloro-5-Thiophenyl |
| 19 | $-CH(CH_3)_2$ | -H | 2-{5-(1,2,3-Thiadiazolyl)-thiophenyl |
| 20 | $-CH_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 21 | $-CH_2CH_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 22 | $-(CH_2)_3CH_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 23 | $-C(CH_3)_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 24 | Cyclopropyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 25 | Cyclobutyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 26 | Cyclopentyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 27 | Cyclohexyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 28 | Phenyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 29 | Cyclohexyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 30 | 3-Tetrahydrofuranyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 31 | $-CH_2(CH_3)CH_2CH_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 32 | $-CH_2(CH_3)CH_2CH_3$ | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 33 | 2-(Cyclopropyl)Methyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |
| 34 | 2-(1Cyclopropyl)Ethyl | -H | 4-Fluro -3-methyl - $C_6H_5$ |

The QSAR studies were carried out using 34 compounds of hydroxamate analogue. The statistically significant model was constructed from training set by using eight

parameters. The data set was divided in to training set of 25 compounds and test set of 9 compounds on the basis of their activity.

Quantum chemical descriptors namely HOMO, LUMO, dipole moment (DM) were calculated using GAMESS [22]. van der Waals surface area (VSA) , molar refractivity (MR), solvent accessible surface area (SASA), polarizability (POL), Partition coefficient (logP) were calculated using Mervin logP calculator. The 2D geometry of compounds was drawn in ChemSketch and converted into 3D model in 3D viewer. Energy minimization was done under MOPAC module according to AM1 (Austin Model 1) method using RHF (restricted HartreeeFock: closed shell) wave function.

### 6.3: Results and Discussion

The values of quantum chemical indices and natural logarithm of experimental $IC_{50}$ value of training set is given in Table 2 and test set is given in Table 6. Correlation matrix for descriptors of training set given in Table 3

**Table 2: Values of quantum chemical indices and experimental value of $lnIC_{50}$ of training set**

| Compd | VSA | MR | SASA | POL | logP | HOMO | LUMO | DM | $lnIC_{50}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 6.0866 | 4.2868 | 6.24 | 3.3996 | 0.961 | -0.2462 | -0.0339 | 6.7728 | 6.8024 |
| 2 | 6.3491 | 4.6061 | 6.3343 | 3.6623 | 1.315 | -0.2135 | -0.0433 | 6.4262 | 8.0392 |
| 3 | 6.1647 | 4.6454 | 6.1191 | 3.7402 | 2.598 | -0.1933 | -0.0414 | 4.2424 | 7.6497 |
| 4 | 6.3631 | 4.6009 | 6.4264 | 3.6855 | 0.888 | -0.2045 | -0.0367 | 9.1421 | 9.8574 |
| 5 | 5.9753 | 4.1938 | 6.1289 | 3.2883 | 0.904 | -0.2298 | -0.0448 | 7.7259 | 7.0031 |
| 6 | 5.796 | 4.1975 | 6.1255 | 3.2846 | 1.02 | -0.2438 | -0.0542 | 7.6089 | 8.7948 |
| 7 | 5.7643 | 4.1975 | 5.8296 | 3.2915 | 1.044 | -0.1461 | -0.039 | 8.3645 | 7.9374 |
| 8 | 5.8181 | 4.1975 | 5.9459 | 3.2918 | 1.068 | -0.2156 | -0.0582 | 8.4473 | 5.6699 |
| 9 | 6.0542 | 4.2671 | 6.2168 | 3.3706 | 1.353 | -0.2406 | -0.0408 | 5.1837 | 8.0064 |
| 10 | 6.053 | 4.2671 | 6.2184 | 3.3639 | 1.329 | -0.2204 | -0.0462 | 7.4782 | 5.9915 |
| 11 | 6.0678 | 4.2701 | 6.2316 | 3.363 | 1.444 | -0.2451 | -0.0482 | 7.3547 | 4.8675 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 12 | 5.8508 | 4.2701 | 5.9275 | 3.3734 | 1.444 | -0.2175 | -0.0488 | 5.5005 | 8.5526 |
| 13 | 5.902 | 4.3257 | 5.9437 | 3.433 | 1.175 | -0.2247 | -0.0119 | 6.2454 | 9.3826 |
| 14 | 5.8832 | 4.3263 | 5.9283 | 3.4306 | 1.62 | -0.1225 | -0.0371 | 1.6783 | 9.7172 |
| 15 | 6.0742 | 4.2945 | 6.2261 | 3.3911 | 1.269 | -0.2056 | -0.0212 | 7.5077 | 7.4025 |
| 16 | 5.8734 | 4.2946 | 5.9116 | 3.3995 | 0.334 | -0.1566 | -0.0475 | 1.932 | 6.3969 |
| 17 | 6.0869 | 4.2946 | 6.1758 | 3.3814 | 0.334 | -0.2394 | -0.0508 | 5.6045 | 7.1778 |
| 18 | 5.7072 | 4.1821 | 5.7949 | 3.3595 | 1.654 | -0.2527 | -0.0526 | 1.8346 | 7.1778 |
| 19 | 6.0945 | 4.4011 | 6.2802 | 3.4391 | 1.487 | -0.2492 | -0.0761 | 4.8421 | 6.697 |
| 20 | 5.9221 | 4.1357 | 6.1201 | 3.2122 | 0.665 | -0.2543 | -0.0376 | 1.5427 | 4.8675 |
| 21 | 5.7737 | 4.2056 | 5.8657 | 3.2995 | 1.201 | -0.245 | -0.0398 | 1.4663 | 4.6052 |
| 22 | 6.1395 | 4.3342 | 6.3489 | 3.4447 | 2.265 | -0.2482 | -0.0377 | 3.2164 | 4.2485 |
| 23 | 5.9504 | 4.3298 | 6.0264 | 3.3986 | 2.025 | -0.2035 | -0.0571 | 8.362 | 5.7038 |
| 24 | 6.0125 | 4.2446 | 6.1579 | 3.3285 | 1.192 | -0.2037 | -0.0532 | 7.8314 | 4.7875 |
| 25 | 6.0789 | 4.3085 | 6.2031 | 3.4004 | 1.366 | -0.1316 | -0.0342 | 3.3367 | 4.8675 |

**Table 3: Correlation matrix for descriptors of training set**

| | **VSA** | **MR** | **SASA** | **POL** | **logP** | **HOMO** | **LUMO** | **DM** | **lnIC$_{50}$** |
|---|---|---|---|---|---|---|---|---|---|
| **VSA** | 1 | 0.7762 | 0.895 | 0.7289 | 0.1426 | -0.0477 | 0.1194 | 0.264 | 0.0643 |
| **MR** | 0.7762 | 1 | 0.4965 | 0.9837 | 0.377 | 0.1856 | 0.0786 | 0.1178 | 0.3652 |
| **SASA** | 0.895 | 0.4965 | 1 | 0.432 | 0.0624 | -0.2605 | 0.0133 | 0.3181 | -0.0797 |
| **POL** | 0.7289 | 0.9837 | 0.432 | 1 | 0.4139 | 0.1879 | 0.1405 | 0.0642 | 0.3989 |
| **logP** | 0.1426 | 0.377 | 0.0624 | 0.4139 | 1 | 0.0245 | -0.0435 | -0.101 | -0.0839 |
| **HOMO** | -0.0477 | 0.1856 | -0.2605 | 0.1879 | 0.0245 | 1 | 0.2127 | -0.0635 | 0.2176 |
| **LUMO** | 0.1194 | 0.0786 | 0.0133 | 0.1405 | -0.0435 | 0.2127 | 1 | -0.0752 | 0.2457 |
| **DM** | 0.264 | 0.1178 | 0.3181 | 0.0642 | -0.101 | -0.0635 | -0.0752 | 1 | 0.2068 |
| **lnIC$_{50}$** | 0.0643 | 0.3652 | -0.0797 | 0.3989 | -0.0839 | 0.2176 | 0.2457 | 0.2068 | 1 |

From the correlation matrix it is clear molar refractivity and polarizability have moderately high correlation with ln $IC_{50}$. Partition coefficient and van der Waals surface area has very low correlation. HOMO, LUMO and Dipole moment has moderate correlation.

A multivariate regression was performed using eight (8) different indices by stepwise addition method [23]. Since HOMO and LUMO plays an important role in transition as well as reaction so we consider these two indexes as our initial step. Using these two quantum chemical parameters Model 1 (Table 4) was constructed and shows a lower F value and correlation coefficient.

**Table 4: Regression equations and Fischer F-value using different indices**

| Model No. | Regression equations | r | $r^2$ | F |
|---|---|---|---|---|
| Model 1 | $\ln Ic_{50} = 9.728269 + (7.6253)HOMO + (27.6314)LUMO$ | 0.298 | 0.089 | 3.57 |
| Model 2 | $\ln Ic_{50} = 9.054719 + (8.1413)HOMO + (7.6253)LUMO + (0.1561)DM$ | 0.379 | 0.144 | 7.76 |
| Model 3 | $\ln IC_{50} = 9.304326 + (-0.1809)\log P + (8.2142)\text{HOMO} + (29.2211)\text{LUMO} + 0.1523)\text{DM}$ | 0.383 | 0.147 | 10.3 |
| Model 4 | $\ln IC_{50} = (-10.881087) + (5.9332)\text{POL} + (-0.7994)\text{logP} + (5.1477)\text{HOMO} + (21.2275)\text{LUMO} + (0.1146)\text{DM}$ | 0.547 | 0.299 | 19.9 |
| Model 5 | $\ln IC_{50} = 0.29 + (-17.8827)\text{SASA} + (8.1232)\text{VSA} + (19.6404)\text{Pol} + (-1.4596)\text{logP} + (2.4232)\text{HOMO} + (31.0984)\text{LUMO} + (0.1962)\text{DM}$ | 0.757 | 0.573 | 29.1 |
| Model 6 | $\ln IC_{50} = (-0.9232) + (-18.1715)\text{VSA} + (2.3851)\text{MR} + (8.1614)\text{SASA} + (17.4222)\text{POL} + (-1.4521)\text{logP} + (2.1142)\text{HOMO} + (32.8694)\text{LUMO} + (0.1934)\text{DM}$ | 0.758 | 0.575 | 31.2 |

ln $IC_{50}$- natural logarithm of Ic50, HOMO- value of HOMO energy, LUMO- value of LUMO energy, DM- dipole moment, SASA-solvent accessible surface area, VSA- van

der Waals surface area. MR- molar refractivity, POL- polarizability, LogP-octanol/water partition coefficient.

In the next step together with HOMO, LUMO we introduce dipole moment (DM). As a result it shows a higher F (7.76) and r (0.379) value than previous one. To modify the results we introduce logP (Model 3) in Model 2. It shows a higher $r^2$ and F value. polarizability, solvent accessible surface area (SASA), van der Waals surface area (VSA) are added stepwise in Model 4 and Model 5. Model 5 shows a larger r (0.757) and F (29.11) value than Model 4. From this step it is clear that solvent accessible surface area and van der Waals surface area plays an active role in activity. Finally with Model 5 molar refractivity is introduced and constructed Model 6. Model 6 exhibits slightly greater r value (0.758) but modified F (31.29) value. The model 6 having good correlation coefficient value (0.758) explains 57.5% variance in the anthrax lethal toxin inhibitory activity. A satisfactory QSAR model (model 6) was obtained with LOO cross-validation values of 0.56. The higher F value indicate the model is statistically more significant than the other model but model 6 and model 5 both r, $r^2$, F values are very closes to each other which indicate that little contribution of molar refractivity on regression equation. Therefore steric effect does not improve the activity of hydoxamate inhibitor. Experimental activities and predicted activities of training set represented in Table 5. The predicted biological activities of the compounds in the test set using model 6 are presented in Table 6 with experimental activities and $R^2_{pred}$ for test is 0.267.

**Table 5: Values of experimental $lnIC_{50}$ and predicted $lnIC_{50}$ of training set using model 6**

| Compd No. | $lnIC_{50}$ | $lnIC_{50}$ (Model 6) |
|---|---|---|
| 1 | 6.8024 | 7.128667 |
| 2 | 8.0392 | 7.647253 |
| 3 | 7.6497 | 8.522042 |
| 4 | 9.8574 | 9.905629 |
| 5 | 7.0031 | 6.022826 |
| 6 | 8.7948 | 8.668462 |
| 7 | 7.9374 | 7.763879 |
| 8 | 5.6699 | 6.943415 |
| 9 | 8.0064 | 5.891324 |
| 10 | 5.9915 | 6.143184 |
| 11 | 4.8675 | 5.665155 |
| 12 | 8.5526 | 6.995867 |
| 13 | 9.3826 | 9.097739 |
| 19 | 6.697 | 5.751855 |
| 20 | 4.8675 | 4.816607 |
| 21 | 4.6052 | 6.279258 |
| 22 | 4.2485 | 5.260135 |
| 23 | 5.7038 | 6.02852 |
| 24 | 4.7875 | 5.785872 |
| 25 | 4.8675 | 6.028028 |

**Table 6: Values of quantum chemical indices, experimental and predicted (using Model 6) value of ln $IC_{50}$ of test set**

| Compd No. | VSA | MR | SASA | POL | logP | HOMO | LUMO | DM | $lnIC_{50}$ | $lnIC_{50}$ (Model 6) |
|---|---|---|---|---|---|---|---|---|---|---|
| 26 | 6.1422 | 4.3686 | 6.2372 | 3.4634 | 2.105 | -0.2146 | -0.0449 | 5.8735 | 4.1589 | 5.276187 |
| 27 | 6.0387 | 4.4252 | 6.0776 | 3.5092 | 2.611 | -0.2021 | -0.0567 | 8.7496 | 3.6376 | 6.234704 |
| 28 | 6.1003 | 4.4116 | 6.2636 | 3.4624 | 1.917 | -0.1599 | -0.0541 | 9.5845 | 3.912 | 7.125803 |
| 29 | 6.1722 | 4.3954 | 6.2541 | 3.4881 | 1.08 | -0.2164 | -0.0418 | 4.8182 | 3.9889 | 6.750255 |
| 30 | 5.8979 | 4.3357 | 5.9139 | 3.4241 | 0.809 | -0.1224 | -0.0219 | 3.5729 | 4.7875 | 8.711773 |
| 31 | 5.8965 | 4.3325 | 5.9143 | 3.4135 | 1.947 | -0.2207 | -0.0366 | 3.8311 | 4.2047 | 6.253476 |
| 32 | 6.1393 | 4.3325 | 6.2605 | 3.4206 | 1.947 | -0.2151 | -0.0514 | 8.1483 | 4.2047 | 5.131939 |
| 33 | 6.0846 | 4.3095 | 6.2236 | 3.3995 | 1.463 | -0.2386 | -0.0405 | 1.2967 | 3.9318 | 5.118746 |
| 34 | 5.9663 | 4.3679 | 6.0579 | 3.4706 | 1.939 | -0.2074 | -0.0535 | 7.2957 | 4.1744 | 7.375374 |

Hydroxamate analogue (1) exhibit high activity against anthrax lethal toxin. When hydrogen atom of amino group replaced by methyl pyridinyl, ethyl thophenyl, ethyl imidazolyl group of hydroxamate inhibitor (1) of table 1, exhibited high value of inhibition activity, which is evident from the hydroxamate derivatives (2, 3, 4). When methoxy group is replaced by fluorine atom (8), the inhibition activity was reduced. But fluorine atom present at ortho and meta position of phenyl ring (6, 7) displayed high activity. Introduction of methyl group at the para position of phenyl ring (9) would improve the inhibition activity. But methyl group present at the meta position of phenyl ring (10), the activity decreases. Replacing benzene ring with ethyl phenyl group (13) and Chloro benzyl group (14) displayed outstanding activity. When isopropyl group substituted by ethyl, butyl, t-butyl, cyclopropyl, cyclobutyl and similar groups (20-34), are reduced the inhibition activity towards anthrax lethal toxin. On the basis of these

observations we have designed 17 inhibitors and calculated their activity using Regression model 6 exhibits in Table 7. It is found that designed compounds are significant value of inhibition activity (Table 8).

**Table 7: Chemical structure of designed compounds**

| Compd No. | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| d1 | $-CH(CH_3)_2$ | -H | 4-Nitro-$C_6H_5$ |
| d2 | $-CH(CH_3)_2$ | -H | 4-$N(CH_3)_2$-$C_6H_5$ |
| d3 | $-CH(CH_3)_2$ | 4(Pyridnyl)-$CH_2$-$CH_2$ | 4- $OCH_3$-$C_6H_4$ |
| d4 | $-CH(CH_3)_2$ | 3-$CH_2$-Thiophenyl | 4- $OCH_3$-$C_6H_4$ |
| d5 | $-CH(CH_3)_2$ | -H | $2CH_3$-$C_6H_5$ |
| d6 | $-CH(CH_3)_2$ | -H | 4-$CCl_3$-$C_6H_5$ |
| d7 | $-CH(CH_3)_2$ | -H | 2-(2F-$C_6H_5$) Ethyl |
| d8 | $-CH(CH_3)_2$ | -H | 4-SH-3-$CH_3$-$C_6H_5$ |
| d9 | $-CH(CH_3)_2$ | -H | 4-F-3-($CH_2$-SH)-$C_6H_5$ |
| d10 | $-CH(CH_3)_2$ | -H | 4-$OCF_3$-$C_6H_5$ |
| d11 | 2(2-Oxo pyrimidinyl) | -H | 4-F-3-$CH_3$-$C_6H_5$ |
| d12 | $-CH(CH_3)_2$ | $-CH_2$-(2-Oxopyridinyl) | 4- $OCH_3$-$C_6H_4$ |
| d13 | $-CH(CH_3)_2$ | $-CH_2$-(4NO2-$C_6H_5$) | 4- $OCH_3$-$C_6H_4$ |
| d14 | $-CH(CH_3)_2$ | 4- Methoxy- benzyl | 4- $OCH_3$-$C_6H_4$ |
| d15 | $-CH(CH_3)_2$ | -H | 4($NH_2$-CO-NH)$C_6H_5$ |
| d16 | $-CH(CH_3)_2$ | -H | 4($NH_2$-CS-NH)$C_6H_5$ |
| d17 | $-CH(CH_3)_2$ | -H | 4($NH_2$-CO-NH-NH)$C_6H_5$ |

**Table 8: Designed compounds and their predicted $\ln IC_{50}$ using Model 6.**

| Compd No. | VSA | MR | SASA | POL | logP | HOMO | LUMO | DM | $\ln IC_{50}$ |
|---|---|---|---|---|---|---|---|---|---|
| d1 | 6.0766 | 4.2986 | 6.1874 | 3.3629 | 0.863 | -0.2698 | -0.1244 | 2.0264 | 2.474801 |
| d2 | 6.1779 | 4.3908 | 6.2804 | 3.4637 | 1.006 | -0.1244 | -0.0324 | 11.7828 | 8.379723 |
| d3 | 6.403 | 4.6525 | 6.4824 | 3.7243 | 1.721 | -0.1024 | -0.0307 | 7.5891 | 9.355408 |
| d4 | 6.3271 | 4.5987 | 6.3447 | 3.6736 | 2.192 | -0.2015 | -0.0475 | 7.9294 | 7.219331 |
| d5 | 6.054 | 4.2671 | 6.2052 | 3.3596 | 1.305 | -0.233 | -0.0264 | 7.0286 | 6.522726 |
| d6 | 5.8884 | 4.463 | 5.9371 | 3.5683 | 2.723 | -0.1991 | -0.0876 | 6.7121 | 7.386881 |
| d7 | 5.9017 | 4.3285 | 5.9257 | 3.4362 | 1.291 | -0.2393 | -0.0117 | 5.7886 | 8.740487 |
| d8 | 6.1042 | 4.3735 | 6.2703 | 3.4638 | 1.51 | -0.2388 | -0.0504 | 0.4293 | 5.835871 |
| d9 | 5.8844 | 4.3737 | 5.9317 | 3.465 | 1.287 | -0.1256 | -0.0397 | 3.7106 | 8.637351 |
| d10 | 6.1344 | 4.2391 | 6.2477 | 3.4006 | 1.874 | -0.201 | -0.0389 | 6.0376 | 4.69501 |
| d11 | 6.0601 | 4.3939 | 6.1824 | 3.4247 | -0.601 | -0.1203 | -0.0404 | 5.4162 | 9.896373 |
| d12 | 6.3268 | 4.6096 | 6.334 | 3.6861 | 0.282 | -0.0789 | -0.0368 | 4.9994 | 10.19898 |
| d13 | 6.4303 | 4.6966 | 6.4724 | 3.7443 | 2.563 | -0.2288 | -0.0785 | 14.2525 | 7.458973 |
| d14 | 6.4424 | 4.6887 | 6.5033 | 3.7598 | 2.661 | -0.1998 | -0.0411 | 8.2498 | 7.7299 |
| d15 | 6.1249 | 4.3797 | 6.2087 | 3.4527 | 0.038 | -0.1051 | -0.0224 | 4.954 | 8.993963 |
| d16 | 6.1355 | 4.4751 | 6.2037 | 3.5119 | -0.059 | -0.1956 | -0.0528 | 12.4105 | 10.41185 |
| d17 | 6.1726 | 4.4193 | 6.2776 | 3.4792 | -0.203 | -0.111 | -0.0276 | 5.4492 | 9.507978 |

**6.4: References**


1. Smith H (1954) Observations on Experimental Anthrax: Demonstration of a Specific Lethal Factor produced in vivo by Bacillus anthracis. J Keppie. *Nature*.**173**: 869.

