Molecular modeling study to explore aryl sulphonamide based anticancer Matrix Metalloproteinase -2 (MMP-2) Inhibitors

Submitted by
Saptarshi Sanyal

Class Roll no: 001711402002
Examination roll no: M4PHA19004
Registration no: 116562 of 2011- 2012
Department of Pharmaceutical Technology
Jadavpur Univeristy
Session:

Under the Guidance of
Prof (Dr) Tarun Jha
Department of Pharmaceutical Technology
Jadavpur University

Thesis submitted in partial fulfillment of the requirements
for the degree of Master of Pharmacy

Department of Pharmaceutical Technology
Faculty of Engineering and Technology
Jadavpur University, Kolkata, India
2019

Certificate of Approval

This is to certify that Mr Saptarshi Sanyal, PG-2 (Exam roll - M4PHA19004; registration no 116562 of 2011-2012) has carried out research work on the subject entitled "Molecular modeling to explore aryl sulphonamide based anticancer Matrix Metalloproteinase -2 (MMP-2) Inhibitor" under the supervision of Prof (Dr) Tarun Jha, professor of Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology of Jadavpur University. He has incorporated his findings in this thesis submitted by him in partial fulfillment of requirements for the degree of Master of Pharmacy (Pharmaceutical Chemistry) of Jadavpur University. He has carried out the research work independently and with proper care and attention to our entire satisfaction.

<div align="center">

Forwarded by   **Supervised by**

                                         Prof (Dr) Tarun Jha
Head   **Professor**
Department of Pharmaceutical Technology,   **Natural Science Laboratory,**
Jadavpur University   **Division of Pharmaceutical Technology**
                                         Department of Pharmaceutical Technology
                                         Jadavpur University, Kolkata

Dean
Faculty of Engineering and Technology
Jadavpur University

</div>

Acknowledgement

I am highly obliged and like to express deepest of gratitude to my respected guide **Prof Tarun Jha,** Professor of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology for his guidance, continuous help, and intellectual freedom throughout the time period.

I would like to thank our HOD Professor (Dr) **Pulok Kumar Mukherjee,** Head of the Department of Pharmaceutical Technology, Jadavpur University for his contunious help to fulfill this project.

I would like to thank all my departmental professors, with a special mention to **Prof Biswajit Mukherjee, Prof Amalesh Samanta, Prof Kunal Roy, Prof Sanmoy Karmakar, Dr Saikat Dewanjee** of Jadavpur University for their continuous help and support.

I would also like to thank **Dr Balaram Ghosh**, Associate Professor BITS- Pilani Hyderabad campus, **Dr Shovanlal Gayen**, Assistant Professor, Dr. Harisingh Gaur University, **Dr Nilanjan Adhikari**, Assistant Professor, Adamas University and **Dr. Supratik Kar**, Post Doctoral Research Fellow, Jackson State University for providing me several intellectual inputs throughout the implementation of the project.

I would like to th ank my mentor and a close friend **Sk Abdul Amin**. I would love to mention **Mr Sandip K Baidya** and **Mr Suvankar Banerjee** to make the workplace a very habitable, friendly and a competitively-learning experience. I would like to give a big shout to **Mr Rajat Sarkar**, **Ms Shubha Mondal** and **Mr. Sanjib Das** for helping me with several professional and mental supports the period of Master of Pharmacy.

Declaration of Originality and Compliance of Academic Ethics

I hearby declare that this thesis contains literature survey and original research work by me, Saptarshi Sanyal, as a part of my Master of Pharmacy studies.

All information in this document have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and provided reference all material and results that are not original to this work.

Name: Saptarshi Sanyal

Exam Roll Number: *M4PHA19004*

Class Roll Number: *001711402002*

Registration Number: *116562 of 11-12*

Thesis Title: "Molecular modeling to explore aryl sulphonamide based anticancer Matrix

Metalloproteinase -2 (MMP-2) Inhibitor"

[SAPTARSHI SANYAL]

Content:

1. Introduction

Introduction

1. Introduction:

The battle against macabre diseases has been in prevalence since the inceptions of human intelligence. In this period of the early 21st century, non-communicable diseases such as diabetes, hypertension, cancer, Alzheimer's disease, chronic obstructive pulmonary disorders (COPD) has now shown to be of great concern for the researchers and clinicians alike and these diseases are proving themselves to be as dreadful as the microbe induced diseases. Cancer lays on the 6th position of the overall causes of death in the world according to the top 10 global causes of death published by world health organization (WHO).

Being a severe threat to the physical, mental and financial threat to human-kind, cancer is one of the diseases under the reticle of researcher through-out the world. The target is not only to stop the proliferation of the neoplastic tissue but to decrease the cost and the severe side effects associated to it. International Agency on Research on Cancer (IARC) is a watchful overseer under the umbrella of WHO. The research body publishes cancer-related data known as the Globocan periodically, 2018 being its latest installment. The organization's data can be found on the global scale and target focused on different countries as well. On a global scale, it has been observed that lung, breast, prostate, and colorectal cancer can be the most prevalent of all the prominent newly added cancer patients (**Figure 1.1**). According to a work by Bray et al, 2019 a major observation was that the country with low HDI, as the likes of Southern asia, the Oral cancer is as prevalent as the lung cancer. The reason may be attributed to the excessive amount of tobacco product, both the inhaled and the chewing type. Smoking (active and passive), alcoholism, obesity, and sedentary life style are some of the major factors for the onset of cancer in a normal human being. Treatment of cancer has a multi façade approach. With only one of

Introduction

chemotherapy, radio therapy and surgery might not individually attempt to treat cancer but can make a worthy effort as a collective.

New cases of cancer 2018

- Lung 12%
- Breast 11%
- colorectum 10%
- prosate 7%
- stomach 6%
- liver 5%
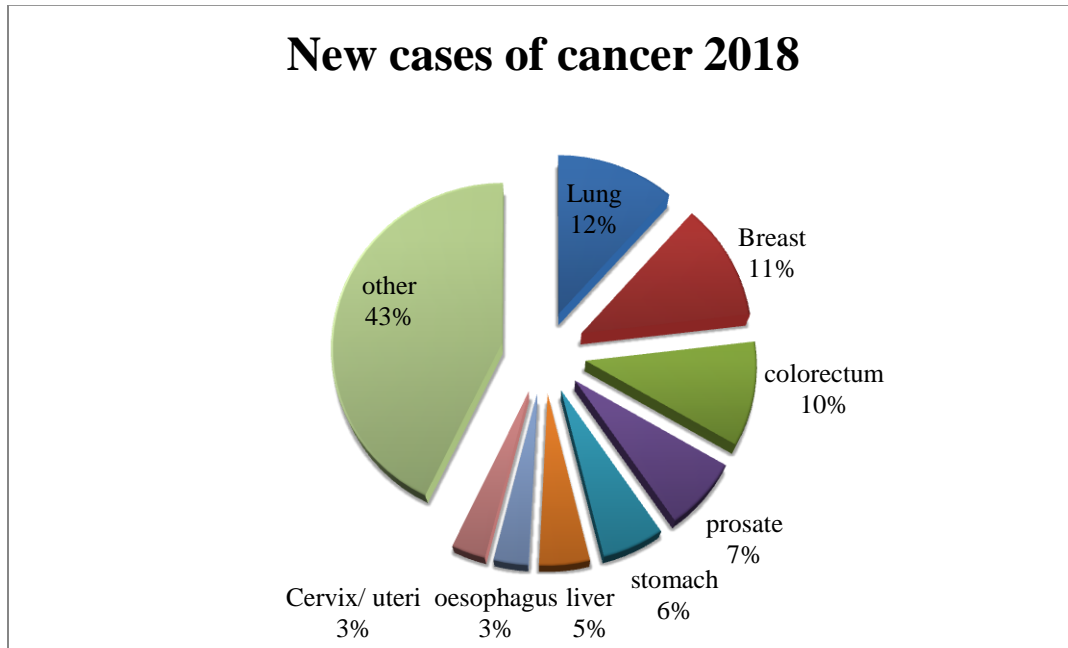- oesophagus 3%
- Cervix/ uteri 3%
- other 43%

Figure 1.1: New cases of cancer on 2018 according to GLOBOCAN. This data is irrespective of gender and represents the global population

A cancer tissue can migrate freely form one portion of the body to another which is known as the neoplasia. The origin of cancerous cell is mostly in the epithelial tissue. However, a particular tissue layer, the extracellular matrix (ECM) prevents the cells to penetrate the blood stream. This degradation of the ECM is carried out by matrix metalloproteinase, a class of 24 zinc-dependent protease enzymes. One of its sub-types, Matrix Metalloproteinase -2 is called the gelatinase as it can break several important matrix components especially the gelatin protein. It is also a major precursor for the formation of vascular endothelial growth factor or (VEGF) that helps in angiogenesis in the tissues. Apart from cancer, MMP-2 is associated with various forms of diseases including diabetes, neuro-degenerative disorders, inflammatory diseases, cardiovascular diseases and pulmonary diseases.

Introduction

In this particular document a thorough study of the MMP-2 protein, its available crystal structure in MMP-2 protein and perform a preliminary molecular modeling study with both small and large number of molecules to explore possible novel inhibitor.

2. Metalloproteins, subtypes and their overview on human physiology.

2. Metalloproteins, subtypes and their overview on human physiology.

2.1. Classification of MMPs

MMPs are a collection 26 different enzymes out of which 24 human enzymes. The MMPs are classified based on their structure and substrate specificity:

2.1.1 Archetypal MMPs

Archetypal MMPs are the subclass of enzymes which has all the common structural feature of MMP enzyme. This subclass can again be further classified as:

2.1.1.1 Collagenases

MMP-1, -8 and -13 are the collagenases as they possess a collagen protein degrading property. Most of the collagen has a triple helical structure that is divided into ¼ and ¾ regions. The hemopexin modules of these enzymes ensure this cleaving which is a major characteristic of these enzymes. It does not require an obvious declaration that the loss of the module diminishes the collagenolysis property of these enzymes (Murphy and Knäuper, 1997). Some proteins present in the extracellular matrix (ECM) like interleukin-8 (Tester et al 2007), pro-Tumor Necrosis Factor α (pro-TNFα) (Gearing et al 1994).

2.1.1.2 Stromelysins

This subclass of MMPs has structural similarities with collagenase. With a wide range of ECM proteins can be its substrate, it, however, cannot process collagenase proteins as they do not have the catalytic ability to diminish the helical structure. The most important use of stromelysins is to activate the proMMPs, procollagenase and proMMP-9 (Barksby et al 2006, Geurts et al 2008). It can cleave the propeptide domain to activate these enzymes. MMP-3 and MMP-10 belong to this subclass.

Metalloproteins, subtypes and their overview on human physiology.

2.1.1.3 Other archetypal MMPs

Four of the other MMPs, MMP-12, -19, -20 and -27 does not fall into any of the previous categories as these enzymes have an ensemble of functions inspite of having all the traits of an archetypal MMP. For instance, MMP-12 is found in macrophages, MMP-19 can be identified in case of rheumatoid arthritis (Kolb et al, 1997) MMP-20 helps in the formation of dental enamel (Lu et al, 2008) An orthologue of MMP-27 has its presence in B-lymphocytes (Bar Or et al, 2003)

2.1.2 Matrilysins

MMP-7 and -26 can degrade ECM proteins like Collagen IV, laminin, entactin etc and they are found in a higher amount in many types of human cancer. MMP-29 can also be a proMMP-9 activator in the diseased state. These two enzymes are called the matrilysins for their matrix degrading activity (Fanjul-Farnandes et al 2010)

2.1.3 Gelatinase

MMP-2 and -9 have an extra site inside the catalytic domain of these enzymes named fibronectin domain. This is unique to only these two enzymes. Denatured collagen or gelatine is the substrate that exclusively binds with this region of the gelatinase. These enzymes degrade collagen type I, IV, V, VII, X, IX fibronectin, etc. (Overall, 2002). Formation of new blood vessels in the tumor, i.e, angiogenesis is governed by gelatinase enzyme with several anti and pro-angiogenic property of these enzymes. Apart from cancer, the presence of MMP-2 and MMP-9 can also be found in several diseased conditions such as bone disease, inflammatory disorders, vascular alteration etc. (Adhkari et al 2017).

Metalloproteins, subtypes and their overview on human physiology

2.1.4 Furin-activated MMPs

These MMPs have a furin group inserted between the catalytic domain and the propeptide sequence. Furin activated MMPs are farther divided into two different segments:

2.1.5 Secreted MMPs

MMP-11, -21 and -28 play an important role in embryogenesis, wound healing and cancer (Rio 2005). Several studies prove a possible relation which states the increased MMP-11 secretions in adipose tissues might be directly correlated to several cases of cancer (Motrescu et al 2008). Other furin activated secreted MMPs (e.g. MMP-28) secret in embryo development (e.g. MMP-21) and in adult tissues like heart, lung, colon, intestine, etc. (Werner et al 2008).

2.1.6 Membrane type MMP

The name of this class expresses some of the function of these enzymes itself. Membrane-type MMPs have an anchor domain that lets the enzyme to be attached to the plasma membrane of the cell and thus these enzymes are the chief proteolytic enzyme for ECM degradation. Membrane-type MMPs (briefly called MT-MMP) are of two subclasses, type 1 transmembrane MT-MMPs and glycosylphosphatidylinositol MMPs (Zucker et al 2003). MT-MMPs are expressed in various organs like brain, lung, colon, testis and leukocytes in normal physiology and also present in case of tumor cells. Activation of proMMP-2 is one of the important activities of MT-MMPs (Sounni and Noel, 2005).

Metalloproteins, subtypes and their overview on human physiology

2.2 Matrix metalloprotein-2, its function:

As discussed already, cancer cells can have some common physiological properties. These cells can proliferate indefinitely, can avoid the apoptosis method of cell death and has the property of metastasis, or moving from one organ to another through the bloodstream. These cells originate for the epithelial tissues; however, the extracellular matrix is a protective layer that must be penetrated to push itself through the bloodstream. Not only that, due to the excess growth of the cancer cells, vascular remodeling occurs in the tissue system. If we look at the cascading system of the enzymatic action (**Figure 2.1**) MMP-2 along with its homologous enzyme MMP-9 is one of the important precursors of a growth factor named aptly as the vascular endothelial growth factor.

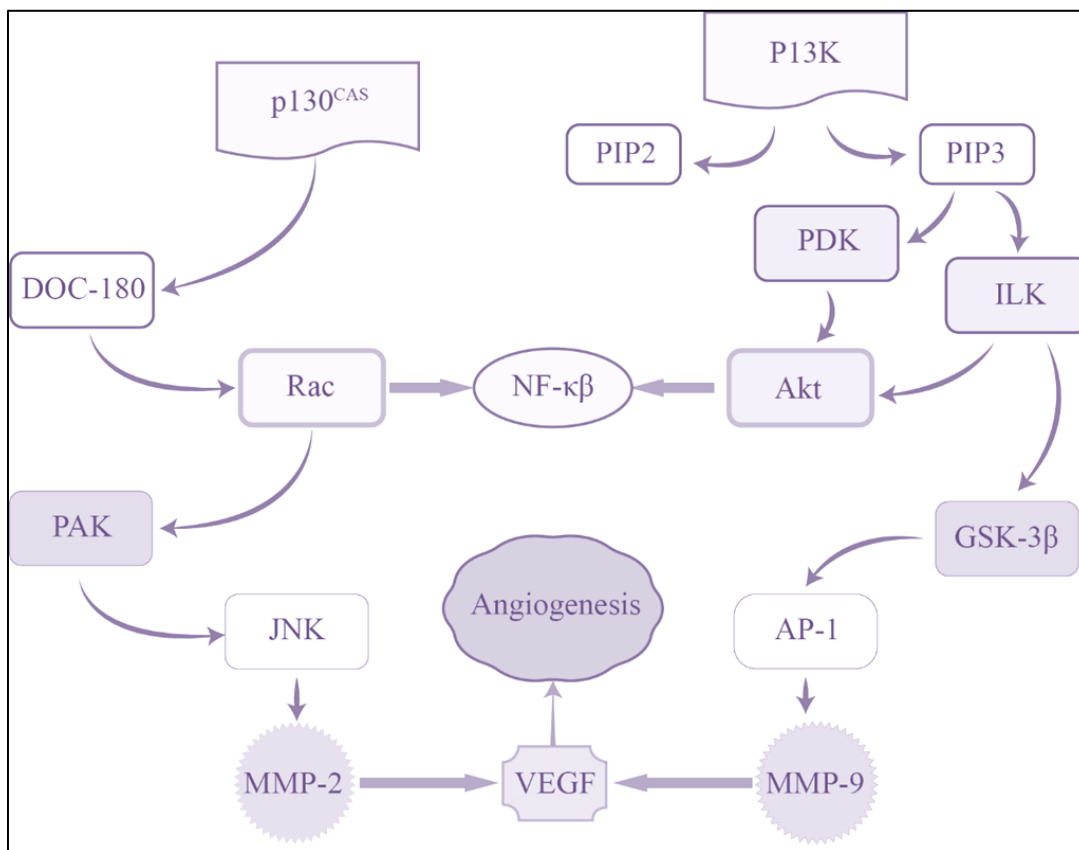# Metalloproteins, subtypes and their overview on human physiology

Figure 2.1: Cascading biochemical pathway of VEGF formation and role of MMP-2 in it.

MMPs, especially MMP-2, has been associated with many more diseases that just cancer. Cardio vascular ailments, AIDS and a series of inflammatory diseases can have major impact by obtaining a specific MMP-2 inhibitor. The **Figure 2.2** shows a number of diseases associated with MMP-2

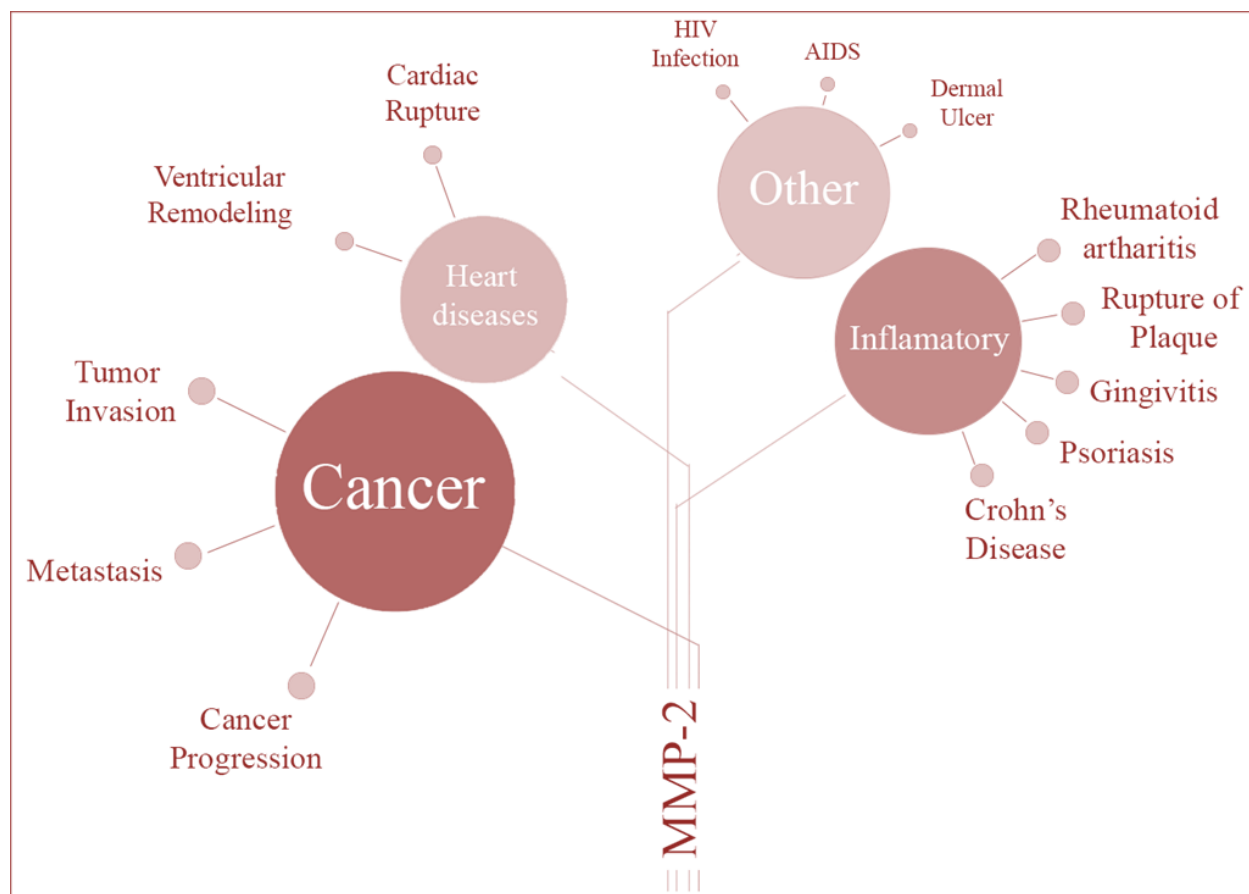# Metalloproteins, subtypes and their overview on human physiology

Figure 2.2: Involvement of MMP-2 in various pathophysiological condition in human body

3. General structural biology involving MMPs and crystal structures of MMP-2

3. General structural biology involving MMPs and crystal structures of MMP-2:

3.1 General Structure of MMPs:

Rational drug designing techniques can utilise the proper knowledge of the structure of a target protein. A drug designing chemist can utilise many new methods by implementing the structure o the ligands. Similar to any other protein, the structural elucidation of MMPs were carried out using nuclear magnetic resonance (NMR) spectroscopy or X-Ray crystallography. With 15 years of research, structures of several MMPs (1-3, 7, 14,16, 20) were identified. A common feature of the proteins were found to have a general structure that consists of a propeptide (~80 amino acid), a catalytic domain (~170 amino acid), a linker peptide (variable length; also known as the hinge region) and a hemopexin group (~ 200 amino acid) (Fanjul-Fernández et al, 2010). However, this is the minimal structure of all the archetypal MMPs; other MMPs may have other unique structural modifications. For all MMPs, the catalytic domain has a common motif of HEXXHXXGXXH (amino acid sequence using single letter format where X is a variable amino acid residue; this X residue helps in the variability of activities of all the MMPs) (Nagase et al, 2005). Also, the propeptide has PRCGXPD which is also known as the "cysteine switch" (Fanjul-Fernández et al, 2010).

3.1.1 Propeptide

The pro-peptide region is made up of nearly 80 amino acids formed into 3 α chains and other connecting loops. MMPs are secreted as enzyme precursor or zymogen and it is 'activated' before it can perform its designated physiological functions. The reason for this protection is important. The enzyme itself is highly active protease enzyme, therefore, the protection is

important to direct the enzyme to its exact substrate that is the gelatine (for MMP-2 and -9) of

the extracellular matrix (ECM). The pro-peptide region needs to be detached and thus controls

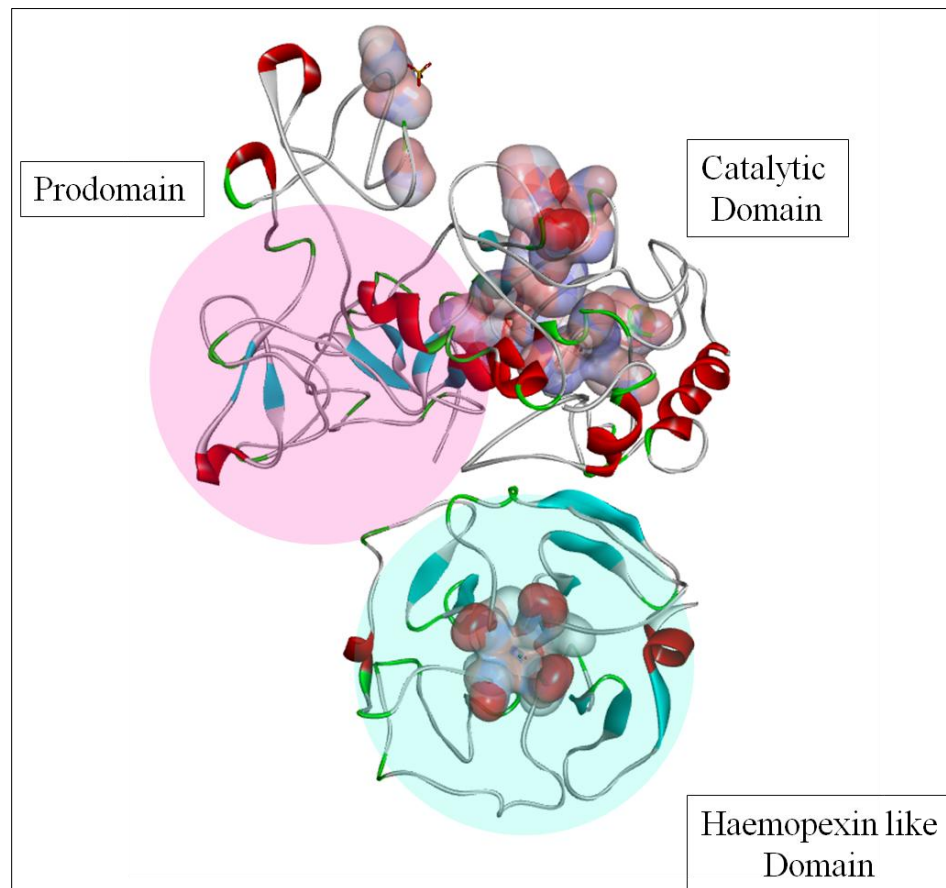the activation of the enzyme (**Figure 3.1**).

Figure 3.1: 3 most inportant parts of an MMP-2 enzyme. The three dimensional surface shown in red, blue and grey represents the catalytic region

The detaching or cleaving work is done at a 'bait region' (Nagase et al, 2005). Though its

location is between the helix 1 and helix 2, the structure is not yet elucidated. This is mainly

because of the flexibility of the region. However, proMMP-2 has its stability due to the presence

of a disulphide bond. A cysteine residue is present near this region which binds with the Zn^{2+} ion

by a sulphydryl (SH) group (Jozic et al, 2004). However, on activation, this link is broken and a

water molecule protects the catalytic zinc ion for the enzyme latency. It has been found that this "cysteine switch" has a motif of PRCGXPD. In some MMPs, this region forms a closed system with the hemopexin domain making it a *closed* structure in the zymogen in contrast to the *open* structure of the activated enzyme. Such a structure can be found for proMMP-1.

3.1.2 Catalytic domain

All of the MMPs show similarity with a crayfish digestive enzyme called the astacin with its catalytic domain. It also holds some structural similarity to the module of snake venom metalloprotease (Bode, 1995). For nearly all of these MMPs, the catalytic domain is spherical in nature with a diameter of 40 Å. This domain contains ~170-180 amino acid residues. Metal ions are an integral part of the enzyme, both catalytically and structurally. It has been observed that in general, this region has one catalytic zinc, one structural zinc and one to three calcium ions. All of these structural ions have its importance in the maintenance of the structural stability of the enzyme. The active site of the catalytic domain is a shallow cleft that lies in the anterior region of the protein. Like any other protein, MMPs have an N-terminal site and a C-terminal site; any substrate binds in this direction and in a horizontal orientation. The active domain can itself be divided into two different subdomains, namely the N-terminal subdomain and the C-terminal subdomain (located in the lower region). The substrate, when enters in the active cleft, is positions itself into these two regions (Tallant et al, 2010)

3.1.2.1 The N terminal region

on the upper portion of the molecule hosts the N terminal region. In its activated form, the enzymes have a salt bridge which conjoins the α amino acid to an aspartate molecule of α helix.

General structural biology involving MMPs and crystal structures of MMP-2

Though the salt bridge has been found to be replaced by simple electronic interactions in many MMP structure, the activity is nearly four folds less than the ones with a salt bridge (Super-active from). It essentially contains a five-stranded β sheets which are oriented from left to right except for the βIV strand which runs right to left. These sheets are all parallel to each other. Two α chains [αA and αB] flank the β sheets (Bode 1995). There are several elements that join a number of loops together which are different from each other in different MMPs (Tallant et al, 2010).

3.1.2.2 The C terminal region

This subdomain is smaller in stature than the N terminal region. It is less regular with its folding and comprises 2 open loops and the C terminal helix (αC). The polypeptide chains that connect the β sheets are termed as LβIIβIII, LβIIIβIV and LβIVβV. The LβIVβV linker, the three oxygen atoms of the main chain carboxylic acid groups, two solvent molecules and carboxylic acid oxygen atom from Asp173 form an octahedral structure with the calcium ion. The LβIIIβIV (known as the "S-loop") binds with a structural zinc ion and three histidine residues (His147, His162 and His175) tetrahedrally and one aspartate residue (Asp149) in a monodentate formation. It has been found that the structural zinc is important as any Zn^{2+} deprived synthetic MMP will result in no activity. The S-loop also creates a second octahedral loop with calcium ion along with main chain carbonyl oxygen atoms of Asp154, Asp177 and Glu180. Apart from these two calcium ions, in the structure of MMP-2, another calcium ion is inherited into the structure by forming a trigonal-bipyramid with LβIαA and LβVαB along with two main chain carbonyl oxygen atoms and carboxylic acid groups. The catalytic zinc ion is housed inside the active-site helix or the αB helix. For all MMPs, amino acid motif around the catalytic zinc ion is

General structural biology involving MMPs and crystal structures of MMP-2

HEXXHXXGXXH (X is any residue). This Zn ion is bound to three histidine residues (including His 218, 222, 228) and water molecules (Adhikari et al 2017). This particular water molecule helps the hydrolysis of the substrate protein.

The active site has 6 different binding positions (S1, S2, S3, and S1', S2', S3' pockets). Among these sites, one of the most important targets of designing potent MMP inhibitors is the S1' pocket. This pocket is hydrophobic in nature and less solvent exposed. This part is of variable size in different MMPs which makes it an important target for drug specificity. MMP-1, 7, 11 and 20 have small pockets, MMP-3, 10, 13 have the large S1' pocket and MMP-2, 8, 9, 12, 14, 16 have medium sized S1' domains (Adhikari et al 2017).

3.1.3 Hemopexin Domain

An important part of the MMP structure can be identified as the hemopexin domain which has approximately 200 amino acid residues. This domain has its structural similarities with a few unrelated proteins, one of which is the hemopexin protein, and thus, the terminology has evolved (Altruda et al, 1985; Jenne and Stanley 1987). The topological analysis through X-Ray crystallography confirms that the chain has four folds making it a propeller-like structure (β propeller). Four antiparallel β (called the 'Blades') of the propeller which surrounds a funnel-like formation. The first and the fourth blades are joined together with a disulphide (-SH-SH-) bond. For MMP-2 protein, this domain is not directly linked with the enzymatic activity. MMP-2 protein without the hemopexin region can express the enzymatic activity by the lysis of the gelatin protein (Nagase et al, 2006).

3.2 Crystal structures of MMP-2 on RSCB PDB

General structural biology involving MMPs and crystal structures of MMP-2

The concept of the structure of the protein has a great significance in rational drug designing. This is true for NMR, X-ray and Electron microscopy, all of the structural data has equal importance to a chemist. The health or fitness of any protein structure is mostly governed by the following three metrics:

3.2.1.1 *Ramachandran outliers*

It is calculated using an online web tool named MolProbility. This dictates if the torsional angles (φ and ψ angles) are unusual nature for certain residues. This is then calculated as a percentage of the total residues present in that structure.

3.2.1.2 Side chain outliers:

This metrics is also calculated using the MolProbility. A side chain is considered to be an outlier when the torsional angle has no similarities to any preferred combination.

3.2.1.3 Clash score:

Clash score is determined by the number of pair of atoms which are in close vicinity to each other and the Van der Waal volume gets overlapped.

MMP-2 has a UniProt ID P08253 (72 kDa) (https://www.uniprot.org/uniprot/P08253) and. Several types of research have been conducted on elucidating the protein structure since 1995 and 11 entries can be found in the protein databank (PDB) of *Research Collaboratory for*

General structural biology involving MMPs and crystal structures of MMP-2

Structural Bioinformatics (RSCB) which is a joint initiative of Rutgers University and University

of California, San Diego, USA. A list of these structures is jotted in **Table 3.1**.

Table 3.1: Summarization of several MMP-2 structures present in RSCB PDB

PDB ID	Method	Resolution (for X-Ray only)	Number of poses (NMR)	Structural Feature
1CK7	X-Ray Diffraction	2.8 Å	-	Full length, including pro domain
1EAK	X-Ray Diffraction	2.66 Å	-	The catalytic site of proMMP-2
1GEN	X-Ray Diffraction	2.15 Å	-	C terminal domain
1GXD	X-Ray Diffraction	3.1 Å	-	ProMMP-2 / TIMP complex
1QIB	X-Ray Diffraction	2.8 Å	-	Active catalytic site
1RTG	X-Ray Diffraction	2.6 Å	-	Haemopexin like domain
3AYU	X-Ray Diffraction	2 Å	-	Active catalytic site
1CXW	NMR	-	50	Fibronectin type II region
1HOV	NMR	-	14	Active catalytic site
1J7M	NMR	-	50	Third fibronectin type 2
1KS0	NMR	-	50	First fibronectin type 2

3.2.2.1 1CK7

Morgunova and co-workers proposed 1CK7 (**Figure 3.2**) structure of *Homo sapiens* MMP-2 in

1999. The structure represents a full length 72 kDa structure of the Gelatinase A or MMP-2.

However, this also includes the prodomain sub-structure and it can be termed as the structure of

the inactivated proMMP-2 protein which is the inactivated form. This structure was developed

using X-ray crystallography method (Resolution: 2.8 Å). The amount of Ramachandran outliers were fairly low (2%), however, when to compare to other structures, the percentile ranks for the outliers was in the median. This structure suffers from poor clash score (54) and 22.60% of side chains remained on the outliers. 80% and 90% of the other structures showed better scores for both the scoring systems. **Figure 5** shows several sub-structural clashes of this PDB submission (Morgunova, 1999).
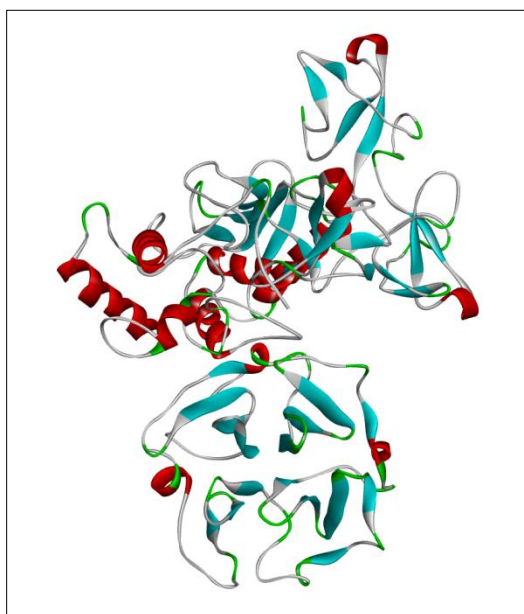

Figure 3.2: 3D crystallographic stucture of PDB: 1CK7

3.2.2.2 1CXW

Initially explored by the Briknarova *et al* (1999) and this structure (**Figure 3.3**) was published in RSCB PDB with ID 1CXW in 1999. This NMR structure has 50 different poses and it showed a very low clash score of 2.0 and only 1 structure (PDB: 3AYU) shows a better score than this

structure. The Ramachandran outlier is low, 1.8%; on the contrary, the side-chain outlier, however, is high (12.40%) and it has a 60th percentile rank. This structure only focuses on the fibronectin type II region that works as an ancillary in gelatin-binding. Therefore, due to its functionality, this inhibition may be worthy of any further exploration.
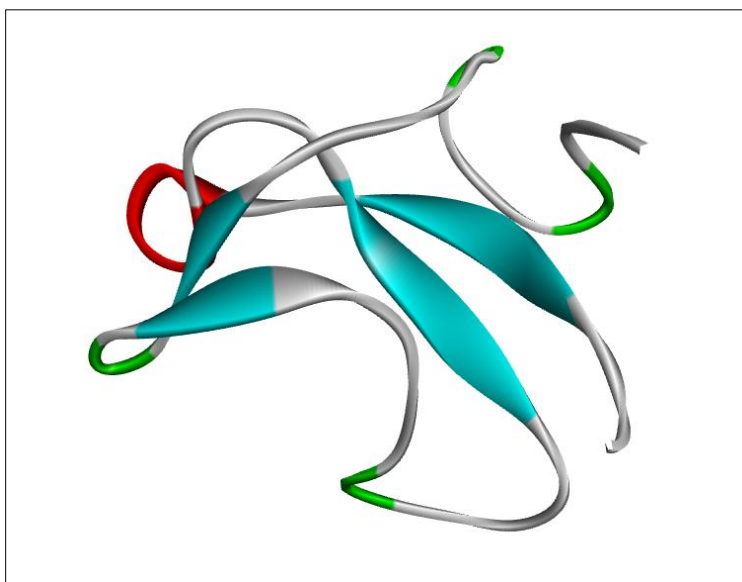
3.2.2.3 1EAK

In 2002 the 1EAK structure (**Figure 3.4**) of the catalytic site of the proMMP-2 protein was published (Bergmann et al, 2002). The catalytic site of the protein was expressed with 2 different bioassembly. The structure has 41 clash scores and 2.60% and 13.50% of Ramachandran and side chain outliers respectively making it a bellow average contender among all the enzymatic structures of MMP-2.

General structural biology involving MMPs and crystal structures of MMP-2
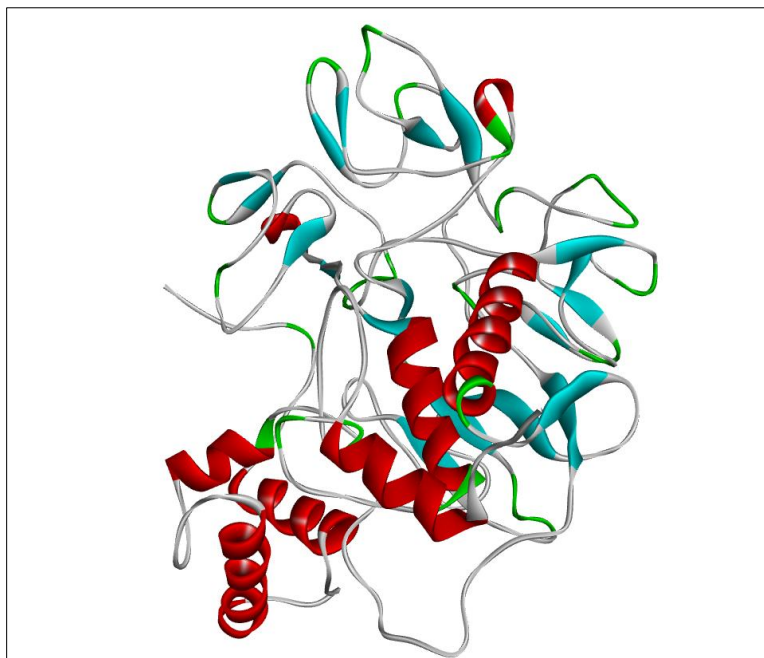
Figure 3.4: PDB structure 1EAK of MMP-2

3.2.2.4 1GEN

The 1GEN is an X-ray crystallographic (2.15 Å) structure with of MMP-2 that has the C terminal domain exclusively (**Figure 3.5**). With less Ramachandran outliers (0.50%; the best among all the structures compared here), a fewer side chain outliers (8.80%; only 3 among 11 structures have better outlier scores) and a plausible clash score (9; 30% of the structures held better score) make this an acceptable structure. On the other hand, lack of the full catalytic domain (only one Zn^{2+} is present here in the structure) makes it unfit for most of the Zn^{2+} binding approach of exploring MMP-2 inhibitors (Lisbon et al, 1995).

General structural biology involving MMPs and crystal structures of MMP-2
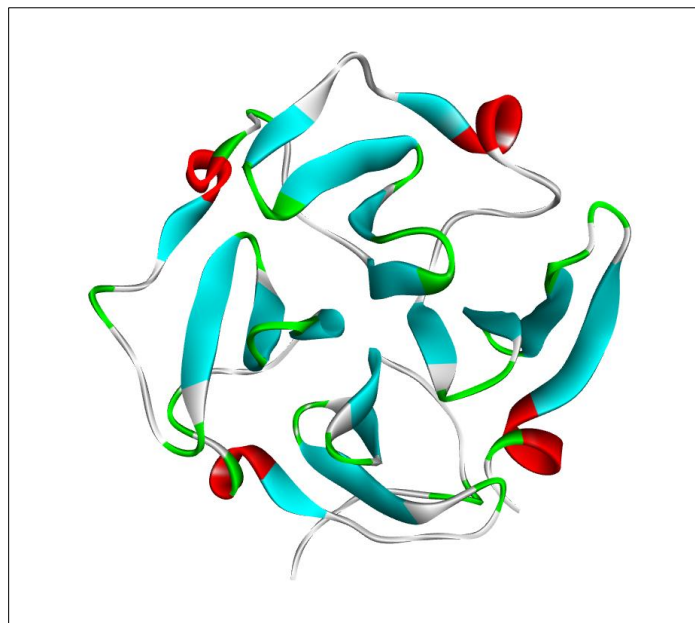
Figure 3.5: PDB structure 1GEN of MMP-2

3.2.2.5 1GXD

Morgunova and co-workers reported this X-Ray crystallographic structure which shows a rough 3.1 Å resolution structure of proMMP-2 / TIMP complex system (**Figure 3.6**). The structure performs poorly with 55 clash score, 5.50% Ramachandran outlier and 15.30% of side chain outliers. At least 80% of the other structures outperform this structure. This, along with poor resolution and lack of any co-crystal structure deems it unfit for structure-based discovery approaches (Morgunova et al, 2002).