2. Hanna P (1998) Anthrax pathogenesis and host responseCurr. *Top. Microbial. Immunol.*, **225**: 13-35.

3. Liu S, Schubert RL, Bugge TH, Leppla SH (2003) Anthrax toxin: structures, functions and tumour targeting. *Expert OpinBiolTher.*, **3**: 843-853.

4. SH Leppla (2006) Bacillus anthrasis toxins. The Comprehensive sourcebook of bacterial Protein Toxins. EdsAlouf JE, Popoff MR (Acadamic, Burlington, MA) 323-347.

5. Young JA, Collier RJ (2007) Anthrax toxin: Receptor-binding, internalization, pore formation, and translocation. Annu Rev Biochem.,**76:** 243-265

6. Brossier F, M Mock(2001) Toxins of Bacillus anthrasis. Toxicon**39**:1747-1755.

7. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC (1997) Crystal structure of Anthrax toxin protective antigen. *Nature*.**385**: 833-838.

8. Bradley KA, MogridgeJ, Mourez M, CollierRJ, YoungJA(2001) Identification of the cellular receptor for anthrax toxin. *Nature*.**414**: 225-229

9. Lacy DB, WigelsworthDJ,ScobieHM, YoungJA, CollierRJ (2004) Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an toxin receptor. *Proc. Natl. Acad. Sci. USA*.**101**: 6367-6372

10. Wigelsworth DJ, KrantzBA, KA ChristensenKA, LacyDB, JurisSJ, CollierRJ(2004) Binding stoichiometry and kineticks of interaction of human anthrax toxin receptor. CMG2, with protective antigen *J. Biol. Chem.*.**279**: 23349-23356

11. Scobie HM, Rainey GJ, Bradley KA, Young, JA. (2003) Human capillary morphogenesis protein 2 function as an anthrax toxin receptor. *ProcNatlAcadsci U.S.A.* **100**: 5170-5174.

12. Leppla SH (1982), Anthrax toxin edema factor: A bacterial adenylatecyclase that increases cyclic AMP concentrations of eukaryotic cells. *ProcNatl Acad. U.S.A.* **79:** 3162-3166.

13. Vitale G,Pellizzari R, Recchi C, Napolitani G, Mock M et al. (1998) Anthrax lethal factor cleaves the N- terminus of MAPKKS and induces tyrosine/threonine phosphorylation of MAPKKs in cultured macrophages. Biochem. biophys. *Res. Commun***248**: 706-711

14. Duesbery NS, Webb CP, Leppla SH, Gordon VM, Klimpel KR, et al. (1998)Proteolytic inactivation of MAP-kinase-kinase by Anthrax lethal factor G.F. *Science.* **280**: 734-737.

15. Pellizzari R, Guidi-Rontani C, Vitale G, Mock M, Montecucco C (1999) Anthrax lethal factor cleaves MKK3 in macrophages and inhibits the LPS/IFNgamma-induced release of NO and TNFalpha.FEBS Lett.,**462:**199–204.

16. VITALE G, BERNARDI L, NAPOLITANI G, MOCK M, MONTECUCCO C (2000)*Biochem. J.*, **352**:739-745.

17. Vitale G, Pellizzari R, Recchi C, Napolitani G, Mock M, et al. (1998) *Biochem. Biophys. Res. Commun.* 248: 706–711.

18. Thomas M, Castignetti D (2009) Examination of anthrax lethal factor inhibition by siderophores, small hydroxamates, and protamine. *J MicrobiolImmunol Infect.*, **42**: 284-289.

19. Chen X, Liu M, GilsonMK (2001) Binding DB: A web-accessible molecular recognition database J. Combi. *Chem. High-Throughput Screen.* **4**: 719-725.

20. Chen X, Lin Y, Gilson MK (2002) The Binding Database: Overview and User's Guide. *Biopolymers NucleicAcid Sci.*, **61**: 127-141.

21. Chen, X., Liu, M., Gilson, MK. (2001) The binding database: Overview and user's guide *Biopolymers/Nucleic Acid Sci.*,**61**: 127–141.

22. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, et al. (1993) General atomic and molecular electronic structure system*J Comput Chem.*, *14*: 1347-1363.

23. Padavala AB, Prasanth VV, Jayanthi SA, Vadlamani A, Chitti S (2010) QSAR analysis and validation studies on substituted spiropiperidines as GlyT1 inhibitors*J. Chem. Pharm. Res.*, **2**: 147-162.

# Chapter VII

# *Molecular dynamics simulation of human bifunctional glutamyl-prolyl-tRNA synthetase*

## 7.1: Introduction

Aminoacyl –tRNA synthetases catalyze the attachment of amino acids to their specific tRNAs in protein synthesis. They are also involved in various biological processes. In higher eukaryotes several of these enzymes are found in a multienzyme complex [1-4]. These multiprotine complexes are composed by nine synthetases (IRS, LRS, MRS, QRS, PRS, KRS, DRS, and bifunctional EPRS) and three auxiliary proteins p18, p38, p43 [5, 6]. These nine synthetases react with Glu, Pro, Ile, Leu, Met, Lts, Gln, Arg, Asp [7]. With the tRNA synthetase of higher eukaryotes, glutamyl and prolyl – tRNAsynthease of higher eukaryotes catalytic activities have been found linked in a single polypeptide [8, 9]. In human glutamyl - prolyl – tRNAsynthease(EPRS), two catalytic domains exhibiting each enzyme activity are linked by a linker peptide that contains three tendemly repeated motifs (EPRS-R1, EPRS-R2, EPRS-R3) of 57 amino acids[9]. In human prolyl – tRNA synthease absent of this linker peptide was still active, suggesting that it is not essential for catalytic activity. Therefore linker region may play a different role in the cell other than the catalytic function [10]. Peptide sequences homologus to these repeats have also been found in other tRNA synthetases. They are present as a single copy in the N-terminal extensions of glycyl, tryptophanyl, histidyl tRNA synthetases, which have been found as free forms and in the C-terminal extension methionyl t-RNA synthetase has been found in the complex form (11-15). These motifs were identified as an antigenic epitope for auto antibodies detected in myositis patients (16). In Human histidyl-t RNA synthetase due to the lacking of this peptide region it is lost antigenicity and catalytic activities, suggesting its importance in these two activities (17). EPRS-R1 contain helix-turn-helix structural motif (18). Jeong et al. solved the NMR structure of multifunctional peptide motifs in human bifunctional glutamyl-prolyl-tRNAsynthetase [19]. Two helices are found in residues from 679 to 699 and from 702 to

721. Two helices are interacting with hydrophobic residues [Val (692), Leu (695), Ala (670), Val (705), and Ala (708)] close to turn involved in helix-helix interaction. The C-terminal loop folds back to interact with the aromatic residues. The hydrophobic interaction between Tyr (727) in the loop Tyr (682), Tyr (719) in the helix are essential for the C-terminal loop formation. Also hydrophobic residues are involved in helix-helix interaction but other are on the surface. These suggest that they may in role in protein-protein or protein-nucleic acid interaction. When non catalytic motifs attached to the catalytic domains of tRNAsynthetase including the repeated motifs of EPRS, exert biological function. One is that they are involved in protein-protein interaction for the multi-tRNAsynthetase complex and another is that they used as a tRNA-binding motif for delivery and efficiency. The RNA binding mode of this helix- turn-helix motifs also found in other homologous RNA binding motifs and used in protein RNA interaction. Repetition of helix-turn-helix to enhanced functional flexibility and binding affinity in molecular interaction.

In this work we intend to obtain the motional properties of protein, its secondary structure and important residues by monitoring RMSD, Radius of gyration and also through principal component analysis.

**7.2: Materials and Methods**

Solution structure (NMR) of the protein (Protein Data Bank code 1FYJ) was used as starting structure [19] was used as a starting structure. A single monomer was solvated with SPC water molecules in a cubic box, having an edge length of 35A°. The simulation was performed using GROningenMAchine for Chemical Simulation [20]. The LINCS algorithm was used to constrain all bond lengths [21]. The simulation was conducted at a constant temperature (300 K), coupling each component separately to a temperature bath using the Berendsen coupling method [22]. A cutoff of 0.9 nm was used for Lennard

Jones interaction and 1.0 nm for Coulomb interaction. The time step was 2 fs, with coordinates stored after every 4 ps. MD simulation was performed for 12ns. Before running simulation, an energy minimization was performed in steepest descent method followed by conjugate gradient method [23, 24]; and this was followed by 1.0 ns of simulation imposing positional restraints on the non-H atoms. The positional restraints were then released and 12ns production run were obtained and analyzed. Analysis programs from GROMACS were used and PCA was performed with the MD trajectory.

### 7.3: Results and discussion

The overall structural stability of the protein during the simulation has been monitored using several parameters like RMSD, radius of gyration (Rg), RMSF etc.

The time evolutions of RMSD of the whole protein of human bifunctional glutamyl-prolyl-tRNA during the full simulation time (Fig 1) shows an initial drift in RMSD which may be due to the difference of crystal structure with solution structure. It is evident from figure 1 that RMSD slightly increases up to 4000ps, and then slightly decreases and again increases up to 9500ps and after that almost become stable.

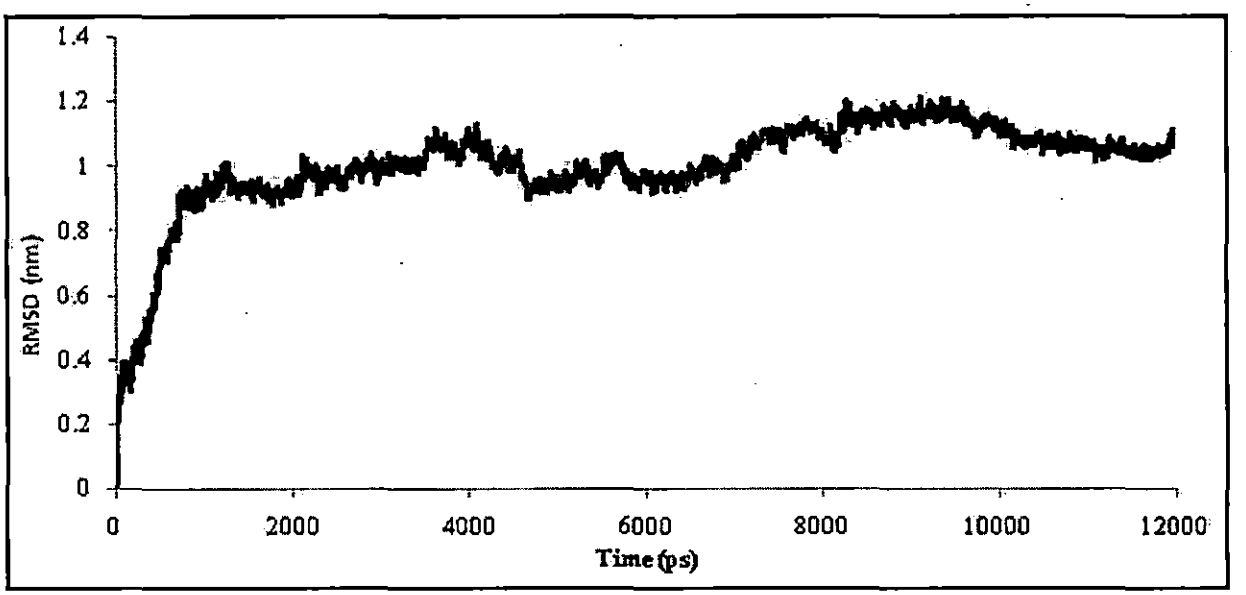



**Fig 1: Time evolution of RMSD during whole simulation of time**

The variation of radius of gyration (Rg) as function of time is presented in figure 2, and from this figure it is clear that Rg show much variation during the simulation time which indicates that the protein is much flexible.

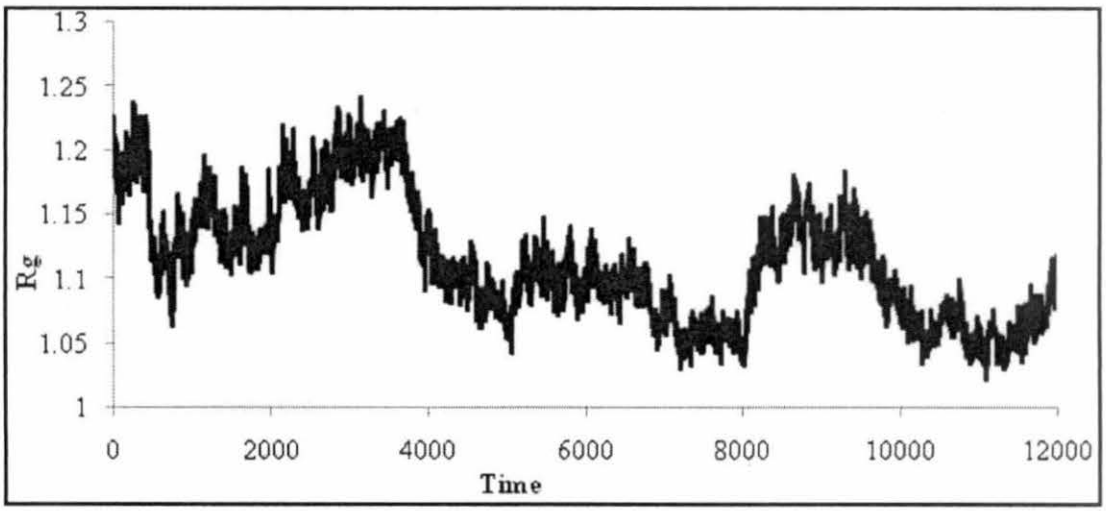



**Fig 2: Time evolution Rg changes in aqueous medium during 12,000 ps Dynamics simulation**

The flexibility of different segments of the protein is also revealed by looking at the root mean-square fluctuation (RMSF) of each residue from its time-averaged position. RMSF of Cα is presented as a function of residue number in figure 3. From RMSF, it is evident that first and last residue fluctuates considerably. Interestingly pronounced fluctuations are observed for both the helixes. The hydrophobic interaction betweenTyr727in the loop Tyr682, Tyr719 in the helix shows less fluctuation.

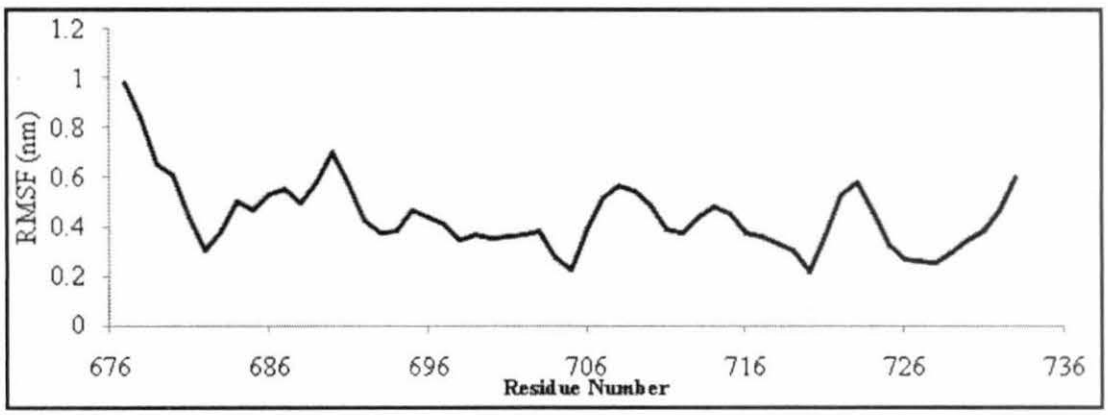



**Fig 3: Plot of RMSF value of Cα atoms value in aqueous medium presented as a function of residue number of human bifunctional glutamyl-prolyl-tRNA synthetase in starting NMR structure**

The flexibility of different segments of the protein is also revealed by looking at the root mean-square fluctuation (RMSF) of each residue from its time-averaged position. RMSF of Cα is presented as a function of residue number in figure 3. From RMSF, it is evident that first and last residue fluctuates considerably. Interestingly pronounced fluctuations are observed for both the helixes. The hydrophobic interaction betweenTyr727in the loop Tyr682, Tyr719 in the helix shows less fluctuation.

RMSD of C terminal α helixes, H1 and H2 are computed and presented in figure 4. It is observed that helix H1 have more fluctuations than helix H2.