General structural biology involving MMPs and crystal structures of MMP-2
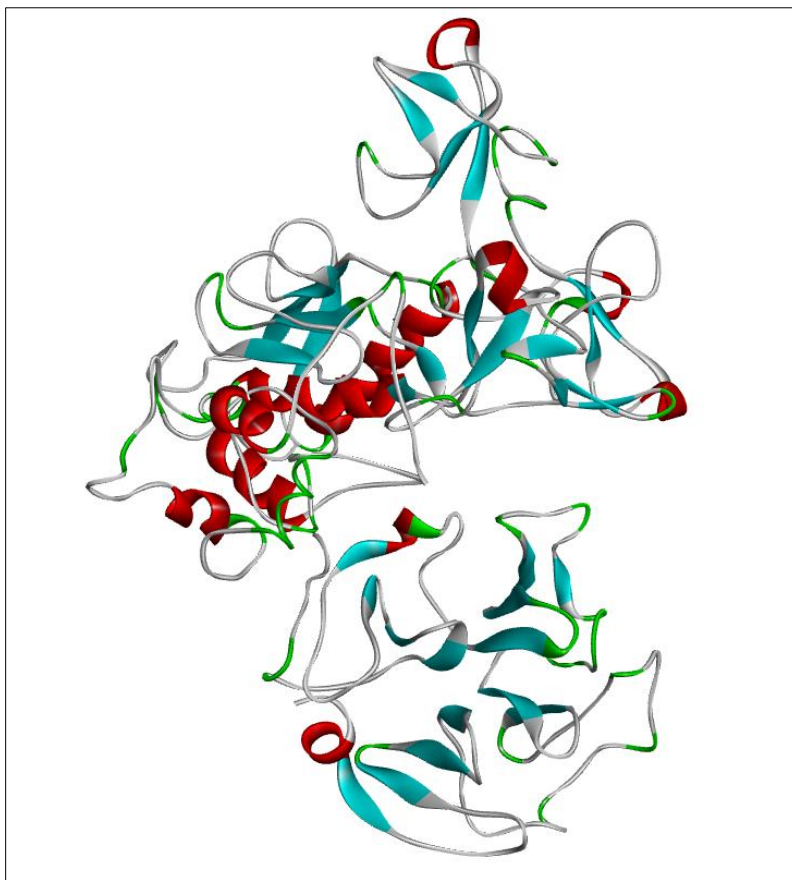
<p style="text-align:center">Figure 3.6: 3D crystallographic structure of MMP-2 (BDB ID: 1GXD)</p>

3.2.2.6 1HOV

1HOV is a set of 14 different NMR first published by Feng *et al* of Pharmacia Corporation, the USA in 2001. These 14 sub-structures was claimed to have an RMSD of 1.02 and a 1.62 Å deviation from the mean position of the structures. The structure (**Figure 3.7**) is not very trusted worthy when it comes just of the number as it holds the least position among all of the discussed structures (62 cash scores, 11.20% Ramachandran outlier, 33.20% side chain outlier; lesser the score the better the performance). However, the NMR structure houses a co-crystallized structure of SC-74020 which is a hydroxamate-based Zn^{2+} binding MMP-2 inhibitor. This, along with a

prominent active site pocket, makes this structure a popular choice for the drug discovery scientists for molecular docking studies (Feng et al, 2002).
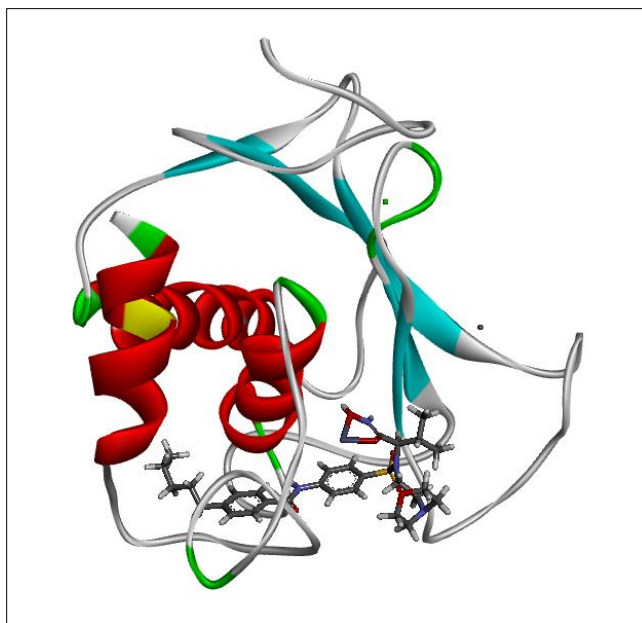
3.2.2.7 1J7M

This structure was released in RSCB by Briknarova *et al* in 2001. This one is also an NMR structure with 50 sub-structure formations. One of such structure is shown in **Figure 3.8**. The fitness metrics of the structure were fair. The clash score was 30 while outliers were 1% (Ramachandran) and 9.8% (side chain) respectively. This structure lacks a proper binding site so this structure is not used for drug designing purpose.

General structural biology involving MMPs and crystal structures of MMP-2
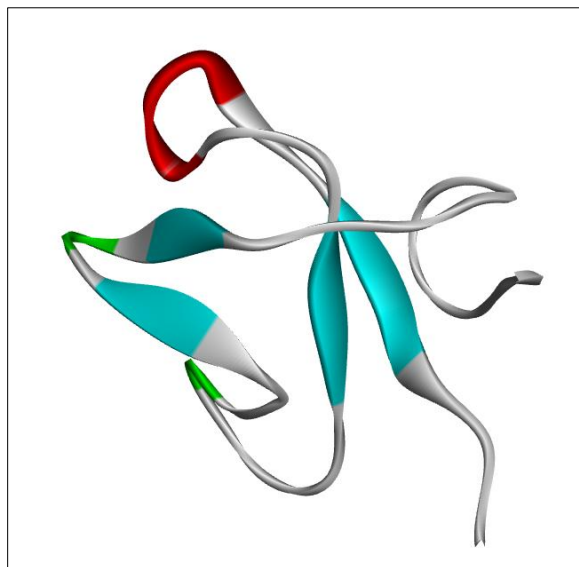
3.2.2.8 1KS0

The gelatin-binding first fibronectin domain of the MMP-2 was explored by Gehrmann et al in 2002, received a universal ID of 1KS0 (**Figure 3.9**). This is an NMR structure with 50 different posses available. The clash score (10) and side chain outliers (10.10%) remain in the median. It has a better Ramachandran outlier (0.9%) than 70% of the other structures. Similar to the previous structure, this structure also has no well-defined pocket for inhibitor binding so gets discarded in the process of selection of protein in structure-based designing.

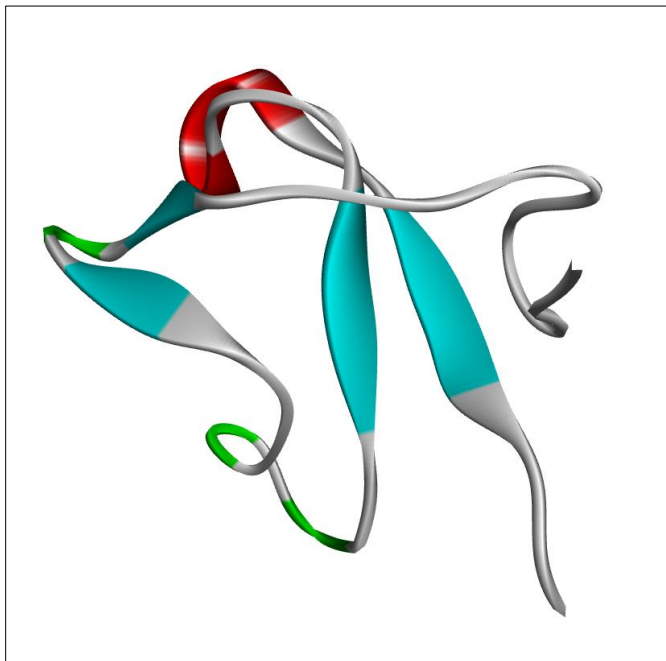# General structural biology involving MMPs and crystal structures of MMP-2

Figure 3.9: 1KS0 structure of MMP-2 fragment

3.2.2.9 1QIB

Dhanraj *et al* (1999) elucidated an X-ray crystallographic structure of MMP-2 catalytic domain and one of its hydroxamate-based inhibitor Batimastat (**Figure 3.10**). The resolution of the structure is 2.8 Å. Ramachandran outlier of 1.90%, side chain outliers of 6.20% and a good clash score of 5 makes it a dependable structure for structure-based drug designing.

General structural biology involving MMPs and crystal structures of MMP-2
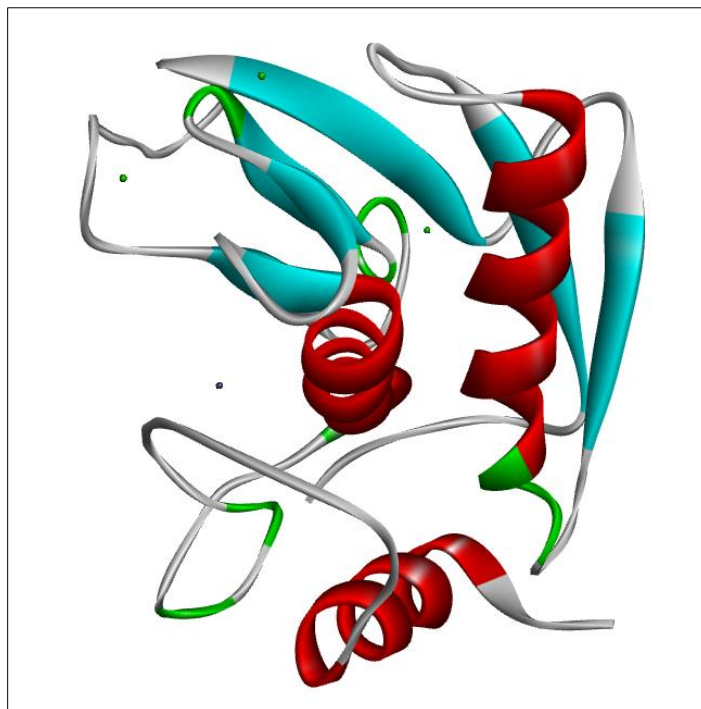
Figure 3.10: 1QIB crystal structure of MMP-2

3.2.2.10 1RTG

Gohlke and co-workers (1996) reported this X-ray crystal structure (**Figure 3.11**) of MMP-2 is was one of the maiden Gelatinase A structure reported in PDB. It has a resolution of 2.6 Å, 9 clash score and 2.0% and 6.4% outliers (Ramachandran and side chain respectively). This structure reports particularly on the C terminal domain, mainly the haemopexin-like domain of MMP-2. Thus, this structure might not be fit for direct docking or molecular dynamic study for identifying any MMP-2 inhibitor.

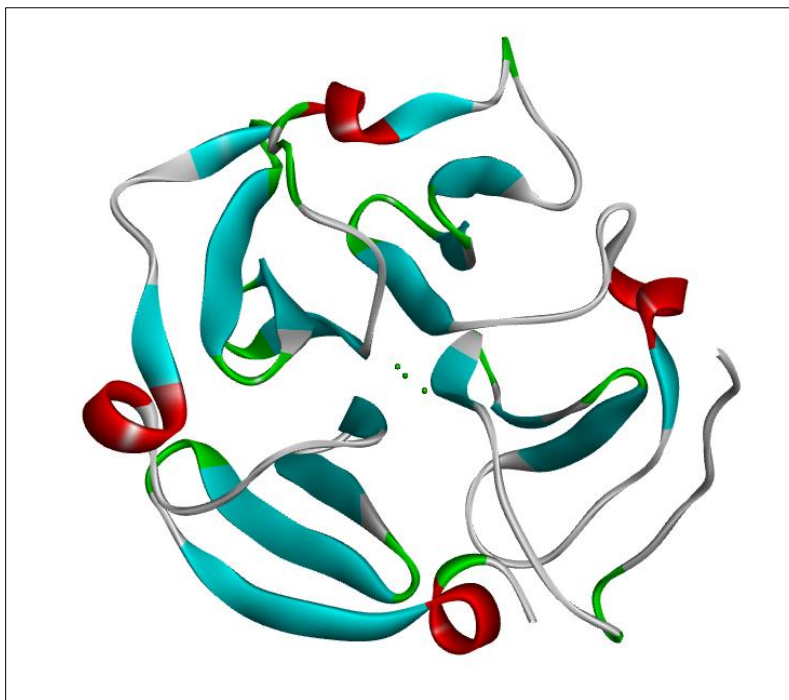# General structural biology involving MMPs and crystal structures of MMP-2

Figure 3.11: 1RTG structure of MMP-2 enzyme

3.2.2.11 3AYU

The Japanese research group comprising Hashimoto and his coworkers (2011) presented a 2.0 Å X-ray crystal structure named 3AYU (**Figure 3.12**). The interesting observation of the structure is that the structure shows few Ramachandran outliers, the clash score also limited to 3. For all these reasons, the structure can outperform any other structures of MMP-2. On the other hand, the structure houses an oligopeptide-based β-amyloid precursor protein-derived inhibitor (APP-IP) as its co-crystal structures which, in turn, opens up the scope of protein-protein flexi-docking. Due to these qualities, the structure has a great potential to be used in molecular docking for possible ligand-protein and protein-protein interactions. Several small molecules also have been docked by several researchers which we have discussed in the later part of this communication.

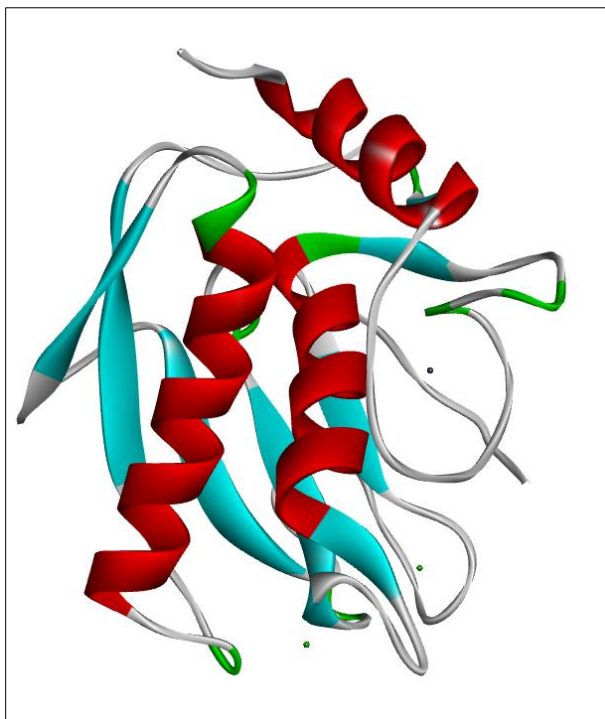# General structural biology involving MMPs and crystal structures of MMP-2

These 11 structures which were available in the PDB might provide an in-depth knowledge of the overall protein structure of Matrix Metalloproteinase 2 and their respective roles in the protein function, only a few (3 to be exact) had been mostly used in docking and molecular dynamics based study for exploration of possible newer drug interaction as well as for finding a novel compound or substructural features. In the following figure, we visualize the comparative analysis of the proteins according to 3 of the most important parameters that determine the quality of the protein structures (Chen et al 2009).

As discussed earlier, the concept of an enriching structure is dependent on 3 parameters namely Ramachandran outlier, Sidechain outliers and clash score. To compare the quality of the models we plot the percentile rank value against the PDB ID in **Figure 3.13.**

35

General structural biology involving MMPs and crystal structures of MMP-2
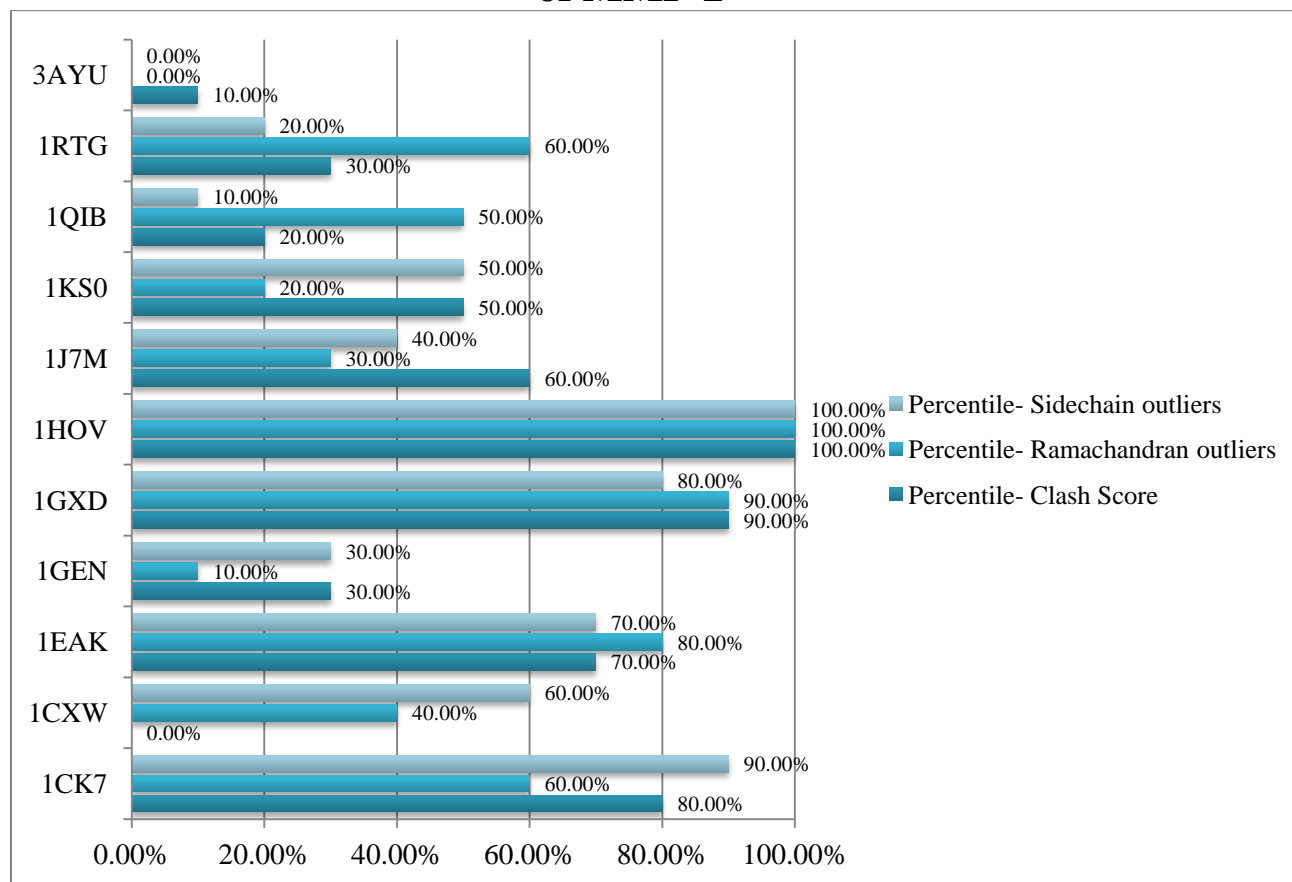

Figure 3.13: Comparison of the quality score of different MMP-2 structures across PDB

# General structural biology involving MMPs and crystal structures of MMP-2

3.3 MMPIs through clinical trial:

Broad spectrum anti-MMP drug had been into clinical trial through several decades. As the MMP-2 can be great target for cancer, this has been studied not only by researchers from academia but in the industrial scenario as well. Many of them showed promising future but no pharmaceutical company could successfully made an MMP-2 inhibitor see the broad market possible for use. A brief list of these potential inhibitors is shown in **Table 3.2.**

Table 3.2: A list of MMP-2 inhibitory drugs through different phases of clinical trial
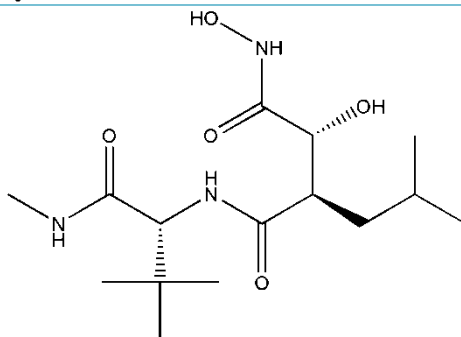
Name	Structure	Description
CSG-27023A		• Discovered by Novartis • Broad Spectrum MMPI • Intended to use for cancer and arthritis • Removed from clinical trial for high musculoskeletal toxicity
Primostat (AG-3340)		• Developed by Agouron • Provided excellent inhibitory and pharmacokinetic property. • Stage III studies showed less improvement in advance stage and hence removed.
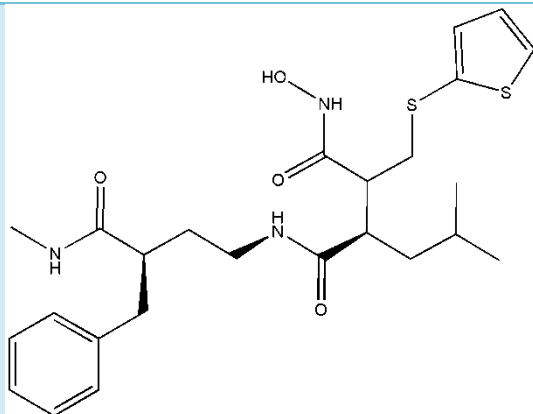
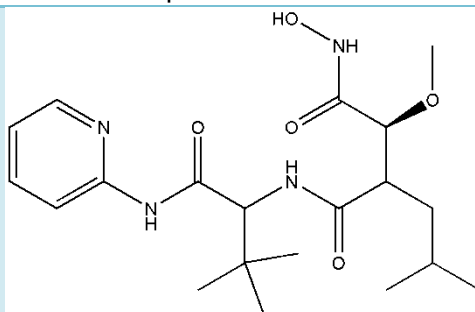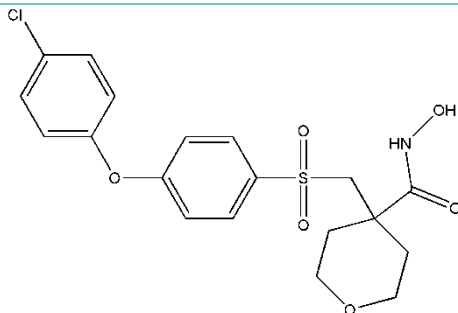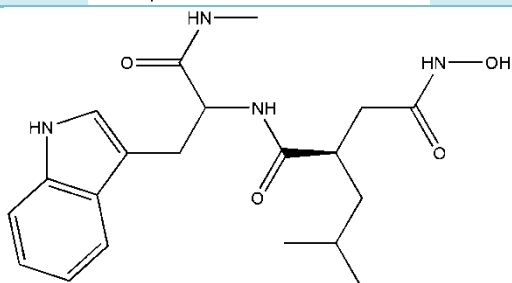|		
Batimastat (BB-94)		• Discovered by British Biotech • Failed in phase II clinical trials
Marimastat (BB-2516)		• Discovered by British Biotech • Failed in the third phase of clinical trial
Solimastat (BB-3644)		• Discovered by British Biotech • Failed in the first phase of clinical trial
RS-130830		• Developed by Roche • Discovered as a potential drug for arthritis • Stopped in clinical trial phase II
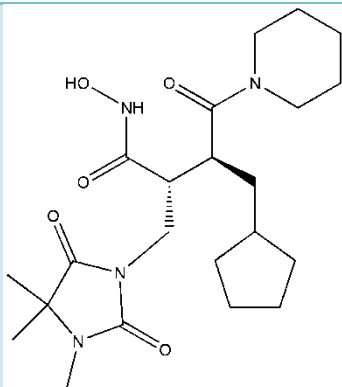
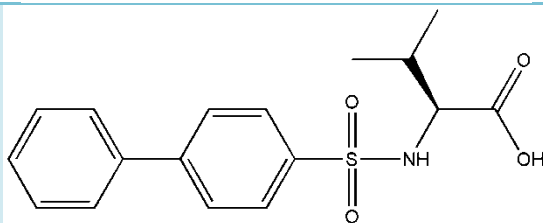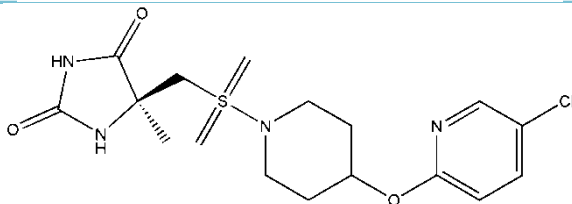|		
RO-32-3555		• Developed by Roche • Discovered as a potential drug for arthritis • Stopped in clinical trial phase III
Ilomastat		• Developed by Glycomed • Intended to target COPD • Withdrawn in clinical trial III
PD166793		• Developed by Pfizer • Cardiovascular disease was targeted • Non selective to TACE
AZD1236		• Discovered by Astra Zeneca • Target was severe COPD • Safety profile was excellent

5. Modeling techniques used in designing MMP-2 inhibitor:

5. Modelling techniques used in designing MMP-2 inhibitor:

Drug designing is an essential part of modern drug discovery as this can reduce the overall financial, operational and human resource cost of drug discovery into a fraction of its original. This has made rational drug designing a strong candidate for both the academic and the industrial researchers. Molecular modeling through multiple QSARs, molecular docking, molecular dynamics and virtual screening can be important methods for this. MMP-2 is quite an important target that can not only be beneficial in diminishing the metastatic properties of cancer tissues, but can can help in several different diseased state. MMP-2 has therefore modal concepts for both ligand based and structure based design strategies. These, can be classified as follows:

5.1 Catalytic domain binding

Being a gelatinase enzyme, the main substrate for MMP-2 is the gelatin, type 3 and type 4 collagen, elastin, aggrecan and cartilage links which are some of the important component of normal ECM (Okada, 2017). As this is the region where the enzyme binds with either of the above mentioned protein, the most effective strategy becomes the competitive inhibition of the catalytic domain of MMP-2. The presence of the vital Zn^{2+} ion and the associated flanking pockets (S1, S2, S3 and S1', S2', S3') proves to be essential for the inhibition. However this inhibitory process can have some alternate strategies as well. **Figure 5.1** briefs the possible designing strategies of inhibition of the drug.

5.1.1 Zinc Binding

Among the 2 distinct zinc ions present at the catalytic domain of the enzyme, both of the m are attached to the protein through three histidine residues. The importance of the catalytic zinc is when the substrate protein gets hydrolysed at the glutamine residue and this particular ion

stabilizes it by forming a gemdiolate formation in the intermediate stage. This has made a numerous chemist target the zinc ion by blocking it with a peptidomimetic (glutamine mimetic precisely) small molecule (Sanyal et al 2019)

However the challenge arises when designing a specific inhibitor. The Zn^{2+} ion is inclusive to all MMPs, thus making it difficult to design a specific inhibitor becomes unjust only by binding to the Zn^{2+} ion. The exclusivity of the MMP-2 protein lies in the variable S1' pocket which is hydrophobic and comparatively longer in nature.

5.1.2 Non Zinc binding

Researchers tend to split into two different ideologies when it comes to binding of the catalytic Zn^{2+} ion. It has been found that most of the inhibitors that poses a zinc binding group would be a pan-category inhibitor and subsequently create more adverse effects. Therefore a non-zinc biding molecule that fits perfectly into the S1' pocket of the enzyme might be more selective to MMP-2 than a molecule with a ZBG.
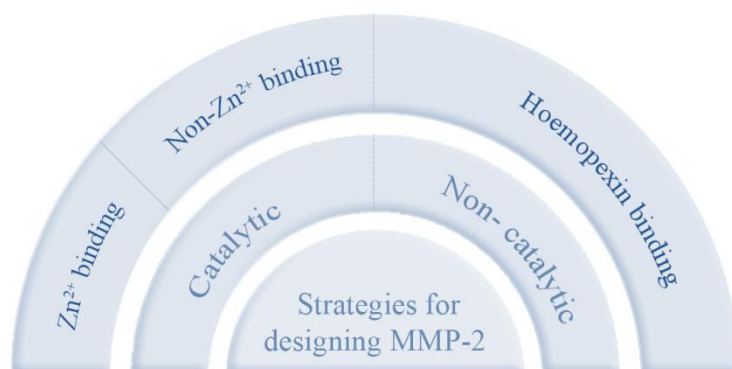

Figure 5.1: Principal designing strategies of MMP-2

Modeling techniques used in designing MMP-2 inhibitor

Instead of any evidence based approach that helped in the field of natural product chemistry, medicinal chemists could make their strategy more target oriented. This target can either be a disease or more precisely an associated protein to the disease. Different types of Molecular modelling techniques can be explained as:

5.2.1 Regression-based 2D-QSAR study

2D QSAR is the classical method of creating a predictive model of biological activity using the the chemical properties of a compound. This method initially demonstrated by Corwin Hans who is also known as the father of classical QSAR. In a simple manner, the relationship can be expressed as

Biological Activity = f(Chemical attributes and properties)

Or, it can be expressed as,

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + \ ... \ + a_n X_n$$

Here X_1, X_2 X_3 … are called independent variables known as descriptors and *Y* is independent variable (biological activity) of the molecule. For most of the cases, the biological activity or Y is the dependent variable which is mostly taken as the negative logarithm of the biological activity. The equation is a statistical equation meant for predicting the biological value. The model is derived using multiple linear regression (MLR) method which can be done in a number of ways such as the forward stepwise method, the backward stepwise method or the genetic algorithm. The descriptors, which are basically several functions applied on different molecular properties, can be calculated based on various physic-chemical aspects. However, for the presence of a myriad of descriptors available, reduction of descriptors is necessary and several machine learning techniques are employed. Genetic algorithm, principal component analysis, lasso-

Modeling techniques used in designing MMP-2 inhibitor

regressions are preferred methods of choice to reduce the number of variables into a rather workable size and this requires keen knowledge on building a proper model. Different validation methods are used to justify the robustness of the model. Mean absolute error, residual value, and squared residual value, leave one out cross validation techniques are some of the important validation methods to figure out the predictability of the model (Todeschini et al 2009).

5.2.2 Classification QSAR study

5.2.2.1 Bayesian Classification study

Bayesian Classification study is a QSAR model which classifies a set of molecules using the Bayes's theorem. It classifies the molecules to be active or inactive on the basis of a predetermined threshold value using molecular descriptors. Using a fragment descriptor known as the ECFP_6, it identifies fragments of a molecule to be favourable or unfavourable for the biological activity. Bayesian classification uses classification instead of estimating the biological activity like the regression-based 2D QSAR technique. Discovery Studio is one of the most used tools used by drug designing chemists, which can be used for Bayesian classification (Adhikari et al 2018).

5.2.2.2 Recursive Partitioning

Recursive partitioning is a classification based QSAR method that uses a machine learning method called 'Decision Tree' and in this way, the method calculates the relationship between biological activity and different molecular properties. The tree is generated based on a specific criterion and farther subdivided into multiple sub-sets (nodes). The tree thus formed is allowed to overbranched and based on a pre-determined cross-validation steps the trees are trimmed off the excessive branches (pruning). Similar to Bayesian classification the quality of the model is

Modeling techniques used in designing MMP-2 inhibitor

judged by several matrices such as sensitivity, specificity and Mathew's correlation coefficient (Cai et al, 2017).

5.2.2.3 LDA QSAR study

Like the Bayesian classification formalism, the linear discriminant analysis (LDA) is another widely used classification technique, which classifies the compounds as active or inactive. This is a pattern recognition method. Here the multi-dimensional variable space is divided into two Linear Discriminant Function separated hyperplanes and this is how the data matrix of dependant variable is divided (Roy et al, 2015).

5.2.3 Hologram QSAR study

Hologram QSAR or HQSAR is a newer QSAR technique which implies the idea of using molecular fragments as independent property to calculate the biological activity. This is extremely important where no protein drug interaction has been studied. Here the 2D structure of the molecule is first broken down into several fragments and an array is generated by assigning certain integer to the fragments. This array is used to generate the model by using Partial Least Square (PLS) technique. The PLS derived equation is

$$A_i = C + \sum_{i=1}^{l} X_{il} \, C_{il}$$

Leave One Out or LOO cross validation is applied to study the robustness of the model. It has been found that this method is equally predictive as 3D-QSAR technique (Ajmani et al 2009).

5.2.4 Pharmacophore Mapping

By definition, pharmacophore can be simply said the minimum structural feature for a molecule to show a specific biological activity. The pharmacophore is an abstract concept which requires

45

Modeling techniques used in designing MMP-2 inhibitor

to identify the minimum clusterization of steric and electronic feature of a molecule (Wermuth et al 1998). There are a number of methods of creating pharmacophore models such as HypoGen and HipHop. Several pharmacophoric features such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic aliphatic, positive charge, negative charge etc. are used and then through the above mentioned algorithm, it is tried to match it with the several parts of the molecular structure to construct a model. The whole model is validated by analysing the cost-effectiveness and predictive R^2.

5.2.5 3D QSAR study

Unlike the classical 2D QSAR methodology, 3D QSAR study determines the activity relationship between the spatial properties of a molecule with its biological activity. Such spatial property includes steric, electrostatic, hydrophobic, hydrogen bond acceptor, hydrogen bond donor. Initially, the most active compound is identified and its energy is minimised. The alignment is generated in a 3D space and a probe is used for calculating the descriptors. The descriptors are then used to generate a a partial least square (PLS) model and Leave One Out (LOO) and Leave Many Out (LMO) cross-validations are carried out (Roy et al 2015).

5.2.5.1 CoMFA Technique

Cramer *et al.* (1988) developed the Comparative Molecular Field Analysis (CoMFA) formalism. The molecules are aligned on a molecule with some definitive structure. This can either be the most active form or the crystal structure present. Also is come cases, docking based structural alignment can also be performed. A grid of 2Å size is generated around the compound and a number of spatial properties, mainly the steric and electrostatic properties are calculated by using a molecular probe. The binding affinity is then predicted using the field value. The result of

Modeling techniques used in designing MMP-2 inhibitor

CoMFA is expressed in a graphical form called contour map. The preferred tool for calculating CoMFA is SYBYL.

5.2.5.2 CoMSIA Technique

CoMFA was farther modified to generate Comparative Molecular Similarity Indices Analysis (CoMSIA). This method uses all the properties of CoMFA and a solvent-exposed hydrophobicity is also calculated. Therefore, the major descriptors are steric, electrostatic, hydrophobic, hydrogen bond acceptor and hydrogen bond donor (Cramer et al 1988)

5.2.5.3 Open3DQSAR study

Open 3D QSAR is an open source and freely accessible tool developed by Tosco and Balle (2010). The concept of development of the model is similar to CoMFA and CoMSIA but the algorithm is where it stands out. Several molecular graphics software such as PyMOL, MAESTRO, Discovery Studio etc. is used to create a Contour Map or the visualisation model.

5.2.6 RI 4D-QSAR study

Receptor Independent 4D QSAR study is used to determine pharmacophoric feature of a compound solely based on the ligand itself. Here, Grid Cell Occupancy Descriptor is used to determine the molecular properties and Genetic Function Approximation- Multiple Linear Regression (GFA-MLR) formalism is used to create the model (Roy et al, 2015).

5.3 Structure based designing principle:

QSAR and chemometry is a ligand based approach where, inspite of the unavailability of a proper protein structure, a drug discovery method can progress. The concept of a well defined molecular structure can have a great amount of value addition for a medicinal chemist.

Modeling techniques used in designing MMP-2 inhibitor

5.3.1 Molecular docking studies

Molecular docking is a method of insertion of a small molecule, i.e. a probable bioactive drug, into a protein structure. In general this process becomes easy when the structure comes with a co crystal molecule along with it. The co-crystal model serves as a template to it. A bunch of target molecule including the separated co crystal is first subjected to energy minimisation followed by experimentation with multiple poses. The final pose is determined using the global minimum (Root mean square distance) RMSD value. The cross validation of the experimentation is done by the RMSD of the actual co-crystal structure form and the docked pose of the same structure.

5.3.2 Molecular dynamics studies

Unlike molecular docking, molecular dynamics deal with a moving molecule or a projectile. The projectile moves through a trajectory, attaches with the protein and finally releases from the protein. The observation of this particular experiment is to understand the possible attachment time in the drug receptor interaction.

5.4 QSAR and other ligand-based approaches:

Molecular modelling using the ligand-based approaches is extremely reliable as it does not depend on any available crystal structure. The available proteins structure of MMP-2 has not been quite perfect in spite of having statistically acceptable entries. QSAR, both 2-D and 3-D, has the potential to explore a wide array of substructural feature and predict novel molecule from scratch.

Kumar and Gupta (2003) were one of the pioneering researchers to report the 2D- QSAR study to find the interrelationship between the MMP-2 inhibitory activity and Kier's valence molecular

Modeling techniques used in designing MMP-2 inhibitor

connectivity index ($^1\chi^v$) of these derivatives and the electrotopological state of nitrogen (S_N) and sulphur (S_S) atoms using a few hydroxamic acid derivatives. An indicator variable which had a binary value depended on the presence of C_6F_5 or 3-CF3- C_6H_4 groups was also used as a dependent variable. Among several established models, the 2D-QSAR model for MMP-2 inhibitors showed a good statistical significance ($R= 0.928$ and $R^2_A = 0.84$). The authors pointed out a few significant SAR derived from the model. The presence of connectivity index might have a direct relationship with the hydrophobicity of the substituent R group (**Figure 5.2**). As more saturation in the group structure was found to be the favourable factor and lesser value of electronegativity will result in the higher value of the descriptor which is favourable for MMP-2 inhibitory activity.
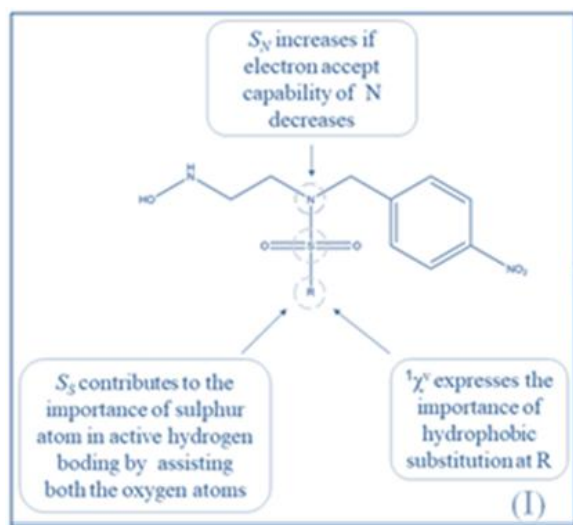
Figure 5.2: Structural exploration of some hydroxamic acid

It was also explored that the electron donating property and the electron accepting property of the sulphur atom and nitrogen atom respectively will produce an acceptable energy level for hydrogen bonding of the sulphonyl group. Also, the presence of a fluorine-substituted substitution can be beneficial. However, the MMP-2 inhibitory potency model showed several

Modeling techniques used in designing MMP-2 inhibitor

similarities with the MMP-9 inhibitory model, thus lacking the selectivity of these compounds over MMP-9.