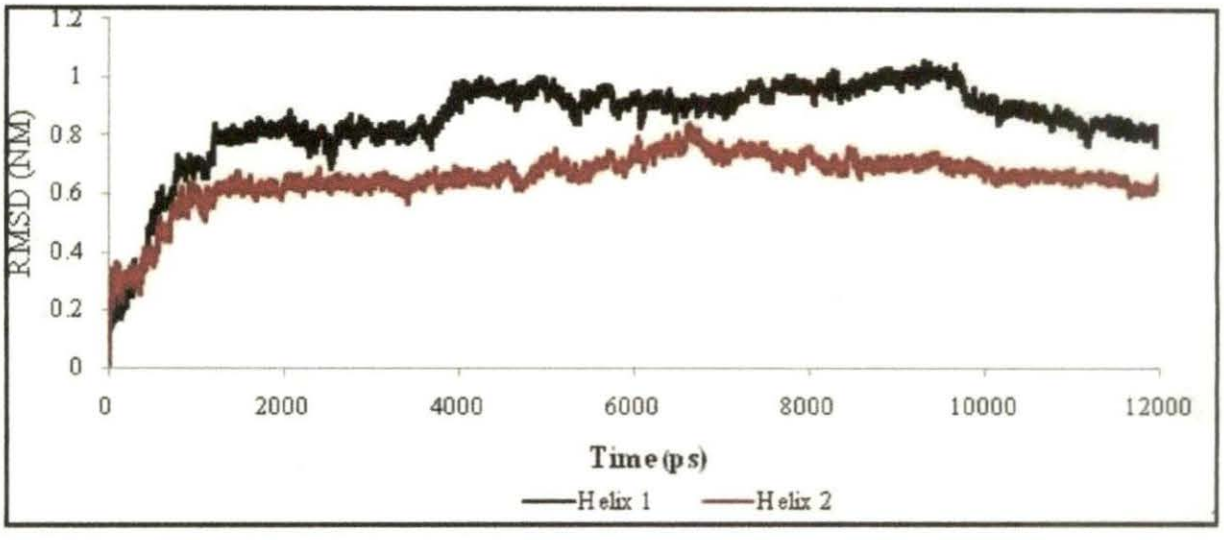



**Fig 4: Plot of time evaluation of RMSD values of Helix H1 and Helix H2**

A common approach in the identification of the major motions of a protein is the use of PCA [25, 26]. PCA reduces the dimensionality of a complex data set and applied to decompose a complex motion of proteins, which are characterized by an eigenvector and an eigenvalue. The eigenvalue for a given motion represents the contribution of the corresponding eigenvector to the global motion of the protein. PCA of the human bifunctional glutamyl-prolyl-tRNAsynthetase simulation reveals that the first 10 eigenvectors account for 88.67% of the global motion and that the first eigenvector

corresponds to 40.46% of the total motion, the second to 13.19%, and the third to a further 11.70% (Fig 5).

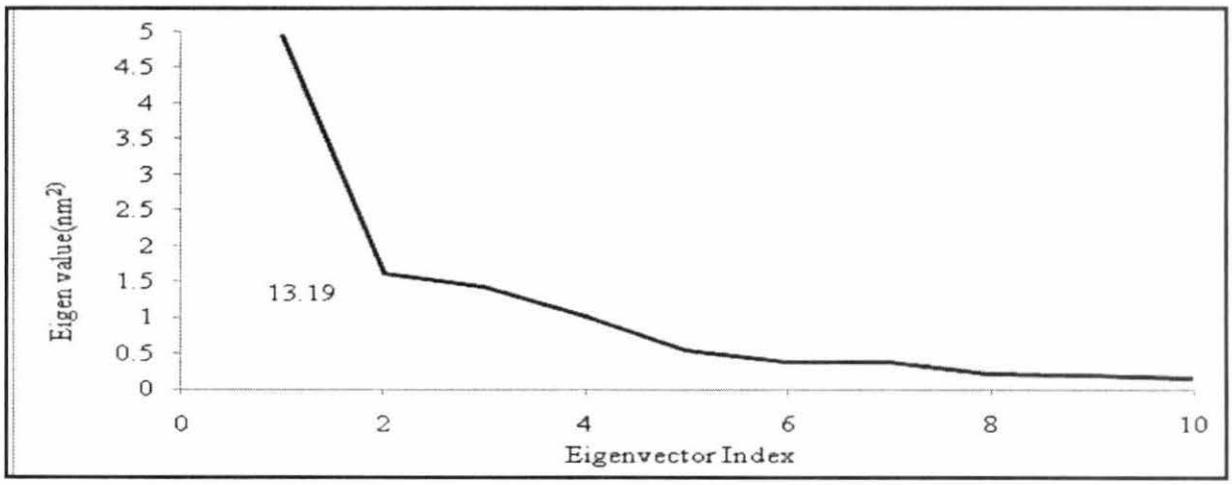



**Fig 5: Plot of Eigen values with Eigenvector indices**

During the simulation, several hydrogen bonds broke and formed. It is found that the number of hydrogen bonds ranged from 17 to 49 during simulation (Fig 6). The hydrogen bond network is weak, because there is a variation in number of hydrogen bonds support the molecule is flexible.

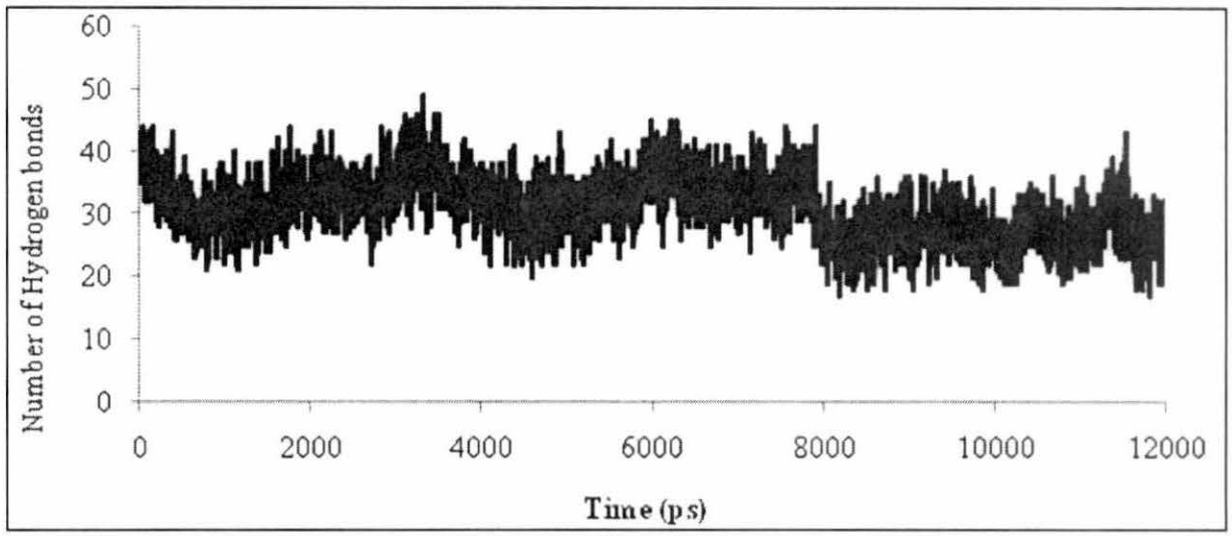



**Fig 6: Number of Hydrogen bonds during the whole simulation of Time**

The probability of sampling the phase space determined by first two principal modes during the simulations of the protein is presented in Figure-9. From this figure it is clear that the protein is sampling different conformational space during the simulation.

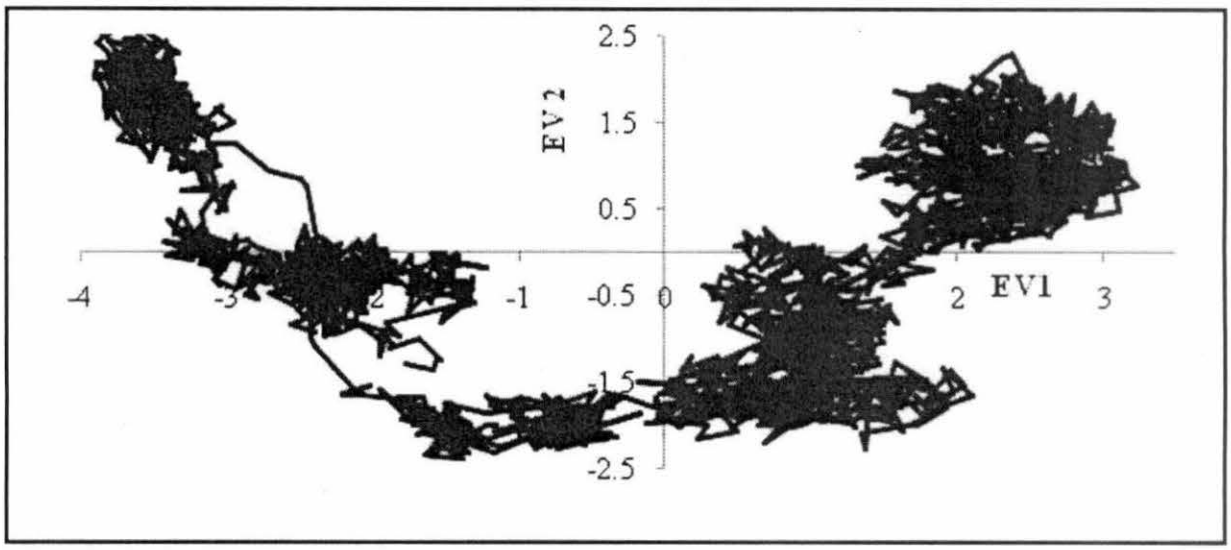



**Fig 9: Plot of Eigenvector 1 with Eigenvector 2**

We further examine the probability of accessing regions of the phase space determined PC1, PC2 & PC3 (Figure-10) and it is clear that the protein show very little arrangement in the XY plane and almost equally arrange in both the plane YZ and XZ.

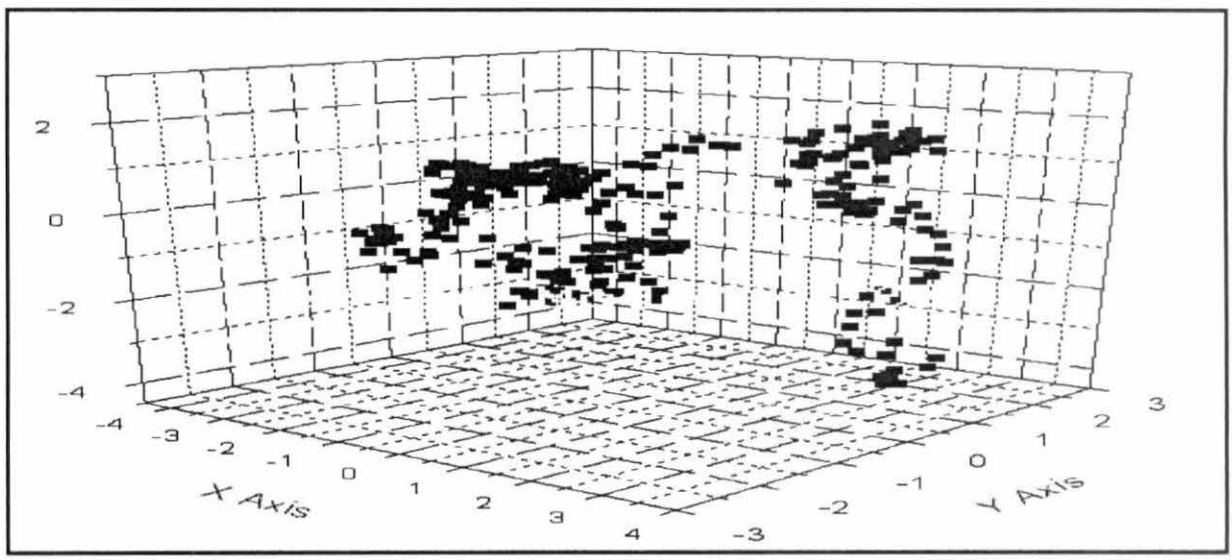



**Fig 10: Conformational sampling: The Probability of sampling the phase space determined by PC1, PC2, PC3**

We have taken some selected snapshots from dynamics trajectory considering time evolution of RMSD as a guideline. The snapshots were taken when the RMSD from initial structure was high, and they are presented in figure 11. From the snapshots, it is clear that there was major changes in the protein conformation which also support that the molecule is flexible.

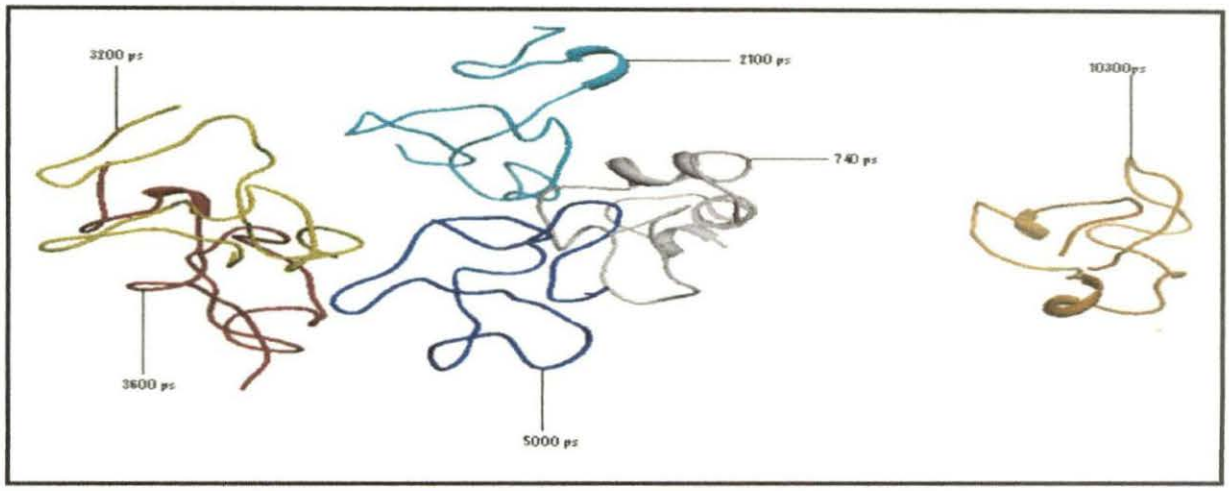



**Fig 11: Snapshots at different time**

From overall study it is found that during the dynamics, the structural variations, as measured by RMSD and the radius of gyration as a function time for the protein suggest that the protein is flexible.

From the probability of sampling the phase space determined by first two principal modes during the simulation the projection of the dynamics trajectory onto the first two PC that the protein traverses one conformational space around the origin and another at the right side of the origin and also at the left side of the origin which are much scattered, indicating high conformational freedom of the protein this is also revealed by different snapshots of the different structures extracted along the simulation trajectory in different times (Fig 11).It is also evident from Figure 9 that conformational freedom is more at the left side of the origin.

**7.4: References**


1. MirandeM (1991) Aminoacyl-tRNAsynthetase family from prokaryotes and eukaryotes: structural domains and their implications.*Prog. Nucleic Acids Res.Mol. Biol.*, **40**: 95-142.

2. DangCV, DangCV (1986)Multienzyme complex of aminoacyl-tRNAsynthetases: an essence of being eukaryotic *Biochem. J.*,**239**, 249-255.

3. KisselevLL,WolfsonAD (1994)The role of the anticodon in recognition of tRNA by aminoacyl-tRNAsynthetases. *Prog. Nucleic Acids Res. Mol. Biol.*, **48**, 83-142.

4. RhoSB, KimJ, LeeJS, MotegiM, KimH et al. (1999) Genetic dissection of protein-protein interactions in multi-tRNAsynthetase complex. *Proc. Natl. Acad. Sci.*U.S.A., **96**: 4488 – 93.

5. QuevillonS,MirandeM (1996)The p18 component of the multisynthetase complex shares a protein motif with the β and γ subunits of eukaryotic elongation factor *FEBS Lett.*, **395:** 63-67.

6. QuevillonS,Argou F, RobinsonJC,MirandeM (1997)The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine.*J. Biol. Chem.*, **272:** 32573-9.

7. M Mirande; D Le Corre; JP Waller (1985) A complex from cultured Chinese hamster ovary cells containing nine aminoacyl-tRNAsynthetases.*Eur. J. Biochem.*, **147**: 281-289.

8. Cerini C,Kerjan P,Astier M,Gratecos D, Mirande M, et al. (1991) A component of the multisynthetase complex is a multifunctional aminoacyl-tRNAsynthetase. *EMBO J.***10:** 4267–4277.

9. R Fett, R Knippers (1991)The Primary Structure of Human Glutarninyl-tRNASynthetase.*J.Biol. Chem.*, **266**, 1448-1455.

10. Heacock D, Forsyth CJ,Shiba K,K KMusier-Forsyth K (1996)Chemical Modification and Site-Directed Mutagenesis of the Single Cysteine in Motif 3 of Class II *Escherichia coli* Prolyl-tRNASynthetase. *Bioorg.Chem.*,**24** ,273-289.

11. NadaS, ChangPK,DignamJD **(1993)** Primary structure of the gene for glycyl-tRNAsynthetase from Bombyxmori.*J.Biol.Chem.*,**268**: 7660-7667.

12. ShibaK,SchimmelP,MotegiH, NodaT (1994) Human glycyl-tRNAsynthetase. Wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation. *J. Biol. Chem.*,**269**: 30049-30055.

13. FrolovaLY, Sudomoina MA,GrigorievaAY,ZinovievaOL,KisselevLL(1991)Cloning and nucleotide sequence of the structural gene encoding for human tryptophanyl-tRNAsynthetase.*Gene.*,**109**: 291-296.

14. LeeCC,CraigenWJ,MuznyDM, HarlowE; CaskeyCT (1990)Cloning and nucleotide sequence of the structural gene encoding for human tryptophanyl-tRNAsynthetase.*Proc. Natl. Acad. Sci.* U. S. A., **87**: 3508– 3512.

15. TsuiFW, SiminovitchL (1987) Isolation, structure and expression of mammalian genes for histidyl-tRNAsynthetas.*Nucleic Acida Res.*,**15**: 3349-3367.

16. HirakantM,SuwaA, TakedaY, MatsuokaY,IrimajiriS, et al. (1996) *J Craft. Arthritis Rheum.*,39, 146-151.

17. Craft J. *Arthritis Rheum.*,**1996**, 39, 146-151.

18. RabenN, NicholsR,DohlmanJ,McphieP, SridharV (1994) C Hyde; R Leff; P Plotz. A Motif in Human Histidyl-tRNASynthetase Which Is Shared among SeveralAminoacyl-tRNASynthetases Is a Coiled-coil That Is Essential for Enzymatic Activity and Contains the Major AutoantigenicEpitop. *J.Biol. Chem.*,**269**, 24277-24283.