A series of 4-aminoprolines and its derivatives were again explored by Gupta *et al.* (2003). This had similarities with their previous work in 2002. The motif of the study was to establish two different models for molecules with four different moieties and also to find the inter co-relationship between the biological activity and Kier's valence molecular connectivity index ($^1\chi^v$) of the derivatives and the electro topological state of the sulphur (S_S) atoms. The first series of nine compounds were used to construct the QSAR model using only the $^1\chi^v$ descriptor. The statistical parameters were well within the validation range ($R = 0.863$ and $R^2_A = 0.710$). This model suggests that the most important influencing factor of the structure of this moiety is the higher hydrophobicity of the substituted 4-amido group. On the other hand, the other two moieties (C, D in **Figure 5.3**) showed dependency on both of these descriptors. However, in even though the models show a good statistical significance, the small pool of data (n=9) makes it hard to determine any concrete inference for both of these models.
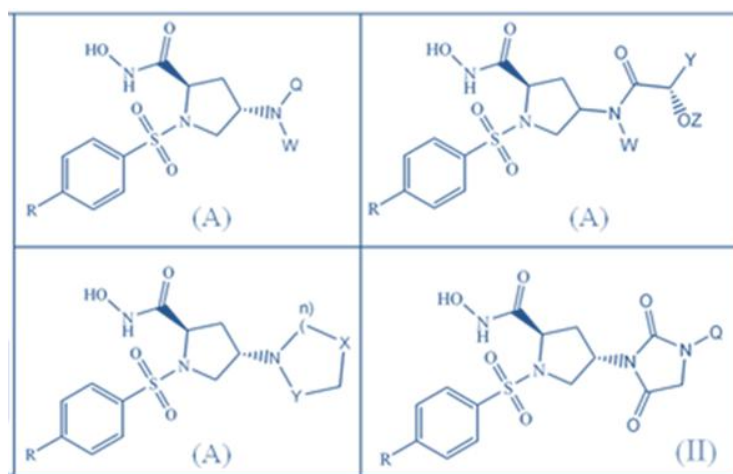

Figure 5.3: Structure of some 6-amino proline derivatives

Modeling techniques used in designing MMP-2 inhibitor

De Melo (2006) performed a number of QSAR study using PLS method on 31 cinnamonyl pyrrolidine derivatives as potential MMP-2 inhibitors. After initial variable calculation, these were reduced based on Ordered Predictor Selection method. The descriptors that were finally selected for setting up the QSAR models were, molecular softness (*SOFT)*, second eigenvalue derived from a resonance integral weighted edge adjacent matrix (*EEig02r)*, component vector to overall polarisability in X-axis (α_{xx}), partial charge of 10th atom by natural bond order method (NBO) *q10NBO*, partial charge of 2nd atom by natural bond order method (*q2NBO)* and E-state topological index of amide group (*SsssN (oth))*. This model had an R^2 value of 0.783 and Q^2_{LOO} value was 0.618, which showed an acceptable score of prediction. The researcher finally concluded that an atom containing a π electron may be beneficial at the oxy- substitution (R$_3$ position; **Figure 5.4**) at the pyrrolidine group. At the same time, modifications at the nitrogen atom at the pyrrolidine group showed scope for a subsequent cinnamonyl-pyrrolidine derivative.
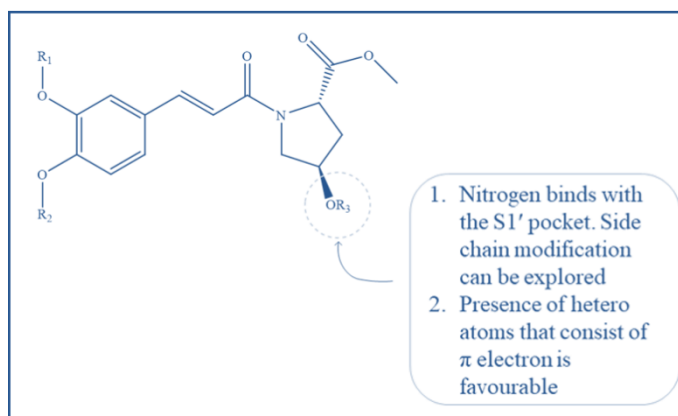
Figure 5.4: Observation of Cinnamonyl pyrolidine scaffold

A QSAR model with autocorrelation descriptors was designed by Fernandez and other (2006). Molecules used the development of this particular model were N-hydroxy-2-[(phenyl-sulfonyl)amino]acetamide (HPSAA) derivatives (scaffolds are shown in **Figure 5.5**). A mixed

Modeling techniques used in designing MMP-2 inhibitor

number of techniques were used for variable reduction. Two unique techniques, Genetic algorithm and Bayesian Regularised Artificial Neural Networking (BRANN) were consecutively applied to the dataset. Orthogonalisation process proposed by Randic was applied to the final model and the high correlation values of different activities of these molecules were eliminated. The positively correlated descriptors were based on the Geary autocorrelation factor of weighted by polarizability (*GATS2p*) and Sanderson electronegativity (GATS6e). On the other hand, 4th eigenvalue and 8th eigenvalue of Moran autocorrelation factor (*MATS4m* and *MATS8m* respectively) when weighted by mass and 3rd eigenvalue when weighted by van der Waals volume (MATS3v). This model has an R^2 value of 0.808 and Q^2_{LOO} of 0. 721. As a mechanistic approach of explanation of autocorrelation descriptors might be troublesome, these authors implored the method of descriptor contribution to compare the contribution of each of the descriptors. In this way, it was concluded that higher polarisability of the substitutable group should enhance the activity along with electronegativity of the group.
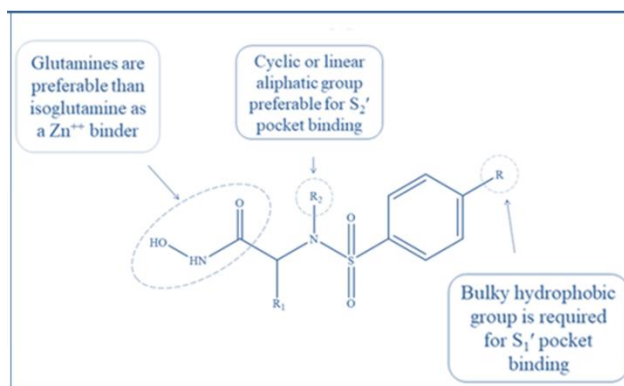
Figure 5.5: Structural features of N-hydroxy-2-[(phenylsulfonyl) amino]acetamide

Jamloki and his colleagues (2006) had a holistic approach towards a number of matrix metalloproteinase inhibitors. They created individual models for MMP-1, MMP-2, MMP- 8 and

Modeling techniques used in designing MMP-2 inhibitor

MMP-9 inhibitory properties of 27 different 5- amino-2- mercapto-1,3,4- thiadiazole based molecules (**Figure 5.6**).
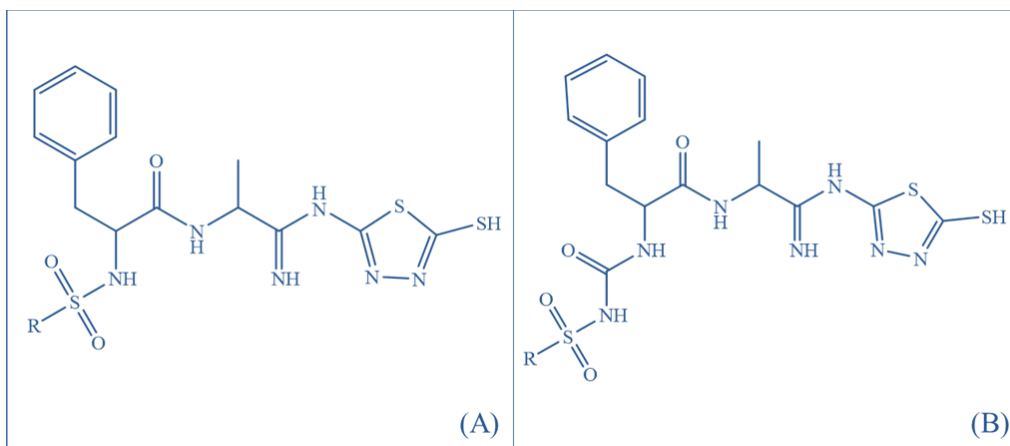
Figure 5.6: Common scaffold of 5- amino-2- mercapto-1,3,4- thiadiazole based molecules

It was inferred from the work that for the presence of terminal branching at the substitution region will be favourable for the MMP-2 inhibitory activity of the molecule. The 2D QSAR model was based on topological and fragmented descriptors of Molecular Operating Environment (MOE) and its QSAR Module. It was found that the binding affinity of MMP-2 could be explained using $^3K_\alpha$ (Kier α modified third kappa shape index), a_nF (number of fluorine atoms) and an indicator variable I which replaces the topological descriptor $^1\chi^V$ (atomic valence connectivity index of order 1). The robustness of the regression model was acceptable ($R^2 =0.797$). the MLR models showed promising insight towards the MMP-2 inhibitory activity.. The regression model showed a positive coefficient for the $^3K_\alpha$ descriptor which denotes the tendency of a molecular branching to be placed onto the extreme sides and can be inferred that this property may increase the inhibitory activity of these compounds. The other interesting factor for activities of the molecule can be explained using the second variable a_nF which quantitatively expresses the number of fluorine atoms present in the molecule. A higher number

Modeling techniques used in designing MMP-2 inhibitor

of fluorine atoms would mean more amounts of hydrogen bonds with the active sites of the various subtypes of matrix metalloproteins. The final indicator variable *I* suggested that the sulphonamide moiety closed to an amide group would result in better functionality. The researches, however, finally concluded that the molecule, 5- amino-2- mercapto-1,3,4-thiadiazole derivatives had comparatively lower selectivity towards MMP- 2 when compared to MMP- 8 and MMP- 9 but better compared to MMP- 1 (Jamloki et al 2006).

Kumaran and Gupta (2007), performed 2D-QSAR studies on piperidine aryl hydroxamic acid analogues (**Figure 5.7**). the calculated hydrophobicity (*ClogP, ClogP^2*) was found to be important for the MMP-2 inhibitory potency of these compounds. Some indicator variable I_1 was used for any R_1 substitution and I_{N1} was used for the presence of NH_2— group at the R_1 position. In the final model, the R^2 value was 0.910 and the R^2_{cv} was 0.860. The model shows that though the hydrophobicity is important for increased activity of these inhibitors, highly hydrophobic characteristics may not be good for the inhibitor potency. This was inferred as *ClogP* showed a positive contribution but the second order of the variable (*ClogP^2*) contributed negatively. In addition, R_1 substitution with a 4- substituted phenoxy piperidinyl group may increase the activity and due to negative coefficient along with I_{N1} may explain that substituent with –NH-group will be decreasing the activity of these compounds (Kumaran and Gupta, 2007).

Modeling techniques used in designing MMP-2 inhibitor
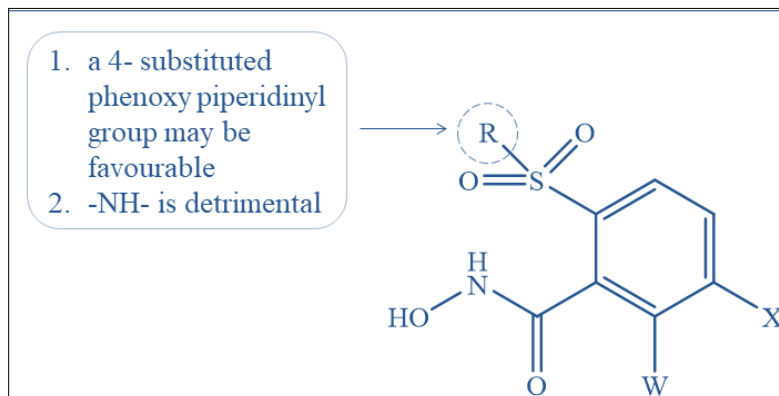

Figure 5.7: Piperidine sulphonamide aryl hydroxamic acid and its explored analogues

67 derivatives of pyrrolidine were subjected to Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Analysis (CoMSIA) 3D-QSAR studies with partial least square (PLS) regression by Zhu *et al* (2009). The squared cross-validation scores (Q^2) were 0.757 and 0.843 respectively for CoMFA and CoMSIA. The 3D contour mappings suggested some interesting properties of the pyrrolidine derivatives (**Figure 5.8**).
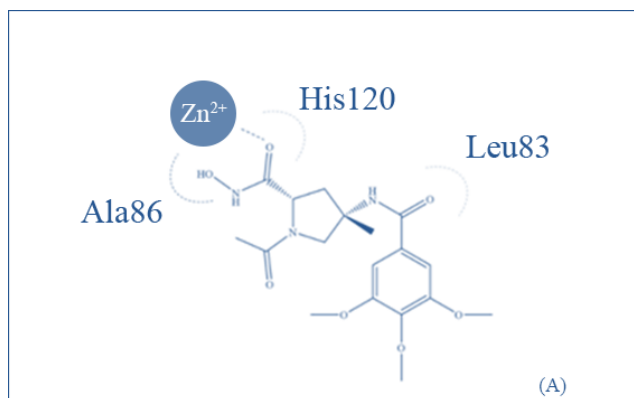
Figure 5.8: Interaction of the highest active molecule with MMP-2 crystal structure

These authors predicted that the derivatives show better potency with steric substitution at the R_1 and R_3 (**Figure 5.9**) regions of the scaffold and produce better enzymatic inhibitory activity. While in the R_2 position such bulky group is not favourable, the COMFA analysis shows that the

Modeling techniques used in designing MMP-2 inhibitor

negatively charged groups would produce a better biological activity. On the other hand, the CoMSIA studies add that the P_1 and P_3 need to be hydrophobic in nature for the better activity while at the R_2 position a hydrogen bond accepting group results in a potent metalloproteinase inhibitory substance. CoMSIA and CoMFA studies unanimously infer that a hydroxamate will be the most acceptable group among the training set compound. The researchers also validated the training set with 17 different test compounds to cross-validate the two QSAR models and both of them were statistically satisfactory (Zhu et al, 2009).
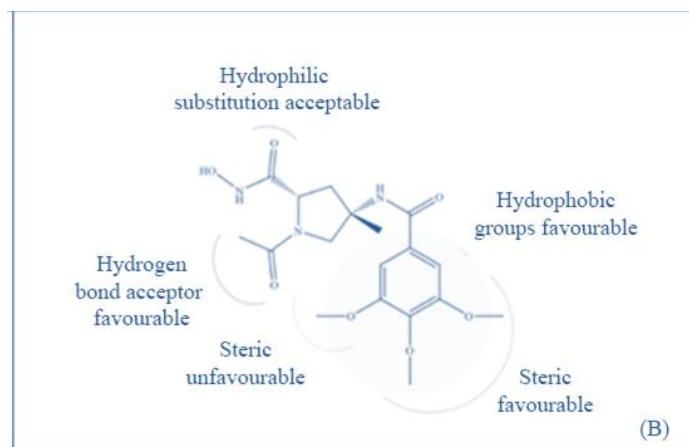
Figure 5.9: Molecular properties derived from 3D QSAR (CoMFA, CoMSIA) studies

Nicoletti *et al.* (2012) from the Department of Pharmacy, The University of Bari, Italy, selected a series of rationally designed derivative of 5- hydroxy 5 substituted pyrimidine -2,4,6 –triones (**Figure 5.10**) as inhibitors of multiple subclasses of MMP which included gelatinase MMP-2, MMP- 9 and collagenase MMP-8.

Modeling techniques used in designing MMP-2 inhibitor

Presence of a 4-methoxy phenyl substitution is selective to MMP-2
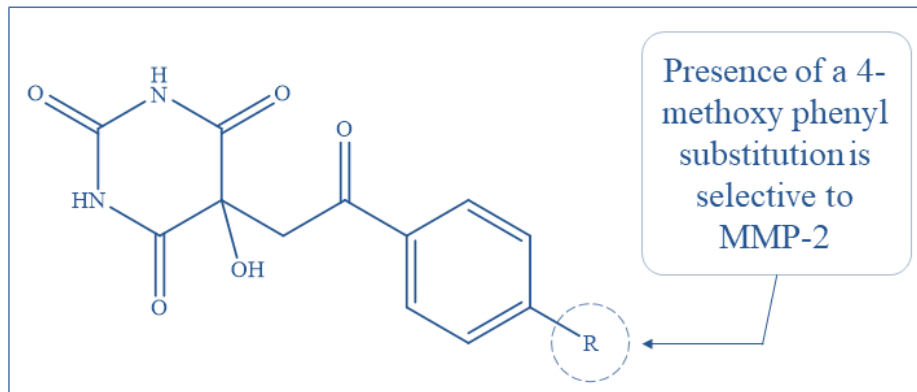
Figure 5.10: 5- hydroxy 5 substituted pyrimidine -2,4,6 – trione derivatives

Synthesis of these derivatives and biological assays were performed by these workers. After that molecular docking with some of the best active molecules with GOLD 5.1 program and QSAR modelling was done for the same. While discussing the theoretical chemistry, these authors, initially, had structured the QSAR model with two different independent variables such as Hammet's substitution constant (σ_p) and the Hansch's substitution constant (π). While the positive coefficient for σ_p denotes a favourable para-substitution, the negative coefficient at the π indicates the unfavourable condition for a non-polar environment. For the better acceptability and more lucid explanation of the 2D-QSAR model, the π was replaced with substituted molar refractivity or *MR*, the variable contributed positively to the model. The model thus obtained had better squared residual values of 0.565 compared to 0.520 of the previous one. This well explains the concept of interaction at multiple hydrophilic regions of the MMP-2 inhibitory site. With the docking interaction with MMP-2 barbiturate complex derived from PDB structure, 1HOV helped in further supporting the 2D-QSAR model. It was observed in the docking study that in the S1´ pocket of the enzyme, one side chain of Glu121 (naming convention according to 1HOV) interacts with the N1 oxygen atom of the compound by hydrogen bonding while another hydrogen bond interaction takes place between C4 and Ala84. Another observation shows that

Modeling techniques used in designing MMP-2 inhibitor

for the best active molecule, the biphenyl group with a long para-COCH$_3$ group binds with the guanidine group of Arg149. This docking study was useful to understand the possible interaction between several fragments of the inhibitor molecule and the enzyme (Nicolotti et al, 2012).

In Jadavpur University, India, Halder along with his co-workers explored 202 structurally diverse compounds in 2013 (Halder et al, 2003). A good number of molecular modelling tools and statistical approaches were selected as a part of their research work. On 2D QSAR methods, they performed regression based QSAR as well as recursive partitioning and Bayesian modelling of classification QSAR study which further intrigued them into pharmacophore mapping and 3D QSAR methods such as CoMFA and CoMSIA. They further modified the 2D QSAR models with stepwise regression, genetic algorithm and spline modelling. The researchers have summarised their work that the biological activity of these molecules considered in their work, showed a strongly positive response to *Jurs$_{FPSA3}$* which is the ratio of atomic charge weighted partial surface area (*PPSA_3*) and solvent accessible surface area (*SASA*) in all of the three models. This represented the probable favourability of the charged fractional partial surface area on the biological activity. Moreover, highly electron-dense polar regions that acts a hydrogen bond acceptor feature might result in the better biological activity of these experimented molecules. Pharmacophore mapping and Bayesian classification studies both substantiate the hydrogen bond accepting feature of these compounds. In addition, the presence of five-membered rings, pyrrolidine in the selected dataset, shows higher hydrophobic feature while being effective for the enhanced biological activity.

The Chinese research group from Nanjing comprising Qiu *et al.* (2014) worked on benzosulfonamide benzenesulfonates for exploring a set of chemicals with potency for selective

Modeling techniques used in designing MMP-2 inhibitor

inhibition of MMP-2 enzyme as a series of the anti-metastatic agent. They synthesised 27 compounds based on the above-mentioned scaffold and comparative ex vivo studies were done against A549, MCF-7, HepG2 and Hela cell lines with gefitinib and celecoxib as standard reference drugs. For a better explanation of the properties of these compounds based on this class, both docking and 3D- QSAR studies were performed by the research groups. The docking study of the highest active compound was elaborate and pointed out important interaction between the molecule and the 1QIB crystal structure of the MMP-2 derived from PDB. It was found that the best active compound (**Figure 5.11**) creates a bonding interaction with TYR155 and GLY152 (named as per 1QIB) through charge transfer with the nitrile atom. In addition to this, the hydrogen bond interaction of the Zn^{++} cation and the oxygen atom of O=S=O group was evident as well.

Modeling techniques used in designing MMP-2 inhibitor
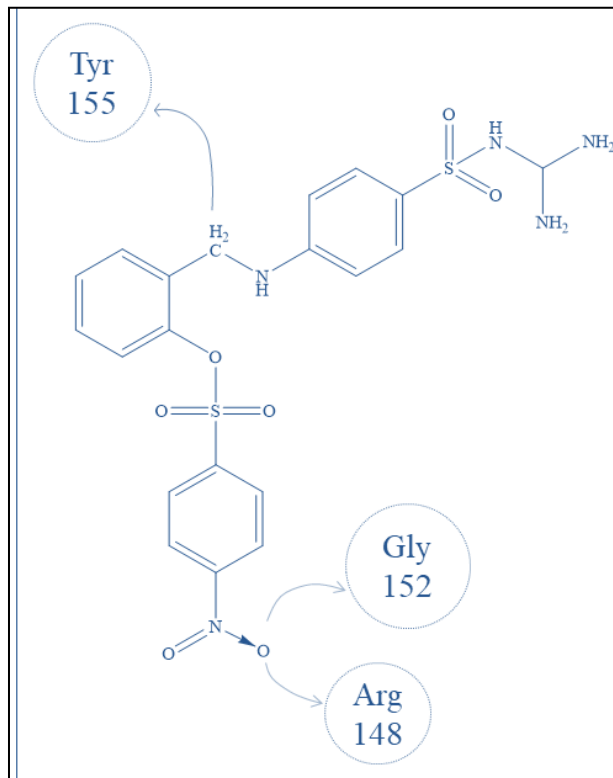
In 2014, a work by K Medeiros Turra and his group from Department of Pharmaceutical Sciences, University of Sao Paolo, Brazil reported a class of b-N-biaryl ether sulphonamide-based hydroxamates **(Figure 5.12)** as an antineoplastic agent as MMP-2 inhibition by pharmacophore mapping. The technique reported by these authors were Receptor-Independent (RI) 4-Dimensional Quantitative Structure-Activity Relationship (4D- QSAR) on 38 molecules (train = 30, test = 8) (Turra et al, 2014). HyperChem 8.0 was used for the minimisation of the energy of these compounds using the MM+ force field. NMR protein data from Protein Data Bank identification code (PDB ID) 1HOV was employed to determine the best-fitted 3D structure. Another programme, MOLSIM also optimised the geometry of the designed inhibitors.

Modeling techniques used in designing MMP-2 inhibitor

The descriptors were determined by Grid Cell Occupancy Descriptors (GCOD) and the statistical model was constructed using Genetic Function Approximation Multiple Linear Regression (GFA − MLR) formalisms. These models were then validated using two methods -LMO, leave many out and Y randomisation processes. It was found that the best model (Model 4) had a squared residual value (R^2) of 0.93, cross-validation value (Q^2) of 0.88 and low error value which proved the robustness of the model. Finally, it was concluded that among all these molecules, the zinc binding group has the minimum required portion for the activity and a hydrophobic group is required for its best fit into the S_1 pocket of the MMP-2 protein. One important aspect of this work, which also suggests 3 new compounds based on the optimum presence of the grid cells.
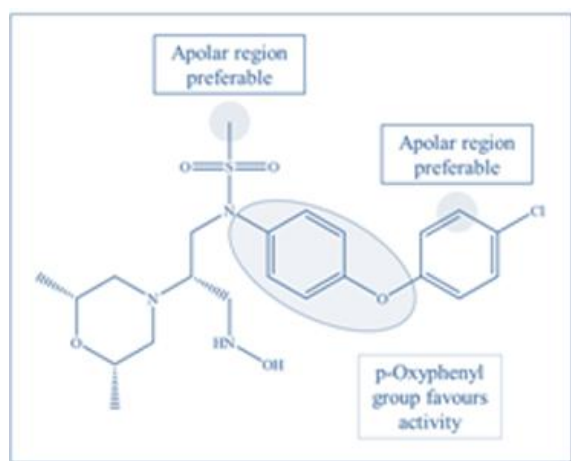
Figure 5.12: Best active molecule and interaction region

Abbasi *et al.* (2015) selected some potential MMP-2 inhibitors with L-tyrosine scaffold (**Figure 5.13.A**). They performed a wide variety of computational methods to establish a robust 2D-QSAR model. 2D-QSAR model using conventional descriptors and semiempirical descriptors (hardness, softness, electronegativity and electrophilicity) were used for designing the model. The total number of molecules that were taken into account were 30 with 20% in the test set and rest in the training set. All these selected molecules were docked into the MMP-2 crystal

Modeling techniques used in designing MMP-2 inhibitor

structure for better understanding of the inter-molecular attraction force between the ligand and the proteins.
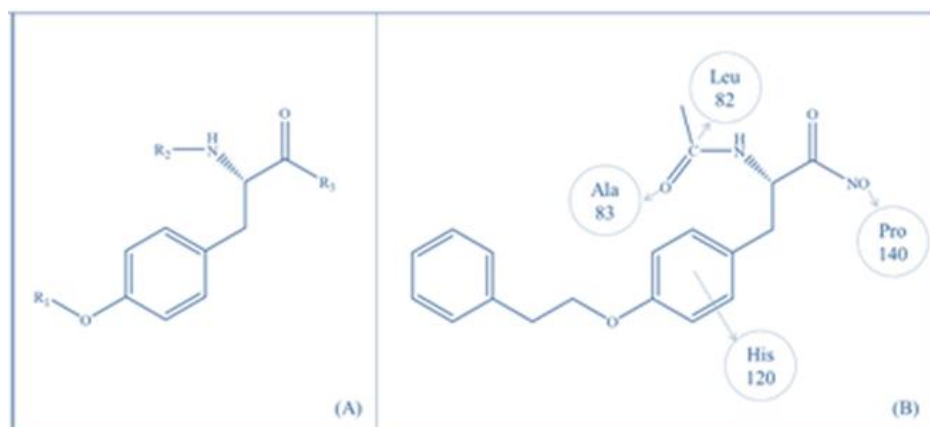
Figure 5.13: A) Skeletal scaffold of L-tyrosin derivatives (B) Docking interaction of the molecule with least docking energy with MMP-2 (PDB ID: 3AYU)

These molecules were initially drawn using Hyperchem 8.0 software and they were geometrically stabilised using empirical molecular mechanics (MM+) force field. Further optimisations of these molecules were done using semi-empirical PM3 method of Gaussian98W. Descriptors for construction of models were calculated by 3 different software – Hyperchem 8.0 for chemical descriptors, DRAGON for all 0D, 1D, 2D, 3D class of descriptors and Gaussian98W for Parameterized Model number 3 (PM3) based semi-empirical descriptors. One of the most important steps of designing the 2D-QSAR modelling is to select the best model with the least amount of descriptors and best statistical validation. This, however, is best to balance between these two factors. Initially, the model was established using the MLR method. The descriptors thus selected were, *IC1* (Information content index (neighbourhood symmetry of 1-order)), *Mor24m* (Radial Distribution Function -3.5 / weighted by atomic Sanderson electronegativity), *RDF115m* (Radial Distribution Function – 12.0 / weighted by atomic masses), *SP20* (Shape profile no. 20) and *Mor15e* (3D-MoRSE –signal 15 / weighted by atomic

Modeling techniques used in designing MMP-2 inhibitor

Sanderson electronegativity). These 24 molecules reproduced an R^2 of 0.958 and Q^2_{LOO} of 0.921. However, the group performed GA- PLS to explore the probability of a more robust statistical model. The result was a model with *IC1* (Information content index (neighbourhood symmetry of 1- order)), *Mor24m* (Radial Distribution Function -3.5 / weighted by atomic Sanderson electronegativity), *Mor15e* (3D-MoRSE –signal 15 / weighted by atomic Sanderson electronegativity) and *Mor32e* (3D-MoRSE –signal 32 / weighted by atomic Sanderson electronegativity). This model had R^2_c value of 0.932 and Q^2_{LOO} of 0.900. AutoDock 4.2 was the choice of software selection for the docking studies of the L-tyrosine derivatives (Figure 15.B). The docking interaction substantiated that the oxygen atom of the carbonyl or sulphonyl groups contributed to the hydrogen bond interaction with Leu82 and Ala83. The main substituents, R_2 and R_3 were two of the vulnerable sites of the scaffold; it showed -OH / -NHOH having better hydrogen bond interaction. R_1 on the other hand provided the better activity with more lipophilicity and optimum amount of chain length for best fitting. A phenylethyl group provided the best activity of the compounds for the π stacking interactions between the phenyl group and Phe86. A -CH_2-CH_2- linker provided the proper chain length for the hydrophobic pocket of MMP-2 (**Figure 30**).

Modeling techniques used in designing MMP-2 inhibitor
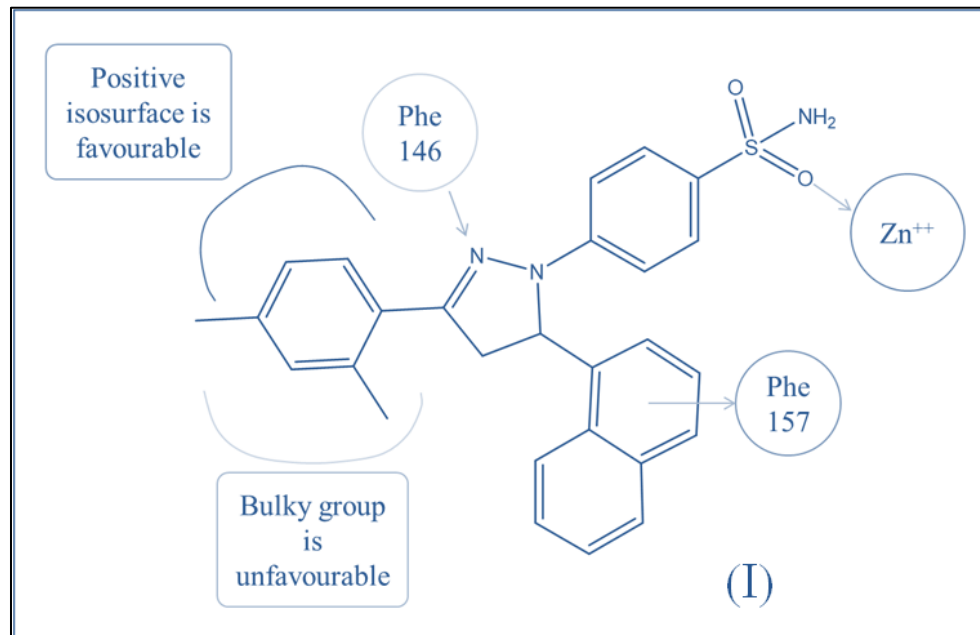
Figure 5.14: Docking interaction between the best active molecule and MMP-2

Yan *et al.* (2015) reported a series of dihydropyrazole moiety after designing, synthesis, inhibitory activities and 3D-QSAR analysis of 25 different inhibitors of both MMP-2 and MMP-9. For computational chemistry, a 3D-QSAR study using Discovery Studio was performed to explore the potential physicochemical parameters for upgradation of the scaffolds at various favourable and unfavourable regions. The docking study was performed to support the probable interaction with a 3D crystallographic structure of MMP-2 (PDB ID: 1QIB).

For MMP-2, the researchers used the Discovery Studio 3.5 to find the docking interaction between all of the 25 molecules and X-Ray crystallographic structure of MMP-2 derived from PDB ID: 1QIB. The RMSD score was minimised to 0.3 Å and the Quanta / CHARMM force field was minimised. The optimised docking poses of these molecules were aligned for lowest CDOCKER_INTERACTION_ENERGY for 3D-QSAR study using PLS regression method. The model was found statistically significant with $R^2 = 0.922$, $Q^2_{LOO} = 0.469$, $R^2_{pred} = 0.834$, *RMSE*= 0.249, *MAE*= 0.214 and *F value* of 101.913. Though the Leave-One-Out (LOO) cross-validation

Modeling techniques used in designing MMP-2 inhibitor

of the model was below the acceptable level of 0.5, the R^2_{pred} value reported an acceptable prediction. The 3D- QSAR model was based on steric and electrostatic effect, polarity and van der Waal interaction and their favourable and unfavourable zones were identified with contour mapping. Through the docking study, it was observed that the best fit one of these molecules (**Figure 5.15**) has 3 distinct interactions with the catalytic site of the protein.
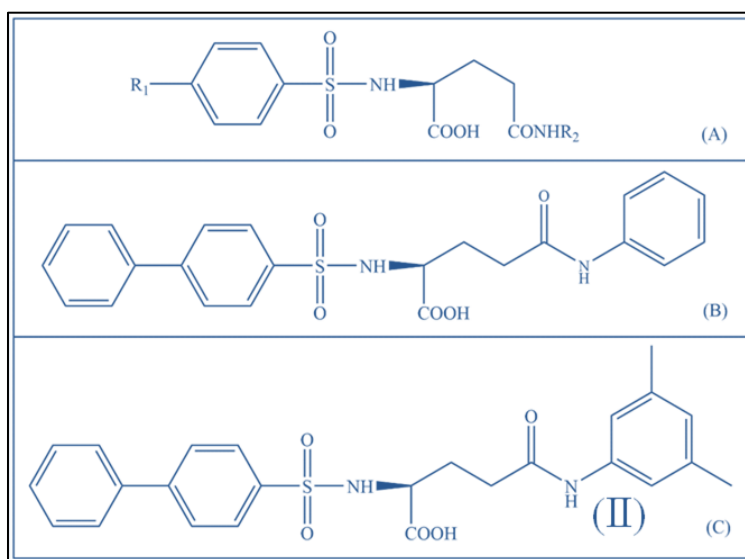
Figure 5.15: (A) Common scaffold of the designing molecules (B) The compound forms a hydrogen bond interaction with the -NH of Phe146, a π-cationic interaction with Phe157 and the oxygen atom forms a coordinate bond with the Zn^{2+} ion. The 3D- QSAR result can be interpreted that the substituted phenyl ring at the 3 positions mostly had electronegative region and negative relation with the Van der Waal interactivity. This provides a future rationale for designing a newer variety of molecules (Yan et al, 2015).

The proteins MMP-2 and MMP-9 show numerous structural similarities. Designing a selective MMP-2 inhibitor is required for efficient management of cancer. Adhikari and co-workers (2016) had designed, synthesised and performed biological assay for obtaining a better selective MMP-2 inhibitor. Common scaffolds are mentioned in **Figure 5.15**. Several computational methods were used for *de novo* designing of a few novel compounds. A vast arsenal of computational formalisms, which are: - 2D QSAR, 3D QSAR, Pharmacophore mapping, Bayesian classification, Recursive partitioning, *De novo* fragment-based lead modification, molecular

Modeling techniques used in designing MMP-2 inhibitor

mechanics combined with generalised Born and surface area salvation (MM- GBSA) and molecular dynamics simulation were used. The total dataset contained 214 molecules which were divided into the training and the test sets according to the formalism applied on to them. For example, 25% of the total molecules were selected for the test set while in case of pharmacophore modelling, 30 structurally diverse compounds were considered as the training set and rest were selected as a test set.

Regression-based 2D QSAR study was done by stepwise MLR method to identify 8 different descriptors, *BIC5* (bond information content neighbourhood symmetry of 5th order)*, HOMA* (harmonic oscillator model of aromaticity)*, RDF110u* (unweighted radial distribution function − 11.0)*, C- 026*(the number of substituent groups which are attached to the phenyl ring)*, Jurs_RPCS* (relative positive charged surface area)*, STD(O,S)* (sum of topological distance between oxygen and sulphur atoms)*, PEOE2* (partial equalisation of orbital electronegativity of second order) and *#rtvFG* (number of reactive groups in molecule). The model showed an R^2 value of 0.766, Q^2= 0.750 and R^2_{Pred} = 0.690 proved the MLR model to be a valid model for statistical prediction. The ligand-based pharmacophore modelling used Hypogen formalism and low RMSD value was considered (<1.50) for identifying good alignment. The hypothesis followed flexible mapping to get most of the considering compounds to be fitted into the model. The correlation coefficient was found to be 0.650.

CoMFA and ComSIA studies were depicted as contour maps. The PLS model was generated with 4 different components and five compounds considered (2 from CoMFA and 3 from CoMSIA) as outliers in the model. For classification based modelling, these compounds were classified initially into 'Selective' (33) and 'Non-Selective' (116) which were farther randomly divided into

Modeling techniques used in designing MMP-2 inhibitor

the test and the training sets. Recursive partitioning model and fingerprint-based Bayesian modelling were done to elucidate the structural requirement for the biological activity of these molecules. Bayesian modelling was done based on eight descriptors such as lipophilicity (*AlogP*) molecular weight (MW), number of hydrogen bond acceptor, number of ring (*nR*), number of hydrogen bond donor (HBD), molecular polar surface area (*MFPSA*) and a fingerprint feature *ECFP_6* (extended connectivity fingerprint of maximum diameter 6). The leave one out (LOO) method and the receiver operating characteristic curve (ROC) were the choices of validation techniques. The true positive, false positive, true negative and false negatives were identified and the model was validated using the test set compounds. It was found to have a ROC value of 0.833 that showed acceptable predictability of the model.

The Recursive Partitioning, on the other hand, used *ES_sum_ssO* (electrotopological state indices of two single bonded oxygen atom), *Jurs_RPCG* (relative positive charge) and *CHI_V_2* (second order Kier & Hall valence − modified connectivity index). A new molecule and its subsequent derivatives were designed by de novo designing and docking study were performed to calculate the highest MM- GBSA binding energy (-57.81 kcal/mol). The molecule was synthesised and it was found that most of the derivatives are selective for MMP-2 over MMP-9. Most of these compounds passed the drug-likeliness study using both Lipinski Rule and Veber rule supporting the predictability of the computational methods used by the workers. This work stands fairly in a unique way because of its *de novo* design. Though *de novo* was adopted earlier by other research groups for designing MMP-2 inhibitor, here the new compound was designed, validated against the existing model and farther correlated the predicted structure with enzyme inhibition assay and two important molecules were found.

Modeling techniques used in designing MMP-2 inhibitor

Zheng *et al.* (2017) performed a 3D QSAR study using CoMFA and CoMSIA tools on a series of galardin derivatives (**Figure 5.16**) as an MMP-2 inhibitor. They mainly discussed the steric effect which is present near the P1' substituent of the compounds. SYBYL 6.8 was used for alignment and calculation.
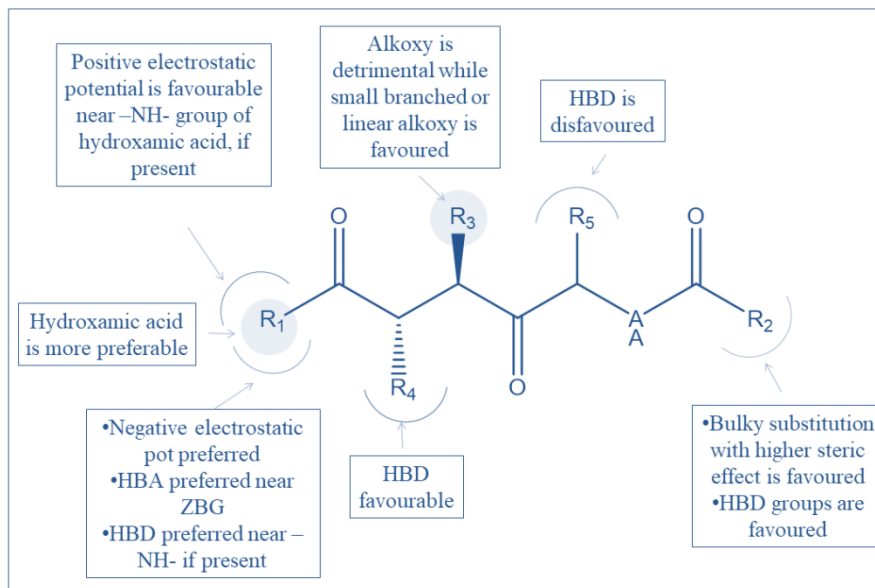
Figure 5.16: Common scaffold of gallardin derivatives and major findings through 3D QSAR studies

For CoMFA calculation, the grid spacing was set to 2.0 Å. An sp^3 carbon probe was used and the Gasteigere-Marsili charge was used for these compounds. For the CoMSIA method, the same carbon probe (sp^3) was used with +1 radius, +1 hydrophobicity and +1 charge. For the PLS statistical analysis, the SEE calculated was high even when the R^2 value was 0.830 and R^2_{cv} was calculated to be 0.522 with all 72 test set molecules. However, in their study, the researcher identified 3 compounds to have excessively high inhibitory values and thus they were excluded to construct a better model. This had an R^2 value of 0.970 and R^2_{cv} 0.712 while the SEE was 0.216. The CoMFA contour suggests the presence of favourable steric group near the S1' pocket and the presence of the negative electrostatic contour proximal to the oxygen atom of

Modeling techniques used in designing MMP-2 inhibitor

hydroxamic acid were confirmed by the CoMSIA technique. Some important findings are noted in **Figure 5.17**.
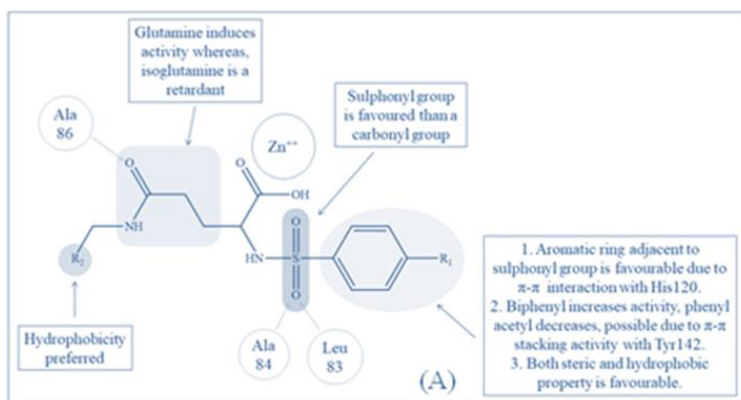
Figure 5.17: Structural features of glutamine derivatives

In 2018, our research team reported a detailed alignment based 3D QSAR study on the chemo-biological aspect of interaction (Adhikari et al, 2017). The common scaffold was detailed in. It was identified by the authors that the previously done 3D QSAR model development by several other research groups led to poor model formation due to less internal cross-validation Q^2 value. These researchers performed pharmacophore based, docking based and Open3DALIGN based 3D QSAR methods on 59 in-house glutamates. Initially, for pharmacophore mapping, ten Hypogen hypotheses were constructed using structural features like hydrogen bond acceptor, hydrogen bond donor, hydrophobic, hydrophobic aromatic, hydrophobic aliphatic, ring aromatic, positive negative aromatic features etc. Initially, the relationship between the biological value and the Fit Value of hypotheses 5 was established. The validation parameters were found to be at an acceptable range ($R^2_A = 0.900$, $Q^2_{LOO} = 0.892$). However, due to less R^2_{PRED} value of 0.492, another variable (*Molecular_SASA*) was introduced which increased the value to 0.582. All of these molecules were subjected to docking to identify the interaction between the molecular substructures and the S1 pocket. It was evident that the biphenyl moiety had π-π stacking

Modeling techniques used in designing MMP-2 inhibitor

interaction with the Tyr142 residue. It also showed hydrophobic interaction with Leu116, Val117, Ala136, Leu137, Ala139, Pro140, Iso141 and Tyr 142.