19. RhoSB, LeeJS,JeongEJ,KimKS, KimYG, et al. (1998)A multifunctional repeated motif is present in human bifunctionaltRNAsynthetase.*J.Biol.Chem.*,**273:** 11267-11273.

20. JeongEJ, HwangGS, KimKH, KimMJ, KimS (2000) KS Kim.Structural Analysis of Multifunctional Peptide Motifs in Human BifunctionaltRNASynthetase: Identification of RNA-Binding Residues and Functional Implications for Tandem Repeats. *Biochemistry.***39**: 15775-15782.

21. Lindahl E, Hess B, et al. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**: 306-317.

22. Berendsen HJC, Postma JPM *et al.* (1984) Molecular dynamics with coupling to an external bath.*J. Chem. Phys* **81:** 3684–3690.

23. Essmann U, Perera L, et al. (1995) A smooth particle mesh Ewaldmethod. J. Chem. Phys., 103(19): 8577–8593.

24. AK Bothra; S Roy; B Bhattacharya; C Mukhopadhyay. *Journal of Biomolicular and Dynamics.***1998,** 15(5), 999-1008.

# Chapter VIII

# Screening of Triazine Derivatives, Inhibitors of MAP-kinase p-38 Alpha, Through Mathematical Modeling And Molecular Modeling

## 8.1: Introduction

The mitogen-activated protein (MAP) kinases comprise a family of serine/threonine kinases that function as critical mediators of signal transduction [1-3]. So far, four different MAP-kinases have been described: the extracellular signal-related kinases (ERKs), the c-Jun N-terminal kinases (JNKs), the p38 MAP-kinases and the ERK5 or big MAP kinase 1 (BMK1). The ERK MAP-kinases are preferentially activated by mitogens, whereas the JNK and p38 MAP-kinases are responsive to stress and inflammatory signals and it is also an important mediator of anti-estrogen growth inhibition in breast cancer [4].

Tamoxifen was shown to activate JNK and p38 MAP-kinases in connection with programmed cell death [5,6]. However, tamoxifen induced apoptosis occurs only at high concentrations and seems to be ER independent as it is not reversible by addition of estradiol [7] and it is also seen that breast tumor often becomes resistant to tamoxifen. This necessitates looking for other inhibitors for P38 MAP kinase for cancer therapy. Triazines might be a good candidate as it is reported to have antitumoral activities [8].Hexamethylmelamine (HMM), a derivative of triazine-A (Figure 1) possess various pharmacological actions against breast, lung and ovarian cancers, but once again causes nausea, vomiting, abdominal cramps, and anorexia. Compound B (Figure 1) was investigated by Moon et. al [9] as a microbial destabilizing agent entity with potent growth inhibition against U937cells(GI50=1μm). Leftheris et al reported compound C (Figure 1) as a potent inhibitor of p38 MAP kinase [10].

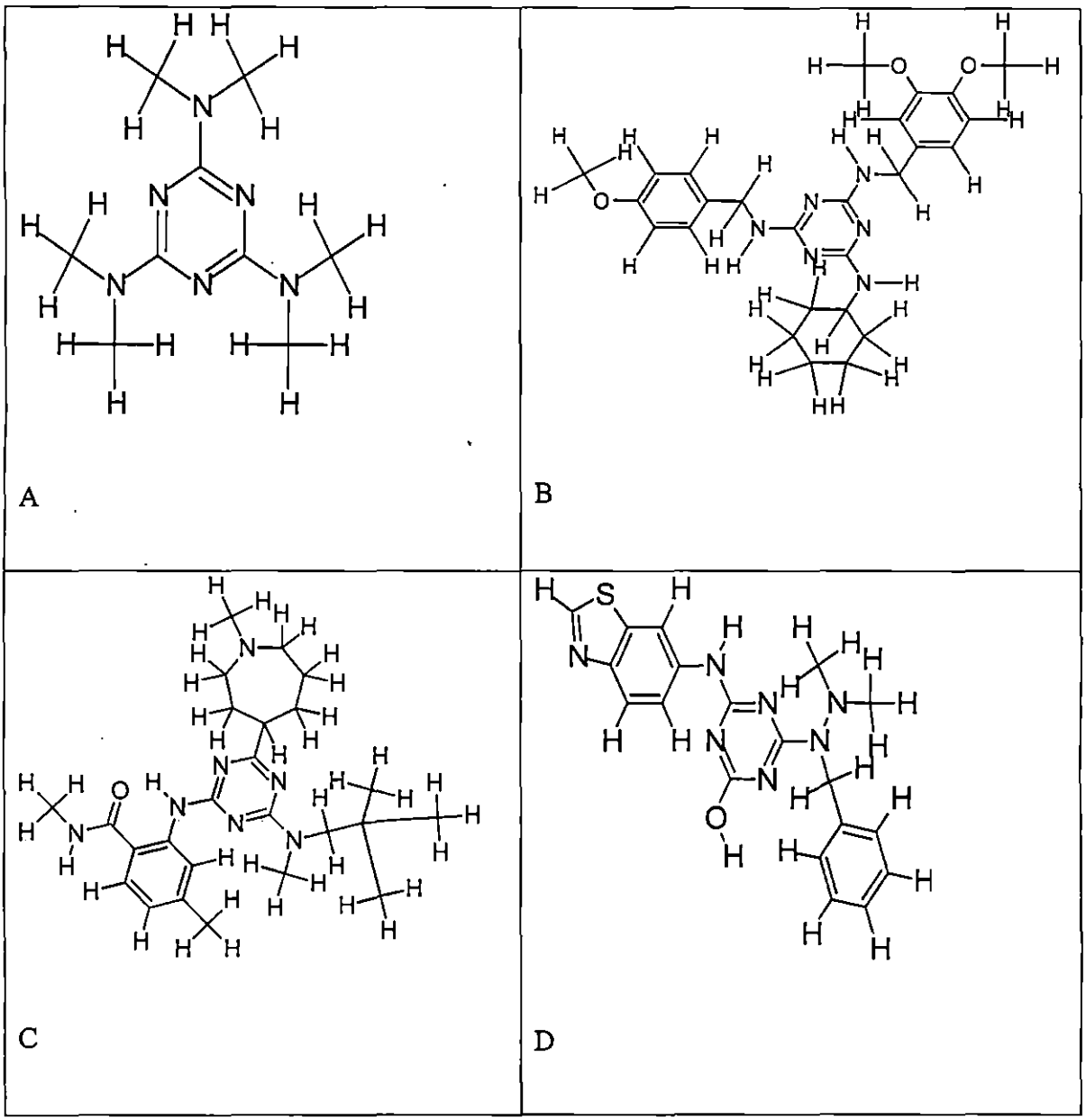



**Figure 1. Reported triaminotriazine compounds with antitumor activity**

Compound E (triaminotrizine derivatives) (Figure 2) was recently reported by "Baindur et al." [11] as a potent VEGF-R2 (KDR) tyrosine kinase inhibitor. It is a common feature for compounds to exhibit antitumor activities by introducing structural units of various aryl amino groups into the triazine scaffold. "Zheng et al." introduced structural units of various aryl amino groups in to the triazine scaffold and determined their biological activity [12]. Taking the experimental activity from the work of "Zheng et al." as dependent variable, we formulate a mathematical model, based on graph theoretical

indices, quantum chemical and structural parameters to design numbers of potent triazine scaffold-based inhibitor.

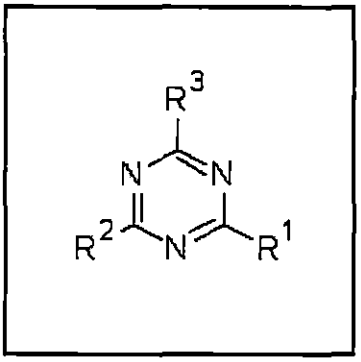



**Figure 2: Structural representation triaminotrizine derivatives**

An important advance in QSAR has been the application of a mathematical technique "Graph Theory" to chemistry [13]. In chemical graph theory [14], molecular structures are represented as hydrogen-suppressed graphs, commonly known as molecular graph, in which vertices and the bonds represent the atoms by edges. The connections between atoms can be described by topological matrices, which can be mathematically manipulated so as to derive a single number, known as topological index (TI) [15].

Topological indices offer a viable way of measuring molecular branching, shape and size [16]. Quantum chemical parameters like, HOMO & LUMO energy, dipole moment is also very helpful to develop the quantitative structure activity relationship.

In this communication we would like to consider different parameters and indices, namely IC (Mean Information Content), SIC (Structural Information Content), CIC (Complementary Information Content) and some quantum chemical parameters namely, HOMO, LUMO, Dipolemoment, Polarisability of the triazine derivatives [12] are calculated and a regression equation is formulated. Some chemically feasible compounds are designed and their theoretical activity is calculated and further binding energies of the

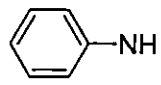
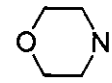
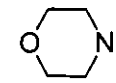
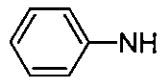
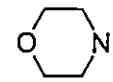
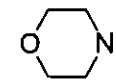
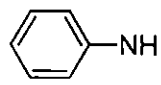
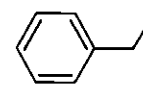
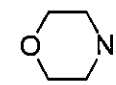
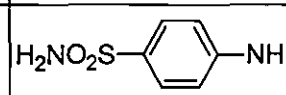
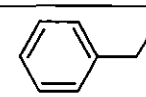
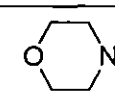
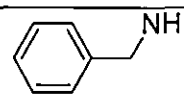
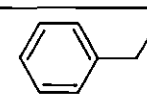
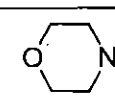
designed inhibitors were calculated through docking the inhibitors into the binding site of p-38 MAP kinase inhibitor. This will help us to screen potent inhibitors.

## 8.2: Materials and Methods

### Biological Data and Structures

The total set of 22 triazine derivatives was divided into 16 compounds in training set and six compounds in test set on the basis of their activity against HT-29 cells. Training set of triazine derivatives (1a -1p) are shown in Table 1. For validation of the trained set, we used a test set consisting of triazine derivatives (2a – 2f) are given in Table 2.

**Table 1. Chemical structures and inhibitory activities against HT-29 cells of triazine derivatives (Training set) by substituting R1, R2, R3 of E.**

| Compound | R1 | R2 | R3 | %Inhibition at μM[α] HT-29 |
|---|---|---|---|---|
| 1a |  |  |  | 80.5 |
| 1b |  |  |  | 74.2 |
| 1c |  |  |  | 90.5 |
| 1d |  |  |  | 68.6 |
| 1e |  |  |  | 50.0 |

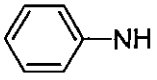
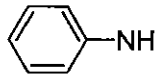
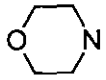
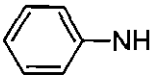
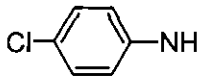
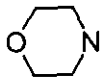
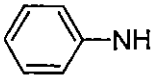
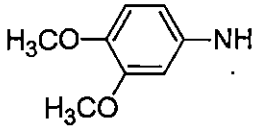
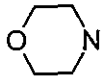
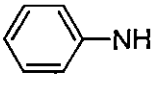
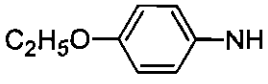
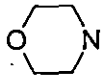
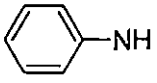
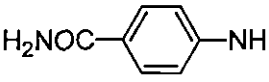
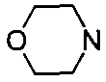
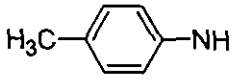
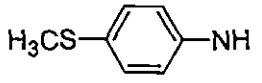
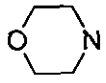
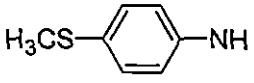
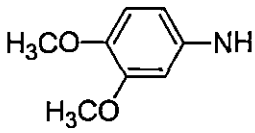
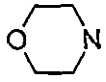
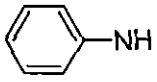
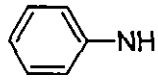
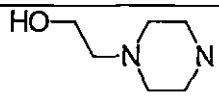
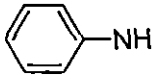
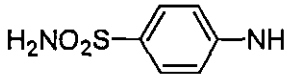
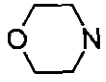
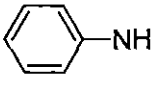
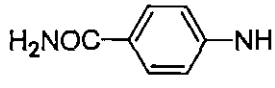
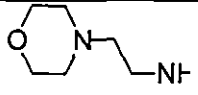
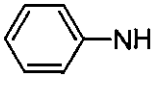
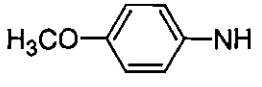
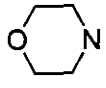
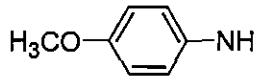
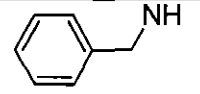
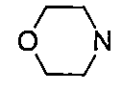
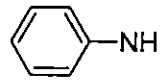
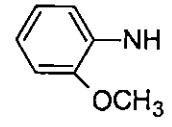
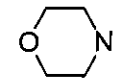
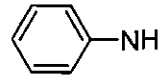
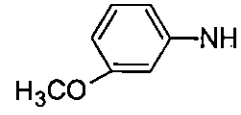
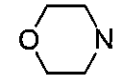
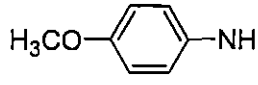
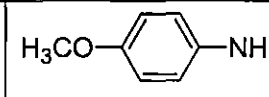
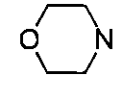
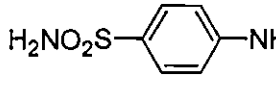
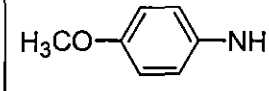
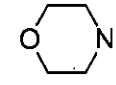
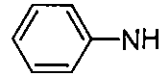
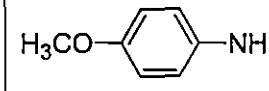
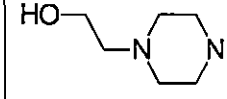
| Compound | R1 | R2 | R3 | %Inhibition at μM[α] HT-29 |
|---|---|---|---|---|
| 1f |  |  |  | 85.0 |
| 1g |  |  |  | 87.1 |
| 1h |  |  |  | 76.4 |
| 1i |  |  |  | 87.5 |
| 1j |  |  |  | 87.8 |
| 1k |  |  |  | 54.1 |
| 1l |  |  |  | 81.4 |
| 1m |  |  |  | 76.1 |
| 1n |  |  |  | 80.4 |
| 1o |  |  |  | 74.6 |
| 1p |  |  |  | 100.00 |

**Table 2.Chemical structures and inhibitory activities against HT-29 cells of triazine derivatives (Test set) by substituting R1, R2, R3 of E.**

| Compound | R1 | R2 | R3 | %Inhibition at μM[a] HT-29 |
|---|---|---|---|---|
| 2a |  |  |  | 88.4 |
| 2b |  |  |  | 89.3 |
| 2c |  |  |  | 87.4 |
| 2d |  |  |  | 80.6 |
| 2e |  |  |  | 83.2 |
| 2f |  |  |  | 77.3 |

**Topological Indices**

A major part of the current research in COMPUTATIONAL chemistry, chemical graph theory, and quantitative structure-activity/property relationship studies involves topological indices. Topological indices (TIs) are numerical graph invariants that quantitatively characterize molecular structure.

Information theoretic topological indices are calculated by the application of information theory to chemical graphs. An appropriate set of n-elements is derived from a molecular graph G depending upon certain structural characteristics. On the basis of a equivalence

relation defined on A1, the set A is partitioned into h disjoint subsets Ai of order (i=1, 2,... $h_i \sum_{i=1}^{n} n_i = n$). A probability distribution is then assigned to the set of equivalence classes.

$$A_1, A_2, A_3, \ldots\ldots, A_h$$

$$p_1, p_2 \ldots\ldots\ldots\ldots\ldots p_h$$

Where $p_i = \frac{n_i}{n}$ is the probability that a randomly selected element of A will occur in the ith subset.

$$W(G) = \sum_{i=1}^{N} \sum_{j=i}^{N} \mathbf{D}_{ij}(G) \tag{1}$$

The mean information content of an element of A is defined by Shanon's relation [17].

$$IC = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

The logarithm is taken at base 2 for measuring the information content in bits. The total information content of the set A is then n*IC.

In this method chemical species are symbolized by weighted linear graphs. Two vertices $u_0$ and $v_0$ of a molecular graph are said to be equivalent with respect to the rth order neighborhood if and only if corresponding to each path u1,u2,....ur of length r , there is a distinct path v1,v2,.....vr of the same length ,such that the paths have similar edge weights ,and both u0 and v0 are connected to the same number and type of atoms up to the rth order bonded neighbors.

Once partitioning of the vertex set for a particular order of neighborhood is completed, ICr is calculated from the appropriate equation. Basak, Roy and Ghosh defined another information theoretic measure, Structural Information Content (SIC ) [18], which is calculated as

$$SIC_r = \frac{IC_r}{\log_2 n} \tag{3}$$

Where IC is calculated from equation and n is the total number of vertices of the graph.

Another information-theoretic invariant, Complementary Information Content (CIC) (19Basak*et al.*, 1983), is defined as

$$CIC_r = \log_2 n - IC \tag{4}$$

CIC represents the difference between the maximum possible complexity of a graph (where each vertex belongs to a separate equivalence class) and the realized topological information of a chemical species as defined by ICr., HOMO, LUMO, Dipole moment, Polarisability are calculated using Mopac (20James *et al.*, 2007) and Arguslab [21].

We have taken x-ray structure of MAP kinase p-38 [22] from the protein data bank (PDB entry 1kv2) for calculating binding free energy of the designed molecules. We have performed docking by using Arguslab and obtained the binding free energy of training molecules and designed molecules.

**Preparation of protein**

We have taken x-ray crystal structure of the MAP kinase p-38 from the protein data bank (PDB ID: 1kv2) (http://www.pdb.org) [22]. The crystal structure of the complex of ligand with protein is available in Pdb. From this we get the binding site. Missing atoms were

repaired by the SPDBV software package and then it is minimized by steepest descent method.

**Preparation of ligand**

All triazine inhibitors used for docking study were collected from the published work of Zheng*et al.* []. Using draw mode of Chemsketch the ligand molecules were drawn and three dimensional optimizations were done and then saved in mol file. Geometry optimizations of the ligands were performed by ArgusLab 4.0.1 software. [21].

**Docking Studies**

Docking simulations were performed by selecting "ArgusDock" as the docking engine. The selected residues of the receptor were defined to be a part of the binding site. A spacing of 0.4 Å between the grid points was used and an extensive search was performed by enabling "High precision" option in Docking precision menu, "Dock" was chosen as the calculation type, "flexible" for the ligand and the AScore was used as the scoring function. The binding site box size was set to 60 x 60 x 60 Å, so as to encompass the entire active site. The AScore function, with the parameters read from the AScore.prm file, was used to calculate the binding energies of the resulting docked structures. The docking poses were saved for each compound according to their dock score function.