A 2D-QSAR analysis was also performed using XP descriptors obtained from the docking study and it showed acceptable R^2_A (=0.718) and Q^2_{LOO} (=0.695) values. The docking alignment was further used as a template of 3D-QSAR analysis. A methyl probe was used to calculate the CoMSIA descriptors. The grid parameter was set to 2.0 Å spacing and hydrophobicity, charge and hydrogen bond acceptor value of +1 each. On the other hand, Open3DQSAR study was performed using a 5 Å grid box with 0.5 Å spacing. For both of these cases, the PLS method was used to construct the final statistical model. The researchers finally compared all the 3 alignment methods i.e. pharmacophore based, docking based and open3DALIGN based, based on the several statistical parameters and it was found that the docking based alignment showed the best possible model. The result is briefed in **Figure 5.18**.
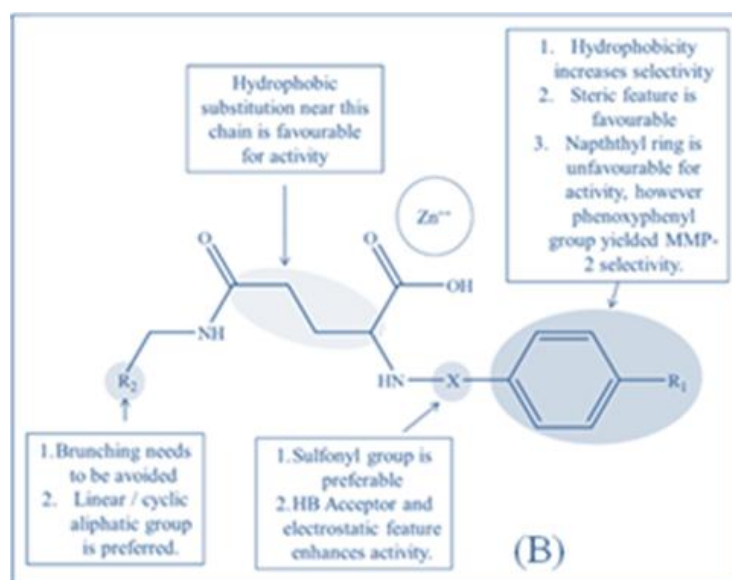
Figure 5.18: Glutamine derivatives used in multi-QSAR model

Modeling techniques used in designing MMP-2 inhibitor

Again in 2018, Adhikari *et al.* worked on 222 different molecules to explore the structural requirements for a biologically active MMP-2 inhibitor (Figure 20) [6]. They worked on regression based 2D QSAR, classification based 2D QSAR, 3D-QSAR and pharmacophore mapping techniques. All the final models were judged using 24 in house glutamates to identify the predictability of the model. In this work, the researches implored a method of lasso regression for reducing the variables from the pool of 674 DRAGON descriptors. The whole dataset was split into training and test set comprising 167 and 55 compounds respectively. The final model showed a good R^2 value of 0.772 and Q^2 value of 0.736 while the R^2_{pred} was calculated as 0.684. For LDA-QSAR, the splitting was different; one set of active and other, inactive, were identified. 113 and 109 inhibitors fit into these categories respectively.

It was also used for Bayesian classification method and ten active and inactive substructural fragments were identified. For pharmacophore based studies, Hypogen algorithm of Discovery Studio (DS 3.0) was used where, two hydrogen bond acceptor, one hydrophobic and one ring aromatic feature were used. Based on the cost analysis and the best correlation coefficient (0.990) the best hypotheses were determined to get the hypotheses 2. On the other hand, CoMFA was performed using PLS and an R^2 of 0.922 was calculated. Nevertheless, other values like the Q^2 and R^2_{pred} were found to be within acceptable criteria. With multi QSAR studies on this 222 molecule, a few important structurally important features were explored. It was suggested that the sulphonamide linker moiety is important as well as sp^2 hybridised carbon with high van der Waals volume enhances the potency of the selected MMP-2 inhibitors.

Another research work was carried out by Adhikari *et al.* (2017) for the exploration of in-house glutamate to identify probable MMP-2 inhibitors. Here, these researchers used regression-based

71

Modeling techniques used in designing MMP-2 inhibitor

2D QSAR, hologram QSAR and 3D-QSAR techniques (CoMFA and CoMSIA). Moreover, a quantitative activity-activity relationship (QAAR) study was performed to find the selectivity towards MMP-2 of these molecules over MMP-9. These formalisms were carried out on 59 different compounds. In 2D QSAR study, initially, a stepwise multiple regression (SMLR) was performed. Apart from this, the machine learning technique of Artificial Neural Networking (ANN) was used to construct a non-linear model. For the SMLR method, the statistical parameters were $R^2 = 0.958$ and $Q^2 = 0.947$ while for the ANN model, it was slightly better with $R^2 = 0.979$ and $Q^2 = 0.963$.

The Bayesian classification identified the good and bad fingerprint using the *ECFP_6* descriptor. 50 different HQSAR models were initially generated and the best model was selected depending on the Q^2 value. As the study used a repertoire of molecular modelling technique, an elaborate inference was able to be drawn. The work suggested that for MMP-2 inhibition, the presence of a glutamine group will be a preferable zinc binding group compared to an isoglutamine structure (Figure 21). On the other hand, while a hydrophobic and bulky group is suitable for S1☐ binding, the linear cyclic aliphatic group will be good for binding into the S2☐ pocket.

Research lab of Jadavpur University (Jha et al 2017) performed multi-QSAR studies on 67 different glutamic acid derivatives as potential matrix metalloproteinase 2 inhibitors. We performed regression based 2D QSAR, classification based LDA QSAR, Bayesian classification, Hologram QSAR, 3D QSAR CoMFA and CoMSIA and also Open3DQSAR studies. Statistical evaluations for each of them were done individually. In the regression-based 2D QSAR study, 4 molecular descriptors were found to be the most effective parameter for the statistical model. These descriptors were, *SpMAD_Dzs* [Spectral mean absolute c*VCH-6* (Valence chain, order 6)]

72

Modeling techniques used in designing MMP-2 inhibitor

and VE3_Dt (logarithmic coefficient sum of the last eigenvector from detour matrix). In this step-wise multiple linear regression model (S- MLR), 41 molecules were taken as the training set and 13 molecules were taken as the test set and 13 molecules were taken for the validation set. Statistical robustness of the model was supported by the *R* squared value of 0.969. The LDA QSAR model was based on five different descriptors, *ATSC2s* (Centered Broto-Moreau autocorrelation - lag 2 / weighted by I-state), *AATSC5s* (Average centred Broto-Moreau autocorrelation - lag 5 / weighted by I-state), *MATS7s* (Moran autocorrelation - lag 7 / weighted by I-state), *MATS3* (Moran autocorrelation - lag 3 / weighted by mass) and *nS* (Number of sulphur atoms). Bayesian classification modelling distinguished the best active compound from the least active counterparts. This also helped to identify biologically active substructure of these compounds.

The HQSAR study reported 48 different models of the fragment distinction for the importance of biological activity and it was found that M46 of the reported model to be the best model with Q^2 value of 0.914. The 3D-QSAR study was done in 3 different ways, CoMFA, CoMSIA and Open3D QSAR. The statistical model based on the R^2 value depicts that the Open 3D QSAR studies ($R^2 = 0.945$) were proved to be a better model for predicting the biological activity when compared to CoMFA ($R^2 = 0.933$) and CoMSIA ($R^2 = 0.923$) studies. 3D QSAR studies helped in identifying the interacting forces responsible for the MMP-2 inhibitory action. The researchers reported that the studies based on the above mentioned QSAR methods along with some earlier reported work found that the presence of biphenylsulphonyl moiety shows better effect than phenylacetyl / naphthylacetyl moieties as per the steric and the hydrophobic characteristics of the

Modeling techniques used in designing MMP-2 inhibitor

S1 pocket. The multi QSAR study unanimously reported the fact that aryl sulphonyl glutamines are more potent MMP- 2 inhibitors than their aryl acetyl isoglutamine counterparts.

The HQSAR and Bayesian classification modelling suggested the importance of biphenylsulphonamido group were better MMP-2 inhibitors than phenylacetylcarboxamido group. The sulphonyl group of these compounds are less flexible than the carbonyl counterparts. Therefore, it has a higher chance to direct the adjacent group into the S1' pocket of MMP-2 enzyme. The carbonyl group, on the other hand, is more flexible and has a tendency to bend towards other directions. On the γ terminal of the glutamic acid, an aryl carboxamido group supported the docking studies of MMP-2 inhibitions (Jha et al 2017).

Docking and other structure-based approaches:

Moroy *et al* (2007) worked on an established MMP-2 inhibitor, Ilomastat (or Galardin). The compound showed excellent pan-MMP inhibitory activity. However, it was rescinded due to its adverse effects in clinical trial. The specificity of these compound towards MMP-2 is entirely dependent of the P1' Group which fits into the unique S1' pocket of the MMP-2. The researchers substituted the P1' group of Ilomastat and performed docking studies to find its binding activity. The carboxyl group creates hydrogen bond with Pro238. On the other hand, the presence of unsaturation in this part will result in lesser flexibility and higher favorability of towards MMP-2. A 2D docking interaction is shown in **Figure 5.19**

Modeling techniques used in designing MMP-2 inhibitor
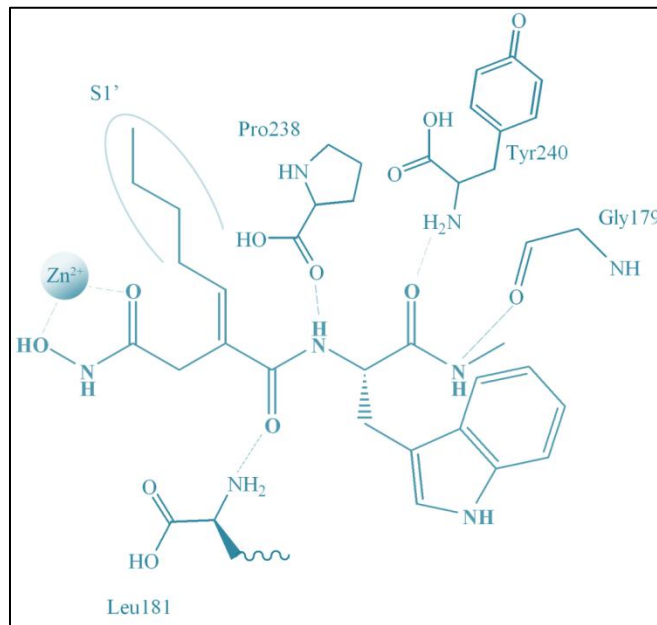
Figure 5.19: Structurally modified Illomastat and its binding interaction with MMP-2 (PDB: 1CK7)

The flexible nature of the protein is one of the major hindrances for designing an S1' pocket bound MMP-2 inhibitor. Durrant et al (2009) performed Molecular Dynamic simulation to explore a series of the molecule on the NMR structure of MMP-2 (1QIB). All the basic residues were protonised and acidic amino acids were deprotonated, also a harmonic constraint was added to maintain the orientation of Zn^{2+} ion. The computational method was performed using LUDI.

Oleic acid is a long chain fatty acid that holds both a zinc-binding group (-COOH) and a long hydrophobic chain -- a structure quite similar to the requirement of inhibition of MMP-2 enzyme. This inspired Moroy and his co-workers [36] to explore oleic acid and some galardin based derivative of it through biological assay and computational approach. The molecular docking in the experiment was performed using AutoDock Vinna. The docking scores were calculated based on van der Waal interaction, electrostatic, H- bonding, solvational and torsional interactions were considered. The crystal structure (PDB ID: 1CK7) was subjected to 'Blind

Modeling techniques used in designing MMP-2 inhibitor

docking' by taking a large grid box that engulfed nearly the full protein (96×126×120 with 0.64 Å spacing) to identify the probable binding site of the drug with the protein. Another grid box of 126×126×126 dimension with 0.60 Å spacing to calculate the binding rigidity. It was observed that the oleic acid may have 2 binding poses with the protein. In one case the S1' pocket gets occupied by the hydrophobic chain, on the other hand, the carboxylic acid chelates with the Zn^{2+} ion. In another pose, the ligand was found to be bound with the 3rd fibronectin type site of the enzyme. However, in the case of the galardin derivative, the –CONHOH group could not provide a satisfactory Zn^{2+} binding activity. The activity was however in micromolar range (~4.3 μM) in various biological assays. A detailed interaction is shown in **Figure 5.20.**
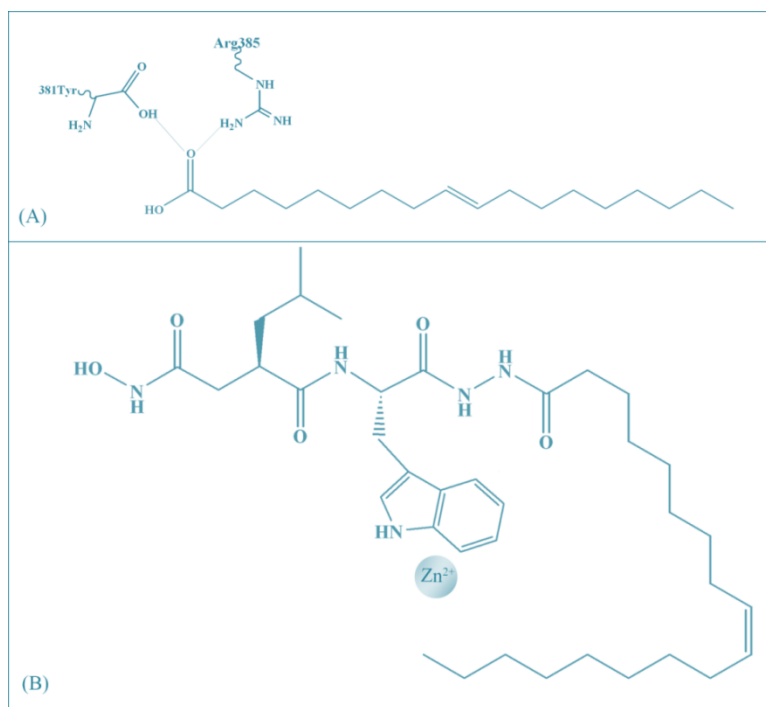
Figure 5.20: (A) Interaction of oleic acid with MMP-2 (PDB: 1CK7). The 2D schematic representation of the interaction shows 2 hydrogen bonds with Tyr381 and Arg385. The zinc-binding was not present. The long and flexible chain of the fatty acid has a chance to be inserted into the S1' pocket of the enzyme. However, the flexibility of this part is more than required. On the other hand, the interaction of oleic acid derivative of galardin with MMP-2 (B) shows poor interaction and no proper Zinc or amino acid interaction. Biologically also this derivative has a poor inhibition quality.

Modeling techniques used in designing MMP-2 inhibitor

In 2011, Pizio and co-workers worked on some of the non zinc-binding inhibitors for reducing the adverse effect arising due to the zinc-binding properties of ZBG containing inhibitors. Maestro and glide program was used to find the binding poses of tautomers of a single compound as depicted in **Figure 5.21**. Several substructural features were identified. The –NH group of methyl amide group was bound to the backbone of Gly162 CO, cyclohexyl amide CO showed interaction with the Tyr223 NH group. The Furamyl amide CO and Leu164 NH, pyrazyl NH formed a hydrogen bond with Ile222 for tautomer 1 and with Tyr229 in case of tautomer 2.
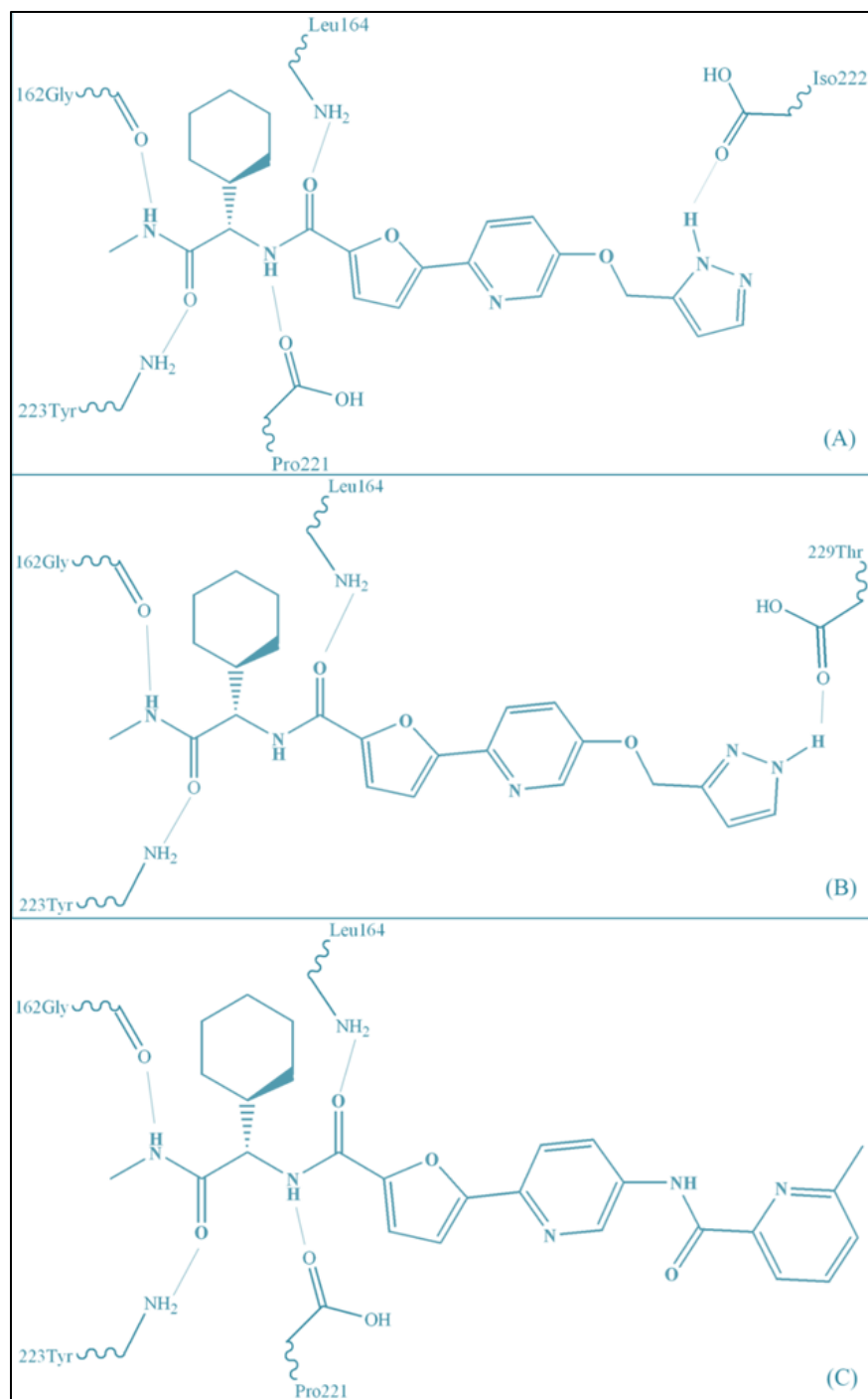
Figure 5.21: Docking interaction with MMP-2 (PDB: 1QIB) and 3 different structures. (A) 5 hydrogen bond interactions can be found in different parts of the compound. This has a sub nanomolar activity. (B) Similar to its tautomer, the second compound showed hydrogen bond interaction with 4 amino acid residues. The hydrogen of the altered pyrazyl group was found to have a hydrogen bond interaction with Thr229 instead of Iso222. (C) The least active compound has shown HB interaction only near the carbohexyl group and the other end showed no interaction. This might be the main cause of a very low activity of the molecule.

Modeling techniques used in designing MMP-2 inhibitor

Durrant and co workers (2011) performed Molecular dynamic simulation on a few pyrone based derivative. PDBID 1QIB was used for this methodology. Prior to the docking all the protein preparative methods were carried out. Partial charges were added, for zinc and all the flanking histidine residues were given partial charges using RESP tool, *ab initio* Hartree Fock method was used to calculate the electrostatic potential (HF/6-31G*). The preferred docking tool was AMBER-99SB. The study concluded that the ligand has two possible conformers, in one case, the S1' pocket remains deeply occupied and for the other state the ligand was not amenable into the deep pocket of MMP-2 enzyme. This study states the fickle nature of the S1' pocket which has led to many of its inhibitors to perform at it sub optimal potency.