**Computer Software**

In this work different software are used like Chemsketch, Matlab-IV, Arguslab 4.01 and several program written by ourselves in Fortran-77 are used.

**8.3: Results and Discussion**

The values of graph theoretical indices, quantum chemical parameters and experimental HT-29 value of training set is given in Table 3 and test set is given in Table 4. Using the

parameters, multivariate regression analysis was performed with the training set and validated by test set.

**Table 3. Graph theoretical indices, quantum chemical parameters, experimental activity of training set.**

| Compound | Homo kcal/mol | Lumo kcal/mol | Dipole moment (debye) | SIC | CIC | Polarisibility ($cm^3$) | %Inhibition at $_{\mu}M^a$ HT-29 |
|---|---|---|---|---|---|---|---|
| 1a | -0.3628 | -0.0314 | 5.287 | 0.4891 | 2.8374 | 37.07 | 80.5 |
| 1b | -0.3528 | -0.0321 | 5.7053 | 0.5012 | 2.8295 | 39.72 | 74.2 |
| 1c | -0.3544 | -0.0262 | 3.4335 | 0.4914 | 2.8555 | 41.94 | 90.5 |
| 1d | -0.365 | -0.0714 | 10.717 | 0.5575 | 2.5464 | 45.99 | 68.6 |
| 1e | -0.3548 | -0.0283 | 3.697 | 0.4709 | 3.0161 | 43.81 | 50 |
| 1f | -0.3565 | -0.0251 | 3.2595 | 0.505 | 2.7345 | 40.07 | 85 |
| 1g | -0.3592 | -0.0295 | 3.4301 | 0.5374 | 2.5549 | 42.01 | 87.1 |
| 1h | -0.3445 | -0.0269 | 4.691 | 0.5161 | 2.7846 | 45.36 | 76.4 |
| 1i | -0.3504 | -0.0259 | 2.7284 | 0.503 | 2.8471 | 44.55 | 87.5 |
| 1j | -0.3368 | -0.0218 | 6.6385 | 0.5625 | 2.4692 | 43.61 | 87.8 |
| 1k | -0.2963 | -0.0272 | 2.8242 | 0.5335 | 2.672 | 46.07 | 54.1 |
| 1l | -0.2975 | -0.0296 | 3.5978 | 0.5351 | 2.7236 | 49.28 | 81.4 |
| 1m | -0.3248 | -0.0232 | 3.3403 | 0.4953 | 2.9044 | 45.15 | 76.1 |
| 1n | -0.3657 | -0.0607 | 6.7406 | 0.5661 | 2.4613 | 44.16 | 80.4 |
| 1o | -0.3156 | -0.0046 | 3.24 | 0.5067 | 2.8894 | 48.29 | 74.6 |
| 1p | -0.3547 | -0.0252 | 4.6082 | 0.5164 | 2.7295 | 42.72 | 100 |

**Table 4. Graph theoretical indices, quantum chemical parameters and experimental activity of test set.**

| Compound | Homo kcal/mol | Lumo kcal/mol | Dipole Moment (debye) | SIC | CIC | Polarisibility ($cm^3$) | %Inhibition at $\mu M^{\alpha}$ HT-29 |
|---|---|---|---|---|---|---|---|
| 2a | -0.3522 | -0.0266 | 3.9354 | 0.5095 | 2.8095 | 44.59 | 88.4 |
| 2b | -0.3542 | -0.0193 | 4.8844 | 0.5164 | 2.7295 | 42.72 | 89.3 |
| 2c | -0.3568 | -0.0254 | 4.8058 | 0.5164 | 2.7295 | 42.72 | 87.4 |
| 2d | -0.3455 | -0.0267 | 3.5659 | 0.5161 | 2.7846 | 45.36 | 80.6 |
| 2e | -0.326 | -0.0593 | 5.7084 | 0.5779 | 2.4404 | 46.68 | 83.2 |
| 2f | -0.3495 | -0.0405 | 5.3335 | 0.5171 | 2.8289 | 47.8 | 77.3 |

**PRESENTATION OF QSAR MODELS**

First, we have taken the SIC and CIC and constructed a regression model is given in equation 5. N is the number of compounds in training set. This equation gives moderate result with chi-square value 4.582513.

$$HT29 = 599.7415+(-480.2107)XSIC+(-99.4646)*CIC \quad (5)$$

$$N=16,\ \chi^2 = 4.582513$$

After that we took the HOMO and LUMO and constructed another regression model is given by equation 6.

$$HT29 =0.7417 +(-252.2386)XHOMO+(292.1922)LUMO \quad (6)$$

$$N=16,\ \chi^2 = 4.673094$$

These two models suggest that structural information and electronic information both are capable of providing predicted values which are not too bad. To improve the model we constructed regression model (equation 7) of comprising SIC, CIC, HOMO and LUMO which further improves the result. This is also evident from the $\chi^2$ test which shows better result.

**HT29=-160.62791 +(-0.3432) XHOMO+(-0.3057) XLUMO+(0. 51795) XSIC+ (2.740963)XCIC** (7)

N=16, $\chi^2$ = 2.851216

Lastly we inserted dipole moment and polarizability to construct a regression model with six parameters SIC, CIC, HOMO, LUMO, Dipole Moment and Polarisibility which gives very good result with chi-square value 0.99211. So the last model represented by equation 8 is very useful for screening these types of inhibitors.

HT29=-1932.1210+[(-0.5853)Homo+0.5539Lumo+(-0.0047)DM+2.3207SIC+0.3043CIC+ (-0.0045)Pol]*1000 (8)

N=16, $\chi^2$ = 0.99211

We have applied these models to several compounds designed by ourselves and predicted their activities. Predicted HT-29 values of training set using different models are given in table 5.

**Table 5:** Experimental and predicted value of the compounds under training set with different models.

| Compound | Experimental %Inhibition at μM[a] HT-29 | Predicted %Inhibition at μM[a] HT-29 (By equation 5) | Predicted %Inhibition at μM[a] HT-29 (By equation 6) | Predicted %Inhibition at μM[a] HT-29 (By equation 7) | Predicted %Inhibition at μM[a] HT-29 (By equation 8) |
|---|---|---|---|---|---|
| 1a | 80.5 | 82.64956 | 83.07904 | 77.41712 | 78.64453 |
| 1b | 74.2 | 77.62477 | 80.35212 | 76.31468 | 84.18933 |
| 1c | 90.5 | 79.74475 | 82.47963 | 78.15458 | 74.25037 |
| 1d | 68.6 | 78.74733 | 71.94627 | 70.93029 | 62.30176 |
| 1e | 50 | 73.61508 | 81.96691 | 70.90522 | 64.96387 |
| 1f | 85 | 85.24913 | 83.33075 | 83.73767 | 80.0625 |
| 1g | 87.1 | 87.55414 | 82.72613 | 91.0948 | 90.21241 |
| 1h | 76.4 | 74.9356 | 79.77793 | 80.65373 | 82.51416 |
| 1i | 87.5 | 75.00981 | 81.55833 | 79.90289 | 88.0083 |
| 1j | 87.8 | 84.02495 | 79.32588 | 94.48669 | 91.25818 |
| 1k | 54.1 | 77.77965 | 67.53237 | 66.8434 | 65.83154 |
| 1l | 81.4 | 71.87896 | 67.1338 | 65.9483 | 66.53845 |
| 1m | 76.1 | 73.0081 | 75.88994 | 69.55018 | 68.51123 |
| 1n | 80.4 | 83.08199 | 75.24929 | 80.87525 | 89.62232 |
| 1o | 74.6 | 69.02569 | 79.00412 | 81.37932 | 81.66174 |
| 1p | 100 | 80.27204 | 82.84749 | 86.00583 | 85.62427 |

Correlation graph of predicted values of HT-29 against experimental values of HT-29 of training set is given Figure 3. $R^2$ value of training set is 0.4181.

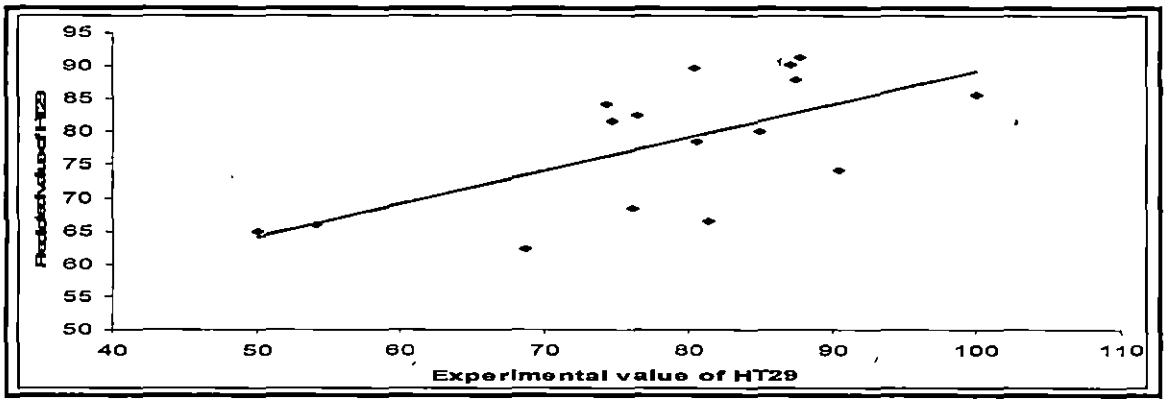



**Figure 3. Correlation graph between Theoretical and predicted activity of training set**

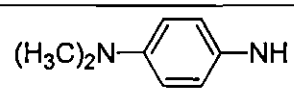
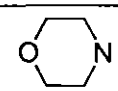
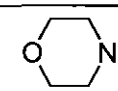
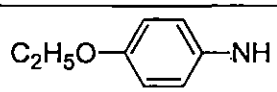
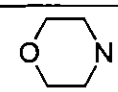
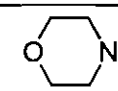
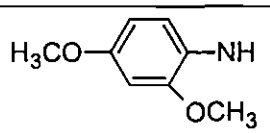
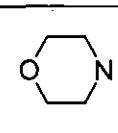
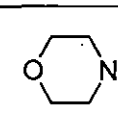
Predicted HT-29 values of test set using different models are given in table 6 and $R^2_{pred}$ for test set using equation 8 is 0.335.

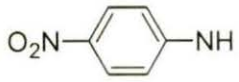
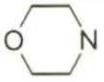
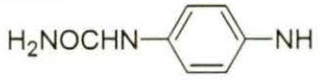
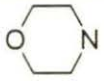
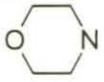
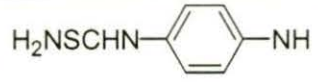
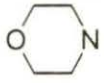
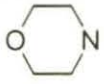
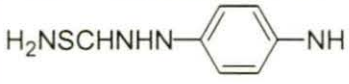
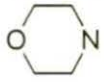
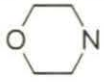
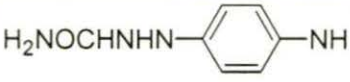
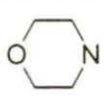
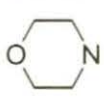
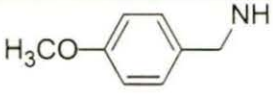
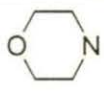
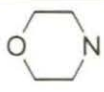
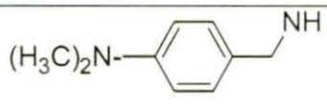
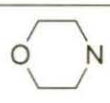
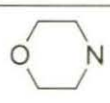
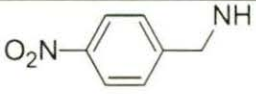
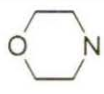
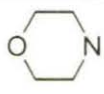
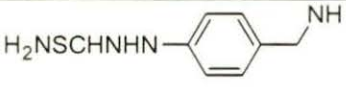
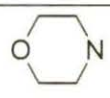
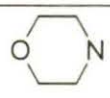
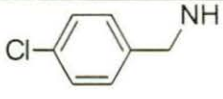
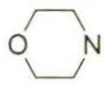
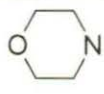
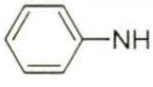
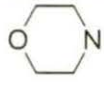
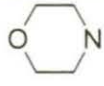
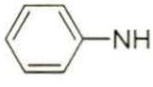
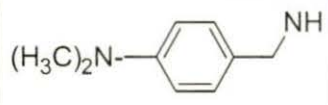
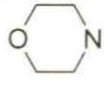
**Table 6:** Experimental and predicted value of the compounds under test set with different models.

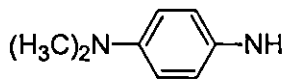
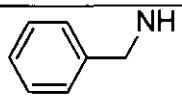
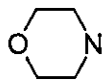
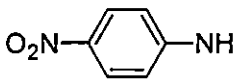
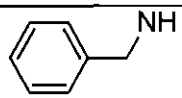
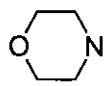
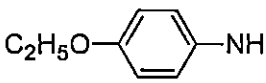
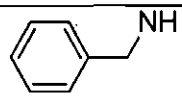
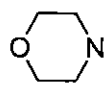
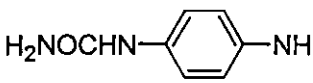
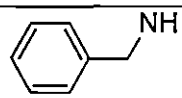
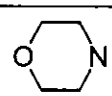
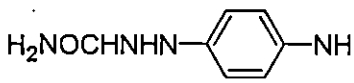
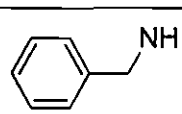
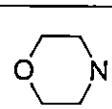
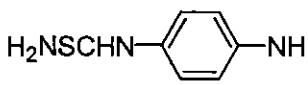
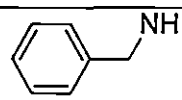
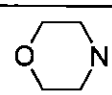
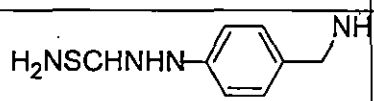
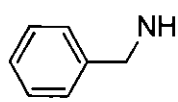
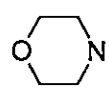
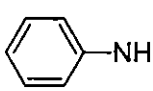
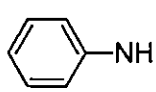
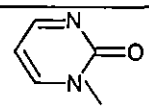
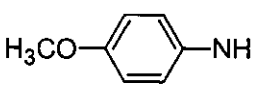
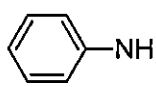
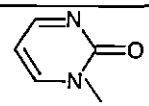
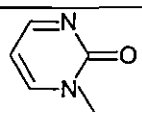
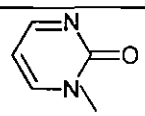
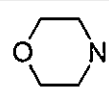
| compounds | Experimental %Inhibition at μM[a] HT-29 | Predicted %Inhibition at μM[a] HT-29 (By equation 5) | Predicted %Inhibition at μM[a] HT-29 (By equation 6) | Predicted %Inhibition at μM[a] HT-29 (By equation 7) | Predicted %Inhibition at μM[a] HT-29 (By equation 8) |
|---|---|---|---|---|---|
| 2a | 88.4 | 75.62831 | 81.80782 | 81.99198 | 86.46399 |
| 2b | 89.3 | 80.27204 | 84.4453 | 89.7197 | 87.30139 |
| 2c | 87.4 | 80.27204 | 83.31875 | 86.68493 | 85.81385 |
| 2d | 80.6 | 74.9356 | 80.0886 | 81.17263 | 88.49817 |
| 2e | 83.2 | 79.4943 | 65.64448 | 69.65482 | 81.69715 |
| 2f | 77.3 | 70.04911 | 77.06531 | 73.67898 | 79.70923 |

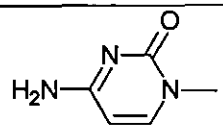
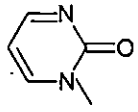
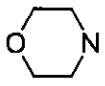
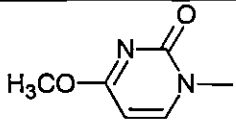
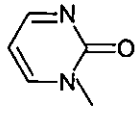
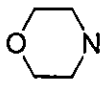
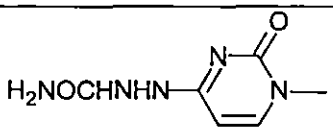
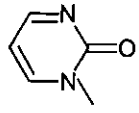
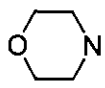
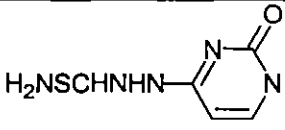
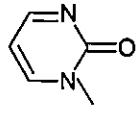
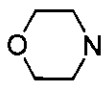
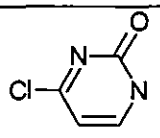
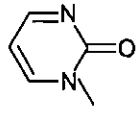
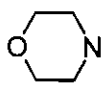
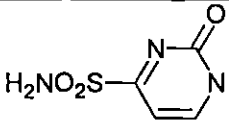
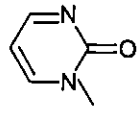
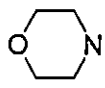
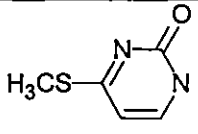
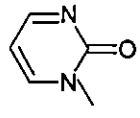
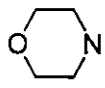
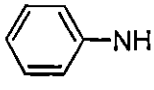
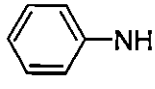
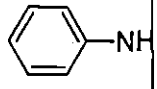
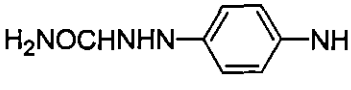
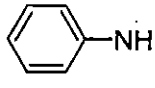
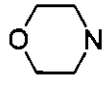
The structures of designed molecules with their predicted HT-29 values with the help of equation 8 are represented in Table 7. So these compounds may act as good inhibitors.

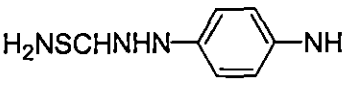
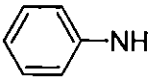
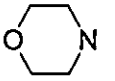
**Table 7. Chemical structures of designed triazine derivatives by substituting R1, R2, R3 of E and their predicted value of HT-29 by equation 8.**

| Compound | R1 | R2 | R3 | %Inhibition at μM[a]HT-29 Model 4 |
|---|---|---|---|---|
| 3a |  |  |  | 56.43 |
| 3b |  |  |  | 70.48 |
| 3c |  |  |  | 79.48 |

| Compound | R1 | R2 | R3 | %Inhibition at μM[α] HT-29 Model 4 |
|---|---|---|---|---|
| 3d |  |  |  | 73.77 |
| 3e |  |  |  | 74.16 |
| 3f |  |  |  | 70.02 |
| 3g |  |  |  | 98.27 |
| 3h |  |  |  | 87.47 |
| 3i |  |  |  | 75.85 |
| 3j |  |  |  | 78.89 |
| 3k |  |  |  | 67.38 |
| 3l |  |  |  | 61.50 |
| 3m |  |  |  | 81.31 |
| 3n |  |  |  | 71.16 |
| 3o |  |  |  | 57.00 |

| Compound | R1 | R2 | R3 | %Inhibition at μM[c] HT-29 Model 4 |
|---|---|---|---|---|
| | | | | |
| 3p |  |  |  | 54.37 |
| 3q |  |  |  | 21.79 |
| 3r |  |  |  | 66.64 |
| 3s |  |  |  | 83.64 |
| 3t |  |  |  | 75.79 |
| 3u |  |  |  | 55.64 |
| 3v |  |  |  | 60.48 |
| 3w |  |  |  | 73.62 |
| 3x |  |  |  | 74.8 |
| 3y |  |  |  | 61.22 |

| Compound | R1 | R2 | R3 | %Inhibition at μM[a] HT-29 Model 4 |
|---|---|---|---|---|
| 3z |  |  |  | 106.74 |
| 4a |  |  |  | 72.52 |
| 4b |  |  |  | 138.95 |
| 4c |  |  |  | 65.79 |
| 4d |  |  |  | 110.58 |
| 4e |  |  |  | 111.01 |
| 4f |  |  |  | 83.93 |
| 4g |  |  |  | 3.03 |
| 4h |  |  |  | 80.09 |

| Compound | R1 | R2 | R3 | %Inhibition at μM[α] HT-29 Model 4 |
|---|---|---|---|---|
| 4i |  |  |  | 29.22 |

Values of graph theoretical indices, quantum chemical parameters of the designed compounds (3a-3z, 4a-4i) are represented in Table 8 and their activities are (predicted by equation 8) is given in same table.

**Table 8. Graph theoretical indices, quantum chemical parameters and predicted activity of designed molecules.**

| Name of compound | Homo kcal/mol | Lumo kcal/mol | Dipole Moment (debye) | SIC | CIC | Polarisibility ($cm^3$) | Predicted %Inhibition at μM[α] HT-29 (By equation 8) |
|---|---|---|---|---|---|---|---|
| 3a | -0.324 | -0.0177 | 7.4078 | 0.4765 | 3.0268 | 42.74 | 56.43 |
| 3b | -0.3261 | -0.0189 | 6.8958 | 0.4941 | 2.9114 | 41.55 | 70.48 |
| 3c | -0.3336 | -0.0157 | 6.5859 | 0.4957 | 2.9157 | 42.36 | 79.48 |
| 3d | -0.3574 | -0.0587 | 5.3187 | 0.5146 | 2.7258 | 39.66 | 73.77 |
| 3e | -0.3145 | -0.0147 | 10.7156 | 0.5522 | 2.565 | 42.29 | 74.16 |
| 3f | -0.3139 | -0.0179 | 8.3862 | 0.5522 | 2.565 | 45.17 | 70.02 |
| 3g | -0.3067 | -0.0216 | 9.6827 | 0.6232 | 2.1781 | 46.6 | 98.27 |
| 3h | -0.3272 | -0.0181 | 10.3029 | 0.5611 | 2.5374 | 43.72 | 87.47 |
| 3i | -0.3531 | -0.0093 | 9.1175 | 0.4852 | 2.9624 | 41.59 | 75.85 |
| 3j | -0.3513 | -0.0109 | 6.7253 | 0.4688 | 3.1117 | 44.62 | 78.89 |
| 3k | -0.3681 | -0.0603 | 9.8262 | 0.5236 | 2.7159 | 41.54 | 67.38 |

| Name of compound | Homo kcal/mol | Lumo kcal/mol | Dipole Moment (debye) | SIC | CIC | Polarisibility ($cm^3$) | Predicted %Inhibition at $\mu M^{\alpha}$ HT-29 (By equation 8) |
|---|---|---|---|---|---|---|---|
| 3l | -0.3076 | -0.014 | 9.4549 | 0.5433 | 2.6752 | 48.47 | 61.50 |
| 3m | -0.3571 | -0.0125 | 7.8039 | 0.5066 | 2.7845 | 40.88 | 81.31 |
| 3n | -0.3123 | -0.0104 | 4.9425 | 0.5161 | 2.7717 | 44.59 | 71.16 |
| 3o | -0.312 | -0.0065 | 5.2811 | 0.486 | 2.9981 | 47.61 | 57.00 |
| 3p | -0.3098 | -0.0034 | 5.9317 | 0.486 | 2.9981 | 47.61 | 54.37 |
| 3q | -0.3429 | -0.0452 | 12.2131 | 0.5062 | 2.801 | 44.54 | 21.79 |
| 3r | -0.3122 | -0.0049 | 6.0474 | 0.5055 | 2.872 | 46.42 | 66.64 |
| 3s | -0.3122 | -0.0049 | 6.0474 | 0.5475 | 2.6185 | 47.16 | 83.64 |
| 3t | -0.3154 | -0.0063 | 9.1228 | 0.5578 | 2.5792 | 48.59 | 75.79 |
| 3u | -0.3105 | -0.0113 | 8.2818 | 0.5475 | 2.6185 | 50.04 | 55.64 |
| 3v | -0.282 | -0.0029 | 5.7743 | 0.5578 | 2.5778 | 51.47 | 60.48 |
| 3w | -0.3241 | -0.0366 | 5.7463 | 0.5771 | 2.2946 | 40.71 | 73.62 |
| 3x | -0.325 | -0.0373 | 6.9574 | 0.5804 | 2.3307 | 43.35 | 74.83 |
| 3y | -0.3714 | -0.0546 | 11.7836 | 0.5658 | 2.3106 | 36.35 | 61.22 |
| 3z | -0.3839 | -0.0667 | 11.4005 | 0.6204 | 2.0471 | 37.11 | 106.74 |
| 4a | -0.3815 | -0.0521 | 11.8090 | 0.5650 | 2.3939 | 38.66 | 72.52 |
| 4b | -0.3744 | -0.065 | 7.4158 | 0.647 | 1.9715 | 41.7 | 138.95 |
| 4c | -0.2563 | -0.0508 | 8.2841 | 0.6439 | 1.9889 | 43.01 | 65.79 |
| 4d | -0.3862 | -0.0611 | 9.9393 | 0.6283 | 1.9781 | 38.18 | 110.58 |
| 4e | -0.3953 | -0.1011 | 12.5408 | 0.6636 | 1.8472 | 40.98 | 111.01 |
| 4f | -0.3627 | -0.0614 | 9.6989 | 0.6044 | 2.1595 | 41.2 | 83.93 |
| 4g | -0.3174 | -0.0081 | 2.4132 | 0.4295 | 3.1329 | 43.07 | 3.03 |

| Name of compound | Homo kcal/mol | Lumo kcal/mol | Dipole Moment (debye) | SIC | CIC | Polarisibility ($cm^3$) | Predicted %Inhibition at $\mu M^{\alpha}$ HT-29 (By equation 8) |
|---|---|---|---|---|---|---|---|
| 4h | -0.3209 | -0.0135 | 8.2205 | 0.563 | 2.5146 | 46.72 | 80.09 |
| 4i | -0.229 | -0.0095 | 7.2959 | 0.5695 | 2.4776 | 48.38 | 29.22 |

We have calculated the binding free energy of the compounds 1a to 1p of training set by docking using the software Arguslab. Values of binding free energy of all the compounds under training set are given in table 9.

**Table 9.** Experimental value of HT-29 and binding free energy of training set.

| Compound | Experimental %Inhibition at □□□ HT-29 | Binding free energy in kcal/mol |
|---|---|---|
| 1a | 80.5 | -5.22 |
| 1b | 74.2 | -6.77 |
| 1c | 90.5 | -9.47 |
| 1d | 68.6 | -9.25 |
| 1e | 50 | -11.66 |
| 1f | 85 | -7.53 |
| 1g | 87.1 | -8.93 |
| 1h | 76.4 | -7.88 |
| 1i | 87.5 | -7.61 |
| 1j | 87.8 | -9.65 |
| 1k | 54.1 | -11.30 |
| 1l | 81.4 | -8.63 |
| 1m | 76.1 | -8.06 |

| | | |
|---|---|---|
| 1n | 80.4 | -9.70 |
| 1o | 74.6 | -7.91 |
| 1p | 100 | -7.94 |

We performed the docking of our designed molecules (3a-3z, 4a-4i) with the protein P38 map kinase ( PDB 1kv2) and calculated binding free energy. It was found that the binding energy lies between –7.3578 kcal/mol (3g) to –11.7924 kcal/mol (4d), which indicates that these are favorable for binding. The correlation between exp% inhibition and free energy of binding is shown in figure 4 and Value of $R^2$ is 0.2607. Low value of $R^2$ indicates that the activity does not depend fully on binding energy

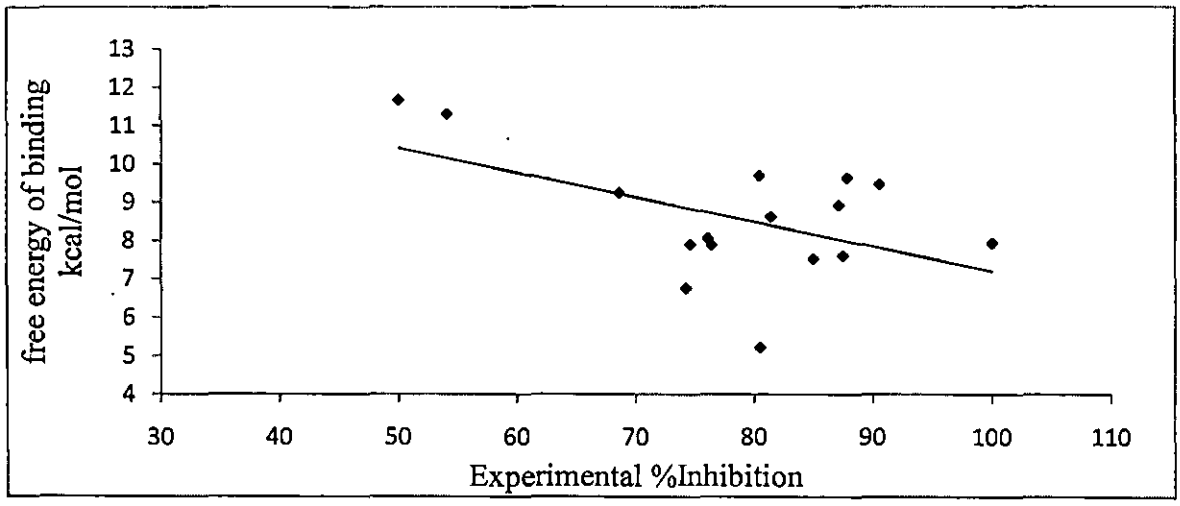



**Figure 4. Correlation graph between free energy of binding and Experimental % Inhibition of training set**

The values of binding free energy together with their predicted activity of designed set are given in Table 10.

**Table 10. Activity of designed molecules and their binding energy with inhibitors**

| Compound | **Predicted %Inhibition at μM[a] HT-29 (Model 4)** | **Binding Energy (kcal/mol)** |
|---|---|---|
| 3a | 56.43 | -7.73 |

| Compound | **Predicted %Inhibition at μM[α] HT-29 (Model 4)** | **Binding Energy (kcal/mol)** |
|---|---|---|
| 3b | 70.49 | -7.91 |
| 3c | 79.48 | -6.74 |
| 3d | 73.78 | -7.74 |
| 3e | 74.15 | -7.47 |
| 3f | 70.01 | -7.65 |
| 3g | 98.27 | -7.36 |
| 3h | 87.48 | -7.75 |
| 3i | 75.85 | -7.81 |
| 3j | 78.89 | -8.24 |
| 3k | 67.38 | -8.17 |
| 3l | 61.51 | -7.52 |
| 3m | 81.32 | -8.55 |
| 3n | 71.16 | -9.70 |
| 3o | 57.01 | -9.46 |
| 3p | 54.38 | -9.63 |
| 3q | 21.79 | -9.16 |
| 3r | 66.65 | -9.24 |
| 3s | 83.65 | -9.24 |
| 3t | 75.80 | -9.75 |
| 3u | 55.64 | -10.61 |
| 3v | 60.48 | -10.01 |
| 3w | 73.62 | -10.61 |
| 3x | 74.83 | -9.39 |
| 3y | 61.23 | -7.75 |
| 3z | 106.75 | -8.57 |
| 4a | 71.52 | -7.65 |

| Compound | **Predicted %Inhibition at μM[α] HT-29 (Model 4)** | **Binding Energy (kcal/mol)** |
|---|---|---|
| 4b | 138.96 | -8.56 |
| 4c | 65.79 | -8.47 |
| 4d | 110.59 | -8.26 |
| 4e | 111.02 | -5.69 |
| 4f | 83.94 | -7.86 |
| 4g | 30.36 | -11.79 |
| 4h | 80.09 | -9.50 |
| 4i | 29.22 | -9.59 |

The snapshot of docked compound 4a with the receptor protein is presented in Figure 5.

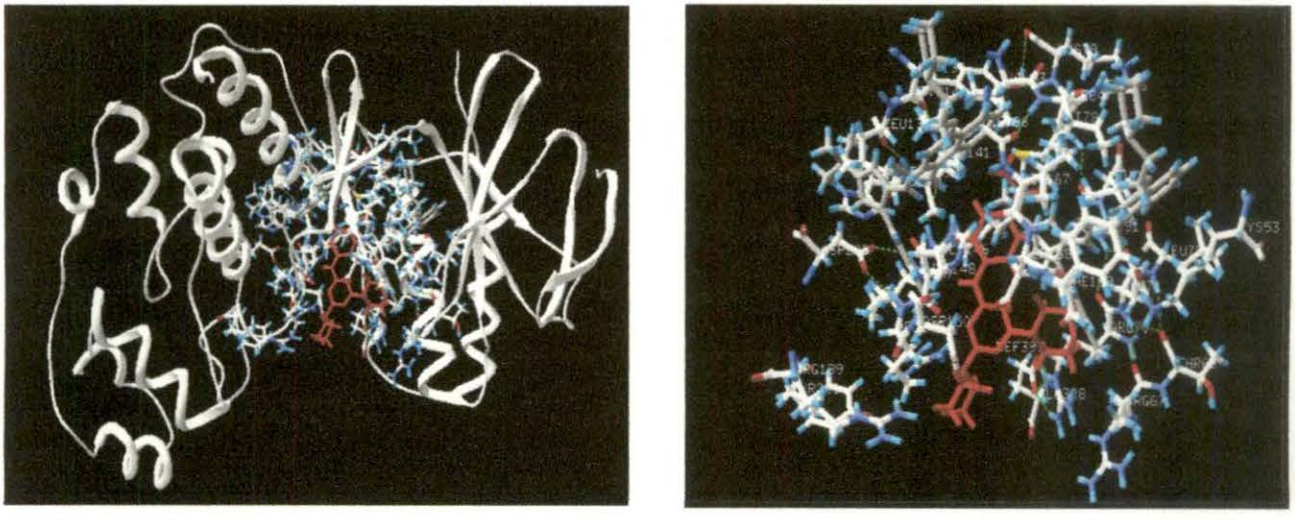


**Figure 5. (A) Ribbon structure of p38 MAP kinase docked with 2,*N*,*N*-dimethyl-*N*' bis(4,6-morpholin-4-yl-1,3,5-triazin-2-yl)benzene-1,4-diamine; (B) protein with side chain with docked ligand**

The triazine derivatives and their inhibition activities to HT-29 cell are depicted in Table 5. Tris (N-morpholino)-1,3,5-triazine have no inhibition activities to HT-29 cells(12). The

mono (N-morpholino) -1,3,5-triazines exhibited higher inhibitory acvtivities against HT-29 cells than their corresponding bis (N-morpholino)-1,3,5 triazine derivatives.

It was seen from our study that the compounds 3c, 3g, 3h, 3m, 3s, 4b, 4f, 4d, 4h, 4e show very high inhibition activities as predicted by our regression equation 8, whereas the compounds 3a, 3o, 3p, 3q, 3u, 4g, 4i exhibit low inhibitory activities. Rest of the molecules in Table 10 has moderate inhibition activities. Thus, from our study it seems that the compounds 3c,3g,3h,3m,3s,4b,4f,4d,4h,4e could be candidates for good inhibitors.

It is observed that all the compounds which shows high activity has at least one morpolino ring and absence of morpholino ring reduces activity value. We designed tris (anilino)-1, 3, 5-triazine (4g) and calculated their activity using our regression model, the predicted value is very low. Therefore presence of one morpholino unit plays significant role in the inhibition activities of the HT-29 cells and it is observed that the presence of three morphilino groups reduces the activity markedly. From our study it was seen that presence of three anilino group also reduces the activity markedly (4g) Replacement of one anilino group by one morpholino group enhances the activity (1p).

It may be concluded that presence of morpholino/anilino ring is essential for the activity of triazinederivates which are MAP kinase inhibitor and insertion of pyrimidine ring in triazine scaffold gives very high predicted activity as evident by compounds 4b, 4d, 4e, 4f, 3z ( Table 8).

**8.4: References**

1. Cohen P (1997) The search for physiological substrates of MAP and SAP kinases in mammalian cells. *Trends. Cell Biol. 7*: 353–361.

2. Ip YT and Davis RJ (1998) Signal transduction by the c-Jun N-terminal kinase.(JNK): from inflammation to *development. Curr. Opin. Cell Biol.*, **10**: 205–219.

3. Davis RJ (1998) Signal transduction by the c-Jun N-terminal kinase (JNK): from inflammation to development. *Curr. Opin. Cell Biol.***10**: 205–219.

4. Kyriakis JM, Avruch J (2001) Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol Rev.***81**: 807-69

5. Obrero M, Yu DV, Shapiro DJ (2002) Estrogen receptor-dependent and estrogen receptor-independent pathways for tamoxifen and 4-hydroxytamoxifen-induced programmed cell death. *J Biol Chem.,277***:** 4569-5703

6. Mandlekar S, Yu R, Tan TH, Kong AN (2000) Activation of caspase-3 and c-Jun NH2-terminal kinase-1 signaling pathways in tamoxifen-induced apoptosis of human breastcancer cells. *Cancer Res.,***60**: 5995-6000.

7. Mandlekar S, Kong AN (2001) Mechanisms of tamoxifen-induced apoptosis. *Apoptosis.,6*, 469-77.

8. Wolf DM, Jordan VC, William L (1993) McGuire Memorial Symposium. Drug resistance to tamoxifen during breast cancer therapy. *Breast Cancer Res Treat,***27:** 27-40.

9. Moon HS, Jacobson EM, Khersonsky SM, Luzung MR, Walsh DP (2002) *J.Am. Chem. Soc.,124*, 11608.

10. Leftheris K, Ahamed G, Chan R, Dyckman AJ, Hussain Z (2004) The discovery of orally active triaminotriazine aniline amides as inhibitors of p38 MAP kinase.*J. Med. Chem.*, *47*: 6283.

11. Baindur N, Chanda N, Brandt BM, Asgari D, Patch RJ (2005) *J. Med. Chem.*,*48*: 1717.

12. Mingfang Zheng, ChenghuiXu, Jianwei Ma, Yan Sun, Feifei Du, Hong Liu, Liping Lin, Chun Li, Jian Ding, Kaixian Chen and Huliang Jiang : Synthesis and antitumor evaluation of a novel series of triaminotriazine derivatives : *Bioorganic & Medicinal Chemistry* ,**2007**,*15*, 1815-1827.

13. Livingstone DJ (2000)The characterization of chemical structures using molecular properties. A survey.*J ChemInfComput Sci.*, **40**:195-209.

14. Ivanciuc O (2003) Graph Theory in Chemistry. In: *Handbook of Chemoinformatics*, Ed.: J. Gasteiger. Wiley-VCH. 103-138.

15. Ivanciuc O (2003) Topological Indices. In: *Handbook of Chemoinformatics*, Ed.: J.Gasteiger. Wiley-VCH. 981-1003.

16. Ivanciuc O, Balaban AT (1999) In Topological Indices and Related Descriptirs in QSAR and QPSR; J. Devillers and A.T. Balaban, Eds.; *Gordon and Breach Science Publishers*. The Netherlands. 59-167.

17. Shanon CE (1948) A mathematical theory of Communication. *Bell Syst. Tech.J.*, *27*: 379-423.

18. Basak SC, Roy AB,Ghosh JJ (1979) Study of Structure-Function Relationship of Pharmacological and Toxicological Agents Using Information Theory. *In*

*proceeding of the llnd International conference on mathematical modeling, University of Missouri at Rolla.* 851-856.

19. Basak SC and Magnuson VR (1983) Molecular Topology and Narcosis. A Quantitative Structure-Activity Relationship (QSAR) Study of Alcohols Usins Complementary Information Content, Arzneimittle-Froschung, *Drug Research.* 501-503.

20. MOPAC2007, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, http://Open MOPAC.net, 2007.

21. http//www.seanet.com/~mthompson/Arguslab.

22. Pargellis C, Gilmore L, Graham AG, Grob PM, Hickey ER (2002)Inhibition of P38 Map-Kinase by utilizing a Novel Allosteric Binding State. *J Regan Nat. Struct. Biol.*, **9**: 268-272.

# Chapter IX

# *Molecular Docking of Triazine analogues*

## 9.1: Introduction

Mitogen activated protein kinases (MAPK) are group of serine/threonine protein kinases that play an important role in signal transduction in many cellular process including growth, differentiation, cell death, and survival [1-5]. There are at least six different groups of MAPK that have been identified in humans cells: the extracellular signal-regulated protein kinases (ERK1, ERK2); c-Jun N-terminal kinases (JNK1, JNK2, JNK3) p38s (p38a, p38b, p38g, p38d); ERK5; ERK3s (ERK3, p97 MAPK, MAPK4); ERK7s (ERK7, ERK8) [6-8]. Each group of MAPKs can be simulated by a separate signal transduction pathway in response to different extracellular stimuli. MAP kinase plays a fundamental role to generate several diseases, such as asthma, osteoarthritis, rheumatoid arthritis, and chronic inflammatory autoimmune disease. The inhibition of MAP kinase would potentially prevent the underlying pathophysiology in the inflammatory diseases [9-11]. Rheumatoid arthritis causes damage of cartilage and deformation of bones [12]. The injury inflammation caused by inflammatory mediators such as Tumor necrosis factor-α and Interlucin-1β [13]. Biosynthesis of these two proinflammatorycytokins regulated by p38 and these two cytokins are associated with the progression of rheumatoid arthritis [14].

Triazine derivatives show wide spectrum of biological activities in antimicrobial effect, Erm (erythromycin-resistance methylase) methyl transferase inhibition, anti-trypanosomal activity, VLA-4 (integrain very late antigen-4) antagonism, estrogen receptor modulation, cytotoxic activity [15-17]. Hexamethyl melamine, a triazine derivative possesses various pharmacological actions against brest, lung and ovarian cancers, severe adverse effect nausea, vomiting, abdominal cramps, and anorexia [18].

Leftheris et al. reported triaminotriazineanline amide as potent inhibitor of p38 MAPkinase [19]. Zheng et al introduced various aryl amino groups in to the triazine scaffold and determined their biological activity [20]. Taking the experimental activity from the work of "Zheng et al." as dependent variable, we formulate a mathematical model, based on binding energy to design numbers of potent triazine scaffold-based inhibitor (Figure 1)

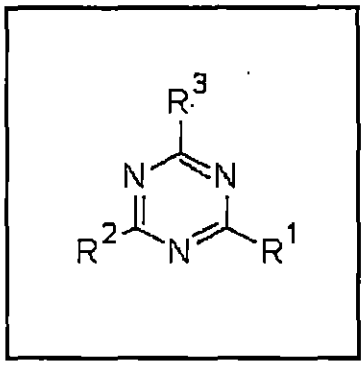
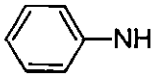
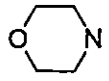
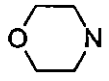
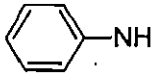
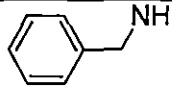
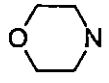


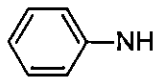
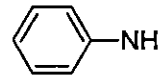
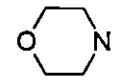
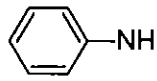
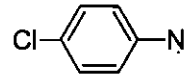
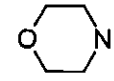
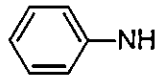
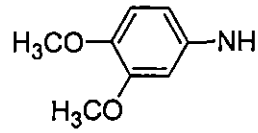
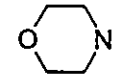
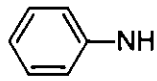
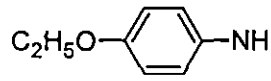
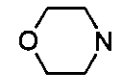
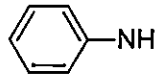
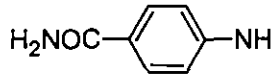
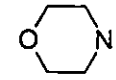
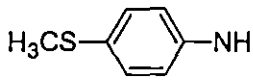
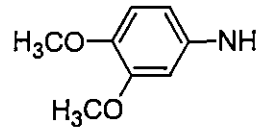
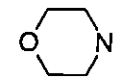
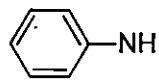
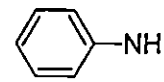
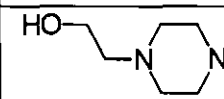
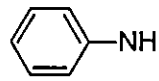
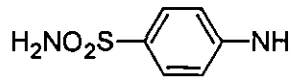
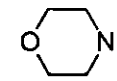
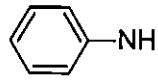
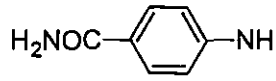
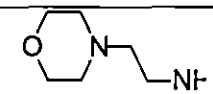
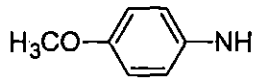
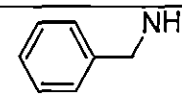
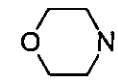
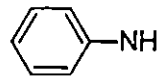
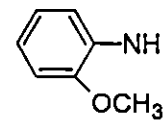
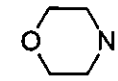
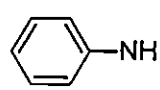
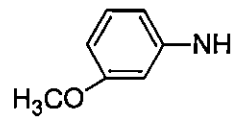
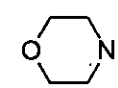
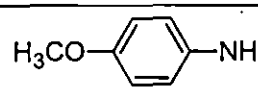
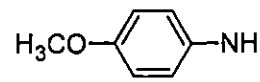
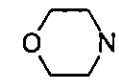

**Figure 1**

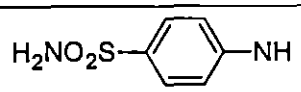
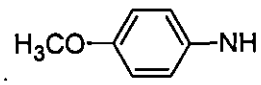
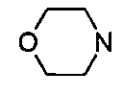
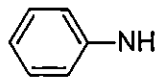
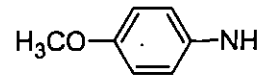
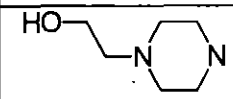
**Structural representation triaminotrizine derivatives**

Triazine derivatives and their corresponding inhibitory activities against HT-29 cells are shown in Table 1

**Table1: Chemical structures and inhibitory activities against HT-29 cells of triazine derivatives by substituting R1, R2, R3 of T**

| Com | R1 | R2 | R3 | %Inhibition at $\mu M^{\alpha}$ln(HT29) |
|---|---|---|---|---|
| 1 |  |  |  | 4.3883 |
| 2 |  |  |  | 4.5053 |

| Com | R1 | R2 | R3 | %Inhibition at $\mu M^{\alpha}$ln(HT29) |
|---|---|---|---|---|
| 3 |  |  |  | 4.4427 |
| 4 |  |  |  | 4.4671 |
| 5 |  |  |  | 4.336 |
| 6 |  |  |  | 4.4716 |
| 7 |  |  |  | 4.4751 |
| 8 |  |  |  | 4.3994 |
| 9 |  |  |  | 4.332 |
| 10 |  |  |  | 4.387 |
| 11 |  |  |  | 4.3121 |
| 12 |  |  |  | 4.4819 |
| 13 |  |  |  | 4.492 |
| 14 |  |  |  | 4.4705 |
| 15 |  |  |  | 4.3895 |

| Com | R1 | R2 | R3 | %Inhibition at $\mu M^{\alpha}$ln(HT29) |
|---|---|---|---|---|
| 16 |  |  |  | 4.4212 |
| 17 |  |  |  | 4.3477 |

*ln(HT29) – Natural logarithm of HT29

**9.2: Materials and Methods**

**Experimental Section**

**Preparation of inhibitor**

Triazine inhibitors used for docking study were collected from the published work of Zheng *et al.* [20]. Using draw mode of Chemsketch the ligand molecules were drawn and three dimensional optimizations were done and then saved in mol file. For docking experiments with AutoDock 4.2, ligand molecules were optimized, and saved as in pdb format with the aid of Arguslab 4.2 [21]. Next ligand is loaded in AutoDock Tool (ADT). Gasteiger charges are added. ADT selected a root with the minimum number of rotatable branches. Root was detected. Next ligand was saved in PDBQT format. Then prepared ligand was used in docking simulation in the next step.

**Preparation of protein**

In the present article crystallographic structure of P38 MAP kinase was downloaded from the Protein Data Bank as PDB file (PDB entry code 1kv2) [22]. Missing atoms were repaired to the crystallographic structure of the SPDBV software package [21]. For docking simulation using Autodock all polar hydrogen was added with the GROMACS modeling package. The resulting structure was optimized by the GROMACS force field.

During minimization, the heavy atoms were kept fixed at their initial crystal coordinates by restraing. Minimization was carried out under a vacuum medium. Electrostatic interactions were calculated using the cut-off method. Finally, solvation parameters were added using the ADDSOL utility of AutoDock 4.2. Default values of atomic solvation parameters were used throughout the calculations. The grid maps of the protein were used in the docking experiments was calculated using the AutoGrid utility program.

**Docking:** Docking studies were performed with AutoDock4.2 using Lamarckian genetic algorithm [23]. The flexible docking procedure was used for a P38 MAP kinase protein and a flexible ligand. A grid of 94, 78, 62 ponts in x,y, and z direction was constructed. A grid spacing of 0.375Å and distance- dependent function of the dielectric constant was used for the calculation of the energy map. The defaults settings were used for all other parameters. The entire calculations were carried out on PC based machines running Linux as operating system. At the end of docking, ligand with most favorable free energy of binding were noted. The best conformer was chosen based on the lowest free energy of binding. The protein with the best conformer is saved as complex and analyzed using Molegro Molecular Viewer and PyMOL [24, 25].

**9.3: Results and Discussion**

Molecules were successfully docked on to the active site of P38 MAP according to the above docking analysis. Results of the docking experiments we calculated free energy of binding for each complex with triazine analogues and P38 MAP kinase are shown in Table 2.

Total 17 compounds were used for regression analysis. A regression equation was performed using one index, free energy of biding.

$$\ln HT29 = 4.507487 + (0.0090) bindingenergy$$

By this equation we calculated predicted ln HT29 activity (Table2).

**Table 2:** Actual and predicted activities of the training set molecules

| Compound | Free energy of Binding (kcal/mol) | % Inhibition at μM$^{\alpha}$ln(HT-29) | Predicted %Inhibition at μM$^{\alpha}$ln(HT-29) |
|---|---|---|---|
| 1 | -8.34 | 4.3883 | 4.4288 |
| 2 | -9.03 | 4.5053 | 4.4224 |
| 3 | -8.67 | 4.4427 | 4.4258 |
| 4 | -8.77 | 4.4671 | 4.4248 |
| 5 | -8.9 | 4.336 | 4.4236 |
| 6 | -6.95 | 4.4716 | 4.4418 |
| 7 | -8.18 | 4.4751 | 4.4303 |
| 8 | -8.67 | 4.3994 | 4.4258 |
| 9 | -19 | 4.332 | 4.3297 |
| 10 | -8.88 | 4.387 | 4.4238 |
| 11 | -9.74 | 4.3121 | 4.4158 |
| 12 | -9.44 | 4.4819 | 4.4186 |
| 13 | -9.32 | 4.492 | 4.41972 |
| 14 | -9.15 | 4.4705 | 4.4213 |
| 15 | -9.11 | 4.3895 | 4.4217 |
| 16 | -9.66 | 4.4212 | 4.4166 |
| 17 | -8.33 | 4.3477 | 4.4289 |

Docking results shows all triazine inhibitors docked in to allosteric of P38 MAP kinase and their calculated free energy of binding is shown in table1. From the table it is shown that lowest binding energy value is obtained for compound 11. Docking structure of compound11 is presented in Figure 2. A repressive figure containing the inhibitor (11) in the binding site of P38 MAP kinase is presented in Figure 3. From the Figure 3, it is clear that the inhibitor is well inside binding cavity.

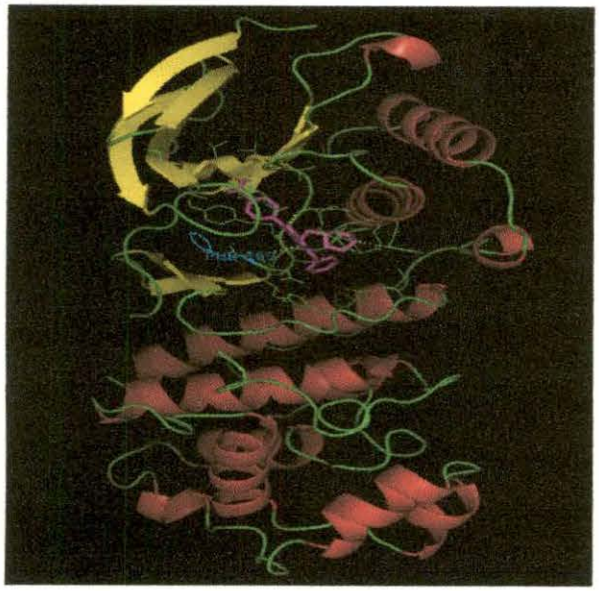


**Figure 2: Compound 11 in a binding pocket of P38 MAP. The inhibitor show in stick model (magenta)**

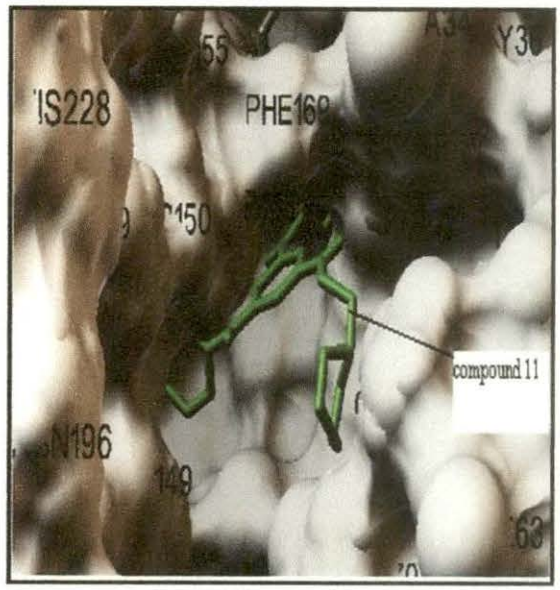



**Figure 3: Docking result of triazine inhibitor (1l)**

Figure 4 shows ligand was surrounded by both polar and apolar hydrophobic group such as ARG67, ARG70, GLU71, LYS53, VAL38, GLU71, PHE169, ASP168, HIS148, LEU7, LEU75, ILE84, ILE166, ILE41. These amino acid residues stabilized both polar and apolar parts of ligands. Figure shows the four hydrogen bond interaction observed between inhibitor and P38 MAP kinase.

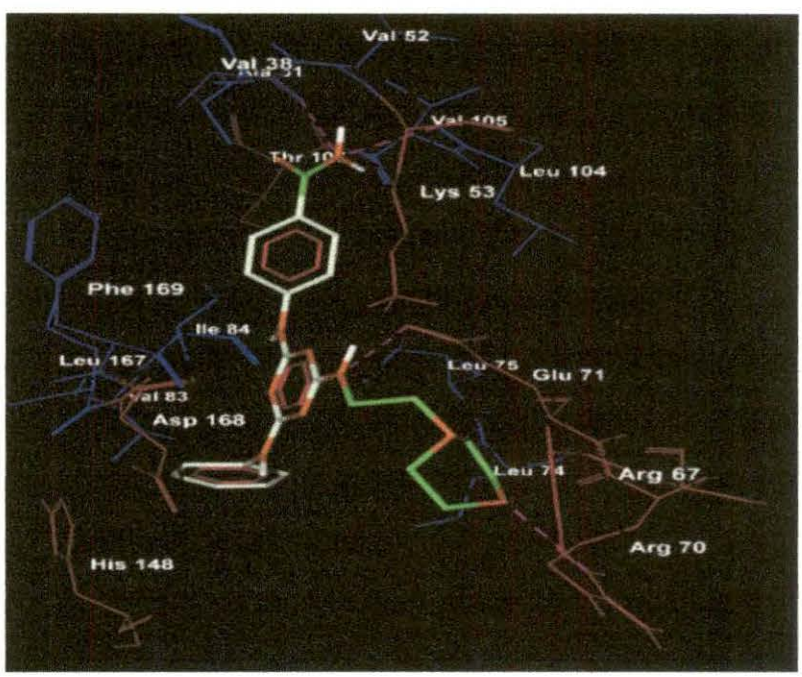
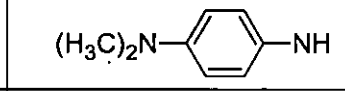
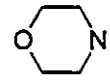
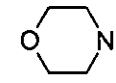
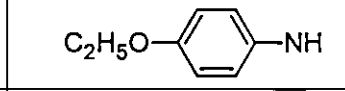
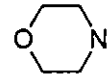
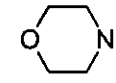
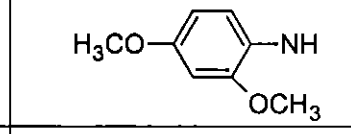
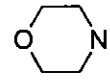
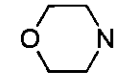
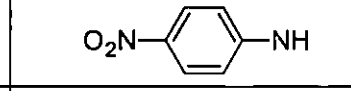
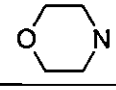
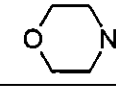
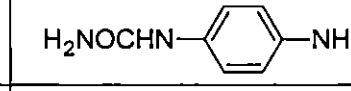
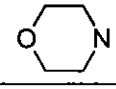
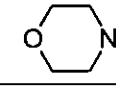
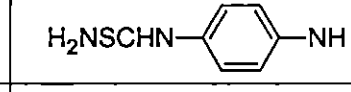
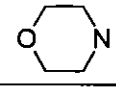
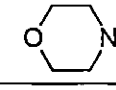
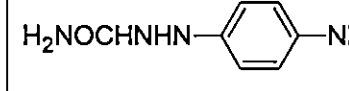
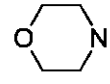
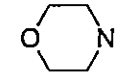
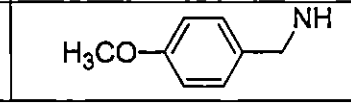
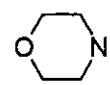
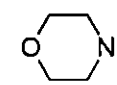
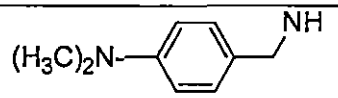
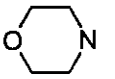
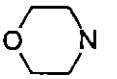
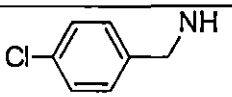
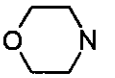
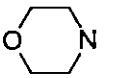
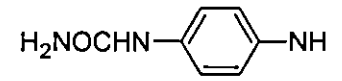
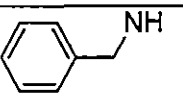
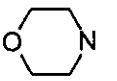
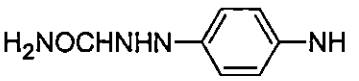
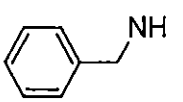
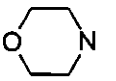
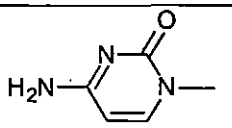
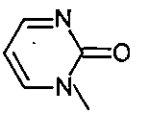
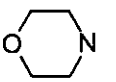
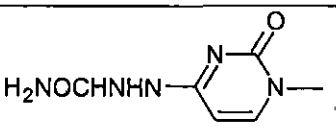
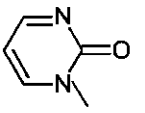
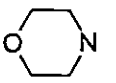
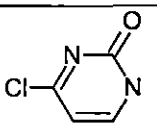
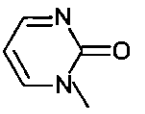
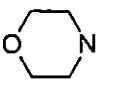
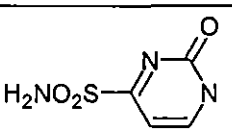
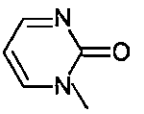
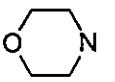


**Figure 4: Docking result of compound11 in the active site of p38 MAP kinase. The inhibitor presented by stick mode (Green is hydrophobic part and orange is hydrophilic part) and important residues in the active site of the enzyme are presented by wireframe.**

It is seen that presence of one morpolino or anilino ring is essential for the activity of triazine derivatives. Also introduced pyrimidine ring in triazine scaffold gives very high predicted activity. Basis of this finding we designed several triazine analogues as more potent p38 MAP kinase inhibitors and Docking simulation was performed. Structures of Triazine analogue and their corresponding binding energy are shown in Table 3.

**Table 3:Chemical structures of designed triazine derivatives by substituting R1, R2, R3 of E and their Docking energies**

| Compound | R1 | R2 | R3 | Binding Energy (kcal/mol) |
|---|---|---|---|---|
| 1a |  |  |  | -8.75 |
| 1b |  |  |  | -7.96 |
| 1c |  |  |  | -9.03 |
| 1d |  |  |  | -9.07 |
| 1e |  |  |  | -8.51 |
| 1f |  |  |  | -9.29 |
| 1h |  |  |  | -8.83 |
| 1i |  |  |  | -9.40 |

| Compound | R1 | R2 | R3 | Binding Energy (kcal/mol) |
|---|---|---|---|---|
| 1j |  |  |  | -9.31 |
| 1k |  |  |  | -9.30 |
| 1l |  |  |  | -9.36 |
| 1m |  |  |  | -8.1 |
| 1n |  |  |  | -9.26 |
| 1o |  |  |  | -8.15 |
| 1p |  |  |  | -7.13 |
| 1q |  |  |  | -8.0 |

Docking analysis also performed for the design triazine inhibitors and inhibitors are docked in to allosteric site of P38 MAP. Their calculated free energy of binding is shown in table 3. From the table it is shown that lowest binding energy value is obtained for lowest binding energy value is obtained for designed compound 1i. Docking structure of compound1i is presented in Figure 5. One of the designed compound (1i) docked with p38 MAP kinase is presented in Figure 6. It is observed that designed compound is also inside the binding cavity (Figure 6).

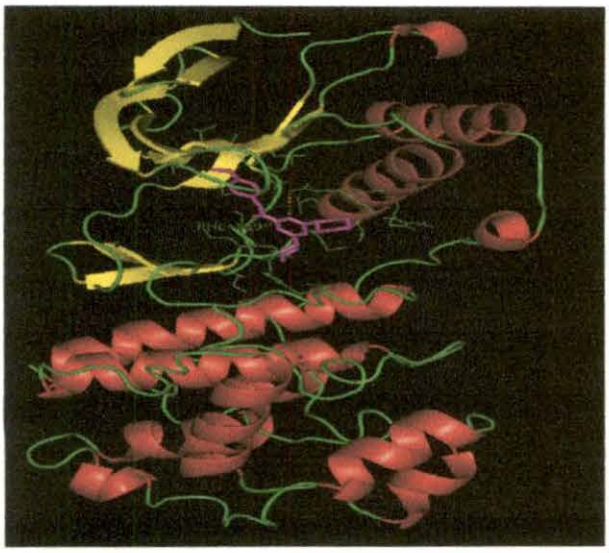


**Figure 5: Compound 11 in a binding pocket of P38 MAP. The inhibitor show in stick model (magenta)**

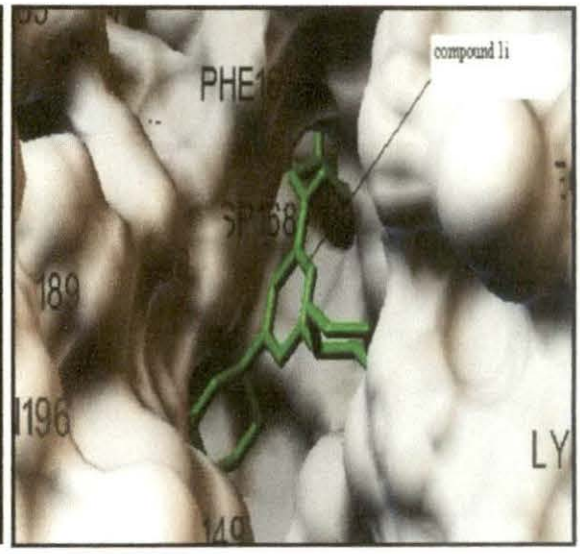



**Figure 6: Docking result of triazin inhibitor (1i)**

Figure 7 shows ligand was surrounded by both polar and apolar hydrophobic group such as Glu71, LYS 53, VAL52, ALA51, PHE169, ASP168, HIS148, LEU75, LEU67, VAL83, ILE84, MET78. These amino acid residues stabilized both polar and apolar parts of ligands,

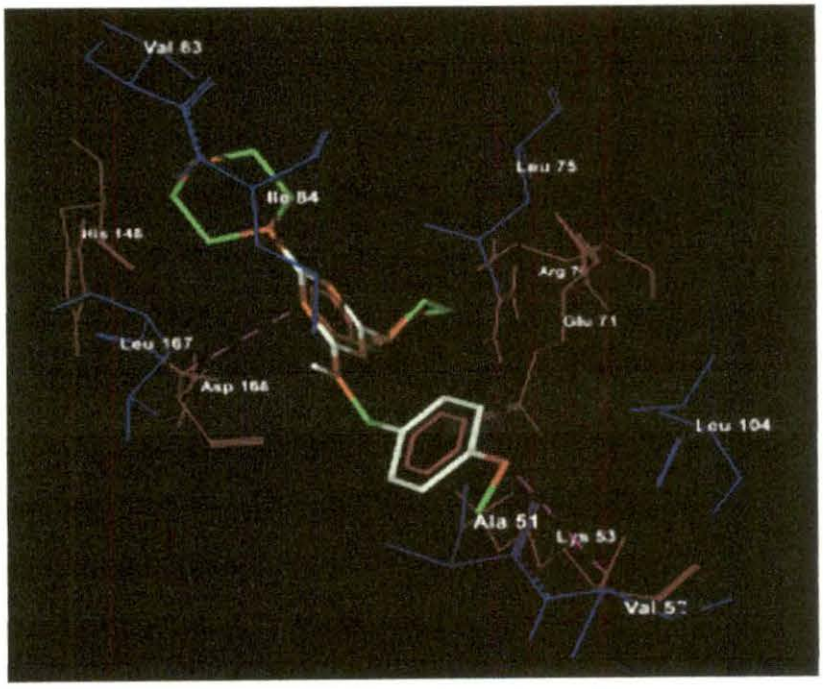



**Figure 7: Docking pose of compound 11 in the active site of p38 MAP kinase. The inhibitor presented by stick mode (Green is hydrophobic part and orange is**

**hydrophilic part) and important residues in the active site of the enzyme are presented by wireframe.**

**9.4: References**


1. Whitmarsh AJ, Davis RJ (1996)*J. Mol. Med.*, **74**: 589- 607.

2. Garrington TP, Johnson GL (1999)*Curr. Opin. Cell Biol.*, **11**: 211-218.

3. Cohen, P. The search for physiological substrates of MAP and SAP kinases in mammalian cells. *Trends. Cell Biol.* **1997**, *7*: 353–361.

4. Ip, Y.T., and R.J. Davis. Signal transduction by the c-Jun N-terminal kinase.(JNK): from inflammation to *development. Curr. Opin. Cell Biol. 1998*, 10:205–219.

5. Davis, R.J... Signal transduction by the c-Jun N-terminal kinase (JNK): from inflammation to development. *Curr. Opin. Cell Biol.* **1998,** 10: 205–219.

6. GonzalezFA, Raden D, Rigby MR, Davis RJ (1992) Heterogeneous expression of four MAP kinase isoforms in human tissues. *FEBS Lett.*,**1992,** 304, 170-178.

7. Mandlekar S, Kong AN (2001) Mechanisms of tamoxifen-induced apoptosis. *Apoptosis.6*: 469-77.

8. Kant S, Schumacher S, Singh MK, Kispert A, kotlarov A (2006) *Biol Chem.*, **281**: 355111-355119.

9. Choy EHS, Panayi GS (2001) Mechanisms of Disease: Cytokine Pathways and Joint Inflammation in Rheumatoid Arthritis. *N Eng J Med.*, **344**: 907-916.

10. Dinarello CA (1991) Inflammatory cytokines: interleukin-1 and tumor necrosis factor as effector molecules in autoimmune diseases.*CurrOpinImmunol.* **3**: 941-948.

11. PalladinoMA,BahjatFR,TheodorakisEA,MoldawerLL (2003)Anti-TNF-alpha therapies: the next generation.*Nat Rev Drug Disc.,* **2**: 736-741.

12. Chen E, Keystone EC, Fish EN (1993) *Arthritis and Rheum.*,**36:**901-910.

13. SalituroFG, Germann UA, Wilson KP, Bemis GW, Fox T, et al. (1999) *Curr Med Chem.*,**6**, 807-823

14. Harada J, Sugimoto M (1999) *Japan J Pharmacol.*,**79**: 369-378.

15. SilenJL, LuAT,SolasDW, GoreMA, MacleanD (1998) Screening for Novel Antimicrobials from Encoded Combinatorial Libraries by Using a Two-Dimensional Agar Format. *Antimicrob. Agents Chemother.* 42, 1447.

16. Klenke B, Stewart M, Barrett MP, Brun R, Gilbert IH (2001) *J. Med. Chem.*, **44:** 3440.

17. Menicagli R, Samaritani S, Signore G, Vaglini F, Via LD (2004) *J. Med. Chem.*, **47**: 4649.

18. Foster BJ, Harding BJ, Leyland JB, Hoth D (1986) *Cancer Treat. Rev.*, **8**: 197.

19. Leftheris K, Ahamed G, Chan R, Dyckman AJ, Hussain Z (2004) The discovery of orally active triaminotriazine aniline amides as inhibitors of p38 MAP kinase.*J. Med. Chem.*, *47*: 6283.

20. Zheng M, Xu C, Ma J, Sun Y, Du F, et al. (2007) Synthesis and antitumor evaluation of a novel series of triaminotriazine derivatives. *Bioorganic & Medicinal Chemistry. 15*: 1815-1827.

21. http//www.seanet.com/~mthompson/Arguslab.

22. Pargellis C, Gilmore L, Graham AG, Grob PM, Hickey ER (2002)Inhibition of P38 Map-Kinase by utilizing a Novel Allosteric Binding State. *J Regan Nat. Struct. Biol.*, **9**: 268-272.

23. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.*, **19**: 1639-1662.

24. Molegro Molecular Viewer (MMV), **2008.** 1.0, Molegro, Denmark.

25. Delano WL (2002) The PyMOL Molecular Graphics System.

# CONCLUSION

# Conclusion

This thesis is focused on evaluation of QSAR analysis some drug molecules and also molecular docking, MD simulation was carried out for some inhibitor and its receptors. The major findings obtained from the studies are given below

Renin is a key enzyme, initiates the enzymatic cascade producing the angiotensin peptides that control blood pressure, cell growth, apoptosis and electrolyte balanced. Binding energy data shows that inhibitors are good binding pose. Hydrogen bonding interactions are important for stability of the complex. All docking results show that common hydrogen bonds formed between P1 moiety and Asp226. Also Gly228 makes a hydrogen bond with the P3' moiety of the 72X. It observed that Asp226 and Gly228 residues are important for binding. Hydrophobic interactions are also crucial for the stability of the complex. LogP value suggests that the inhibitor has hydrophobic environment. P1, P2, P1', P2'and P3' residues of the inhibitor surrounded by hydrophobic residues of the protein such as Thr15, Thr18, Tyr20, Tyr83, Thr85 , Pro118, Phe119, Leu121, Ala122, Phe124, Val127, Thr224, Thr227, Ala229, Tyr231, Met303, Leu252 (B) and Phe253(B).

Type IIa receptor protein tyrosine phosphatases (RPTPs), such as RPTPσ, LAR and RPTPδ, are cell surface receptors which play an important role in neuronal development, function and repair. From the analysis it is clear the motion of the protein is distributed among the PCAs. Hydrogen bonds formed between the hydroxyl groups of SER50 and TYR216 (HB6). Another hydrogen bond formed between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB7) in Ig1-Ig2pro-rich loop. It was found that hydrogen bond between the backbones carbonyls of ILE42 and backbone amide of VAL214 (HB6) remain intact during the whole simulation time.

QSAR analysis was performed on a series of 34 inhibitors of anthrax lethal toxin and validated through six QSAR model. Descriptors are used for multiple regression analysis. A QSAR model (model 6) was obtained with LOO cross-validation values of 0.56. The model 6 predictive ability as differentiate by the testing on the external test set and also useful to explain the relationship between compound structure and biological activities and make easy to design of more potent hydroxamate inhibitors.

Aminoacyl -tRNA synthetases catalyze the attachment of amino acids to their specific tRNAs in protein synthesis. In higher eukaryotes several of these enzymes are found in a multienzyme complex. Jeong et al. solved the NMR structure of multifunctional peptide motifs in human bifunctional glutamyl- prolyl tRNA synthetase. To understand the motional properties and mode of action of the human bifunctional glutamyl- prolyl -tRNA synthetase, molecular dynamics simulation of human bifunctional glutamyl-prolyl-tRNA synthetase, in aqueous environment was carried out using the software, GROMACS. From the time evolution, Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) and Radius of gyration (Rg), it was found that the toxin was relatively flexible. Principal Component Analysis (PCA) was also performed for better understanding of motional properties in reduced dimension. All these observations help us to understand the structure and function of human bifunctional glutamyl-prolyl-tRNA synthetase

In this communication, we came out with a mathematical model that involved accepted quantum mechanical parameters and topological indices. This model was applied for screening derivatives of triazine (MAP-kinase inhibitors). We have been able to predict activities that might be taken to inhibitors prior to synthesis. Thus a screening regime might emerge from this study that would lead to efficacies at a relatively reduced cost and telescoping the time-frame as well.

The mitogen-activated protein (MAP) kinases a group of serine/threonine kinases function as critical mediators of signal transduction. MAP kinase causes several diseases, such as asthma osteoarthritis, rheumatoid arthritis, and chronic inflammatory autoimmune disease. Triazine analogues are inhibitor of p38 MAP kinase. Docking of MAP inhibitors are performed using AutoDock and binding energy for the inhibitors are calculated and regression equation is formed using HT29. Effect of substitution is analyzed. It is found presence of morpholinoor anilino ring is essential. Some compounds are designed and their binding energy is calculated. It is seen that designed compound also inside the binding pocket.

In future research rigid docking can be used and further AutoDock Vina can be used for docking of the compounds whose crystal structure is not known. DFT based QSAR modeling; 4D QAR and Molecular Modeling can be performed.

# *Appendix*

# Appendix

**Appendix 1: Software for molecular descriptors calculationand QSAR analysis**

1. HyperChem
2. ACD/LogP
3. Pallas
4. TOPKAT
5. KOWWIN
6. Dragon (TALETE srl)
7. GAMESS
8. Mervin logP calculator
9. MOPAC
10. Sybyl
11. Cerius2
12. CODESSA

**Appendix II: Software for Docking simulation**

1. AutoDock
2. GOLD
3. DOCK
4. GLIDE
5. ICM
6. FlexX
7. SITUS

## Appendix III: Software for Molecular Dynamics Simulation

1. ABINIT (DFT)
2. AMBER (Classical)
3. GROMACS
4. GROMOS
5. GULIP
6. Hippo
7. MacroModel
8. MOLDY
9. NAMD
10. SIESTA (DFT)
11. MOSCITO
12. VASP (DFT)
13. DL_POLY
14. CPMD (DFT)
15. LPMD