Several chemically modified tetracyclines (CMT) (**Figure 5.22**) were put under several computational methods by Marcial and co-workers to understand the interaction between the CMTs and the MMP-2 enzymes (Marcial et al, 2012). To begin with, the partial atomic charges were calculated using Quantum mechanical (QM) method of DFT approach. Becke, 3 parameter, Lee-Yang-Parr (B3LYP) basis set was used with 6-31g (d) parameters. A DFT optimization was also used prior to docking, an $80\times80\times80$ grid box was calculated and the spacing was set to 0.275 Å. An MD simulation was performed to analyze the fortitude of the binding poses of the molecules. The best active molecule tagged as CMT-3 showed initial Zn^{2+} and flanking histidine molecule showed 2.31 Å, 2.33 Å and 2.36 Å distance, while post-process distance varied between 1.86 Å to 2.26 Å. Also, the lack of any binding group in ring D resulted in a hydrophobic interaction with Asp165 (3.4 Å). Further studies concluded that the coordination of Zn^{2+} ion may not hold equal importance as the occupancy of S1' pocket is. All these calculation were performed by the 1QIB structure of MMP-2.

Modeling techniques used in designing MMP-2 inhibitor
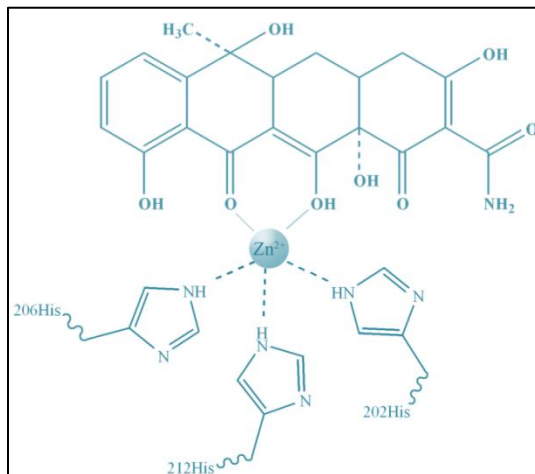
Figure 5.22: Chemically modified tetracyclines (CMT) has the ability to form bonds with the catalytic zinc ion of the MMP-2 protein. The oxygen atom of the hydroxyl group (O11) and carbonyl group (O12) were found to be the bare minimal necessity to create a zinc binding group.

Fabre *et al* (2013) used a computational docking method to support the SAR of the selected triazole-based compound. 19 different compounds with possible inhibitory activity were docked inside the 10th isoform of the 1HOV formation of the protein. The 10th position was the most favourable docking position as the co-crystal structure remains deep inside the target pocket (S1'). Glide was the choice of software for these scholars and all the default settings were maintained to perform a flexible docking calculation. The docking interactions (**Figure 5.23**) clearly expressed the zinc binding capacity of the hydroxamate oxygen atoms. Several other polar coordination and steric interactions were observed throughout the study.
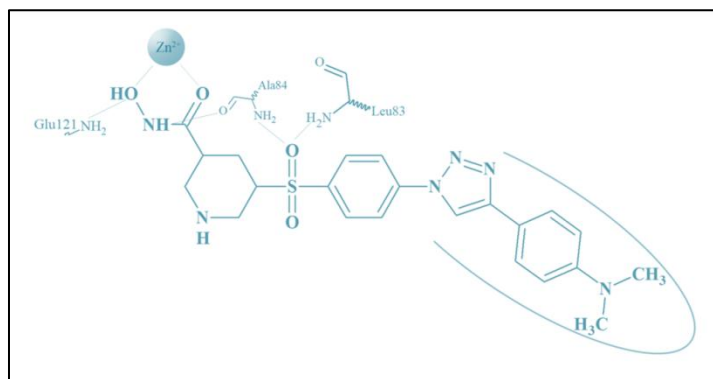
Figure 5.23: Docking interaction of a triazole-based hydroxamate and 10th isoform of the NMR structure of MMP-2 (PDB: 1HOV). Two very important interactions with the Glu121 and Ala84 is shown in the diagram. The green curved line shows the Van der Waal interaction of the group with the S1' pocket of the MMP-2.

Modeling techniques used in designing MMP-2 inhibitor

A molecular docking study was also performed to assess the stability of the binding mode of the preferred hydroxamate derivative throughout the duration of the MD study. It was found that among all the interactions, the hydrogen bond between Glu121, Ala84 and the hydroxamate group as well as the oxygen atom of the sulphonyl group creates at least 90% rate of abundance.

Apart from hydroxamates, several groups can be used as zinc-binding groups for chelating the active catalytic zinc. Tauro *et al* (2013) synthesized a group of possible inhibitors having biphosphonic core. These compound showed acceptable inhibitory action and was docked onto the crystal structure of MMP-2 for predicting the binding activity. The best active compound zinc binding action through only one phosphonic group; the other group remained in an h-bonded state with Leu164 and Ala165. A π-π stacking interaction with His201 was also observed (**Figure 5.24**). Interestingly, it was observed that both large hydrophobic and small hydrophilic molecules were able to inhibit the enzyme and in docking action, most compounds did not bind to the S1' pocket but into S1 and S2 pocket instead. The experiment by this group was quite in contrast to the popular belief of having an S1' pocket binding group in the probable inhibitors and thus makes this a unique piece of work.
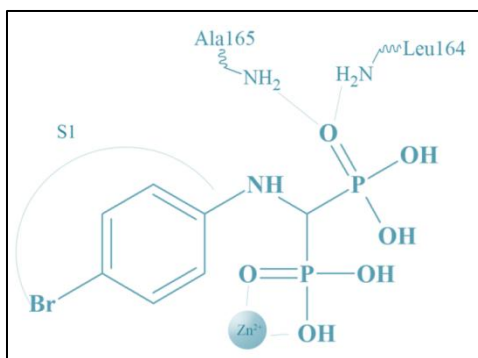
Figure 5.24: A docking interaction of biphophonic acid derivative a pan MMP-2 inhibitor. One of the phosphonic acid group is seen to be interacting with 2 amino acids as mentioned in the figure and the other one has a bidentate chelation with Zn^{2+} ion.

81

Modeling techniques used in designing MMP-2 inhibitor

In virtual screening method, probing of a particular pocket can be particularly useful to identify some yet-to-be-discovered interaction. One such molecule which is active on MMP-8 (PDB ID: 3DPF) was used to generate a pharmacophore model as the superimposition of MMP-8 crystal structure and MMP-2 crystal structure (PDB ID: 1QIB) showed several possible interactions between the MMP-2 and co-crystal molecule of 3DPF. It was observed by researchers Pizio and coworkers (2013) that probable squaramide ring might interact with Pro217, Leu164 and Ala165 (**Figure 5.25**). Imidazole ring can indulge in π-π staking interaction with His201. Additional hydrogen bonds were observed with residues of S1' pockets such as Ala217, Leu218, Thr227 and Thr229. A pharmacophore model was generated with 2 acceptors, 2 donors, 2 aromatic and 1 hydrophobic properties. The result of virtual screening was synthesized and the best active compound was farther docked on the crystal structure to rationalize the SAR. It was found that the furyl ring shows π-π stacking interaction with the His201 residue, the urea function reacts with Leu218, Ala217and Thr227 by hydrogen bonding. On the other hand, the naphthyridine group NH would remain in a hydrogen-bonded notion in carbonyl oxygen of Pro215 residue. All the docking interactions were calculated by Glide tools of Maestro package.

Modeling techniques used in designing MMP-2 inhibitor
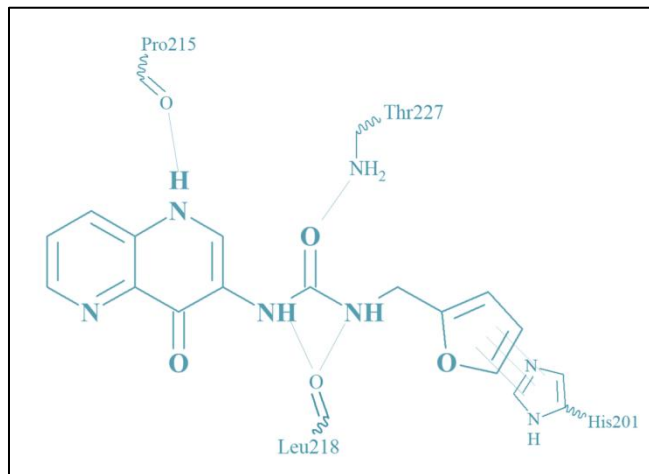
Figure 5.25: Docking interaction of a compound retrieved through virtual screening method. This is a non zinc binding inhibitor with micromolar activity in MMP-2 inhibition. The furyl group shows a π-π staking interaction with the His201. Interesting this is the same enzyme that binds with the catalytic zinc. Other residues mostly show hydrodgen bond interactions.

Ahmed *et al* (2015) worked on several derivatives of the natural product curcumin obtained from *Curcuma longa* and its derivatives difluorinatedbenzylidene curcumin or 3,4 difluorobenzylidene curcumin (CDF)/ It was docked onto NMR structure of MMP-2 1HOV using AutoDock 4.2. Optimised using self docking of the co –crystallized ligand, polar H was added by Kollman charge to MMP-2. Bong of ligand was rotatable, otherwise rigid. Parameter of zinc (r = 1.1 Å, e=0.25 kcal mol-1, +2e charge). Auto grid program generated 40x40x40 grid 0.375 Å spacing. 25 run with 150 population size with Lamarckian Genetic Algorithm (LGA) method and 25 conformers were found. The final 2D interactions for curcumin is shown in Figure 5.26 (A) and the interaction with CDF is shown in Figure 5.26 (B).

Modeling techniques used in designing MMP-2 inhibitor
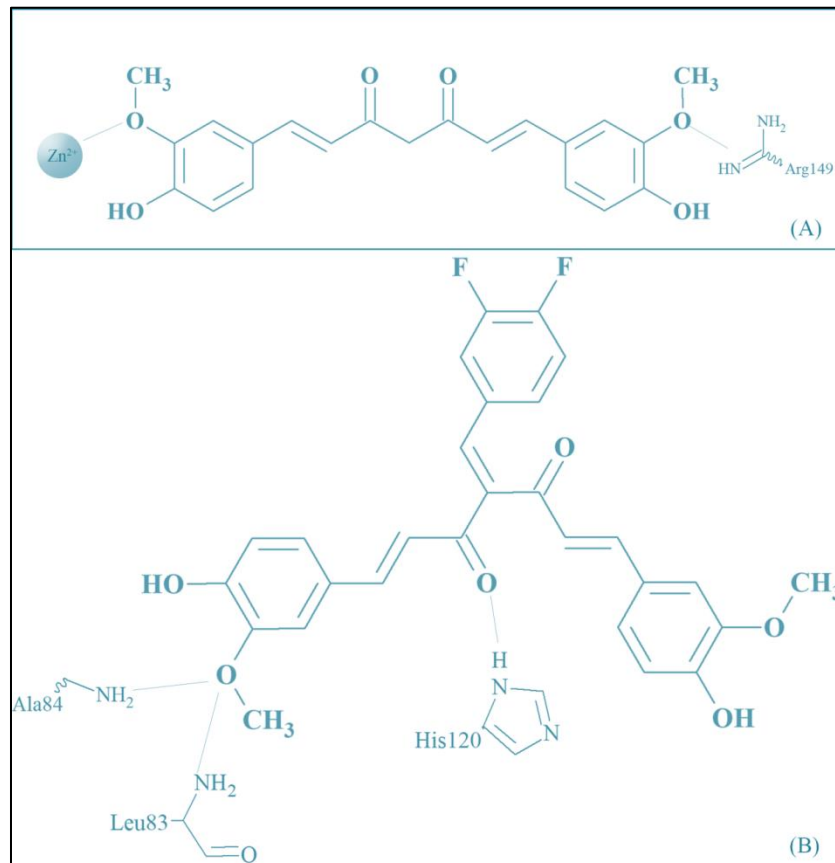
Figure 5.26: (A) The first figure shows a 2D interaction of curcumin and MMP-2 (PDB: 1HOV) enzyme. It can be clearly seen that the zinc ion binds with one of the methoxy groups in Curcumin structure, the other curcumin binds with the Arg149 amino group. (B) the second figure depicts the interaction between CDF and various MMP-2 residues. This is clear that CDF, unlike curcumin, does not bind with zinc.

Two important inferences could be drawn from the docking interactions for curcumin. Here, the catalytic Zn^{2+} ion binds with the methoxy group of the curcumin molecule at one end with 1.9 Å distance and Arg149 at a 2Å distance. It was observed that the ligand goes inside the S1 pocket. For difluorobenzyl curcumin, interaction with S1' residue Ala84 (1.8 Å), Leu83 (2.0 Å) and active site His 120 (2.1 Å) were evident. They also observed that the binding of difluorobenzyl curcumin and the co-crystal hydroxamic acid of the 1HOV crystal structure (SC-74020) had relatable overlapping.

Modeling techniques used in designing MMP-2 inhibitor

Among the (I-3, II-3)-biacacetin of biflavon class were synthesized by Nanjan *et al* (2015) as non-zinc binding groups and both biological assay and computational docking were performed. 1QIB structure was used and the most potent compound was docked using AutoDock package. It was found that the 7-OH group shows hydrogen bond interaction with Thr229, 5-OH and carbonyls showed H-bond interaction with Arg119 (**Figure 5.27**). Apart from these, Phe196, Phe232, Leu197 and Glu122 showed hydrophobic interaction with the molecule. No zinc binding actions were evident in the compound.
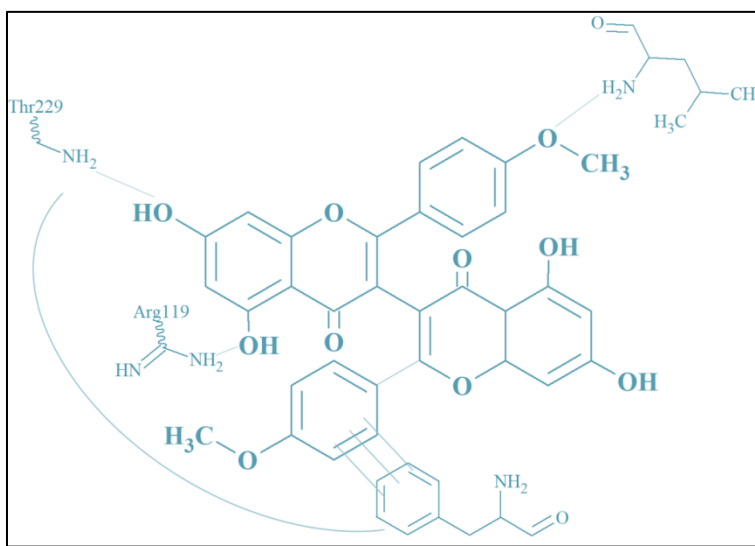
Figure 5.27: Interaction of flavone derivatives with MMP-2 structure (PDB: 1QIB). 3 hydrogen bonds can be seen with different amino acid residues. A π-π staking interaction is also present with the phenylalanine (No. 196). The green curved-line shows the probable S1' pocket binding interaction.

Resveratrol was docked into the 1QIB structure of MMP-2 to identify binding mode interaction of the compound with the protein (**Figure 5.28**). This work by Pandey and his co-workers (2015) from BHU, India showed binding energy of -37.85 kJ/mol binding energy with a K_i of 237.25 nM. The molecule interacted with several hydrogen bonds with amino acid residues of Leu164, Ala165, and Thr227. Ala217 and Ala220 showed polar interaction with the ligand.

Modeling techniques used in designing MMP-2 inhibitor
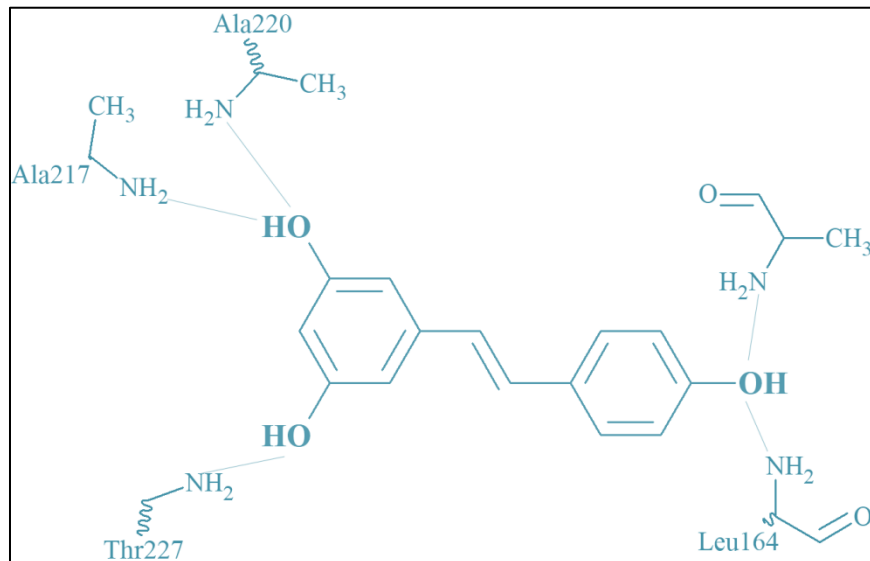
Figure 5.28: Docking interaction of resveratrol with 1HOV. The 2D interaction present here shows a handful of interaction between different residues of the protein and the molecule. However, as there was no catalytic zinc present in the vicinity of the ligand, it can be said that Resveratrol shows a non-zinc binding inhibition.

Li and coworkers (2015) worked on Eckol and its dimer Dieckol to study their inhibitory activity on several biological markers of vessel invasion including VEGF, Akt and also the topic of interest, MMP-2 to find the probable inhibitory activity of this natural product obtained from *Eisenia bicyclis*. The computation method was done on the crystal structure of 3AYU using Docking Server tool which is a freeware and the docking is performed by a remote server computer. The worker rationalized the structure-activity relationship that was observed through this process which is detailed in **Figure 5.29.**

Modeling techniques used in designing MMP-2 inhibitor
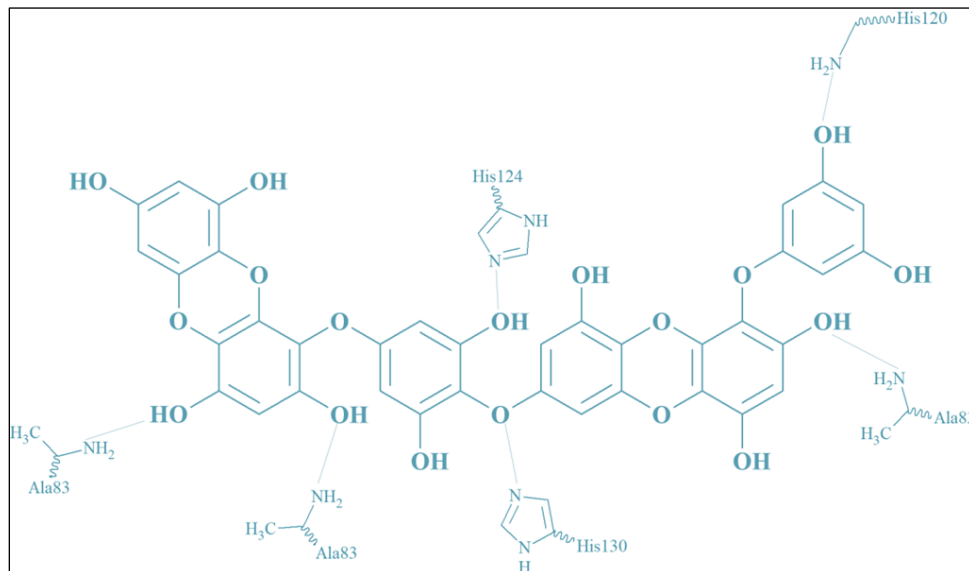
Figure 5.29: Interaction of Dieckol, a natural product derivative, with the 3AYU structure of MMP-2. The ligand shows multiple hydrogen bonds as seen in the picture. However, no zinc binding group is featured in the diagram and thus shows no interaction with the structural zinc.

Scholar Ammazzalorso and his co-workers (2016) worked on some compounds with urea linker group. The synthesized analogues were the non zinc binding type inhibitors (NZIs). Some important derivatives were docked using Glide tool with extra precision for most accurate prediction of the RMSD. Upon docking, it was found that among the selected compounds, 1H had good activity (15 μM) and NH group of the urea interacts with Leu218, whereas, the oxygen atom of the carboxyl group between the two amide groups interact with Thr227. The larger molecule interacts with the Pro221, Ala165 and Leu164 group residue (**Figure 5.30**).

Modeling techniques used in designing MMP-2 inhibitor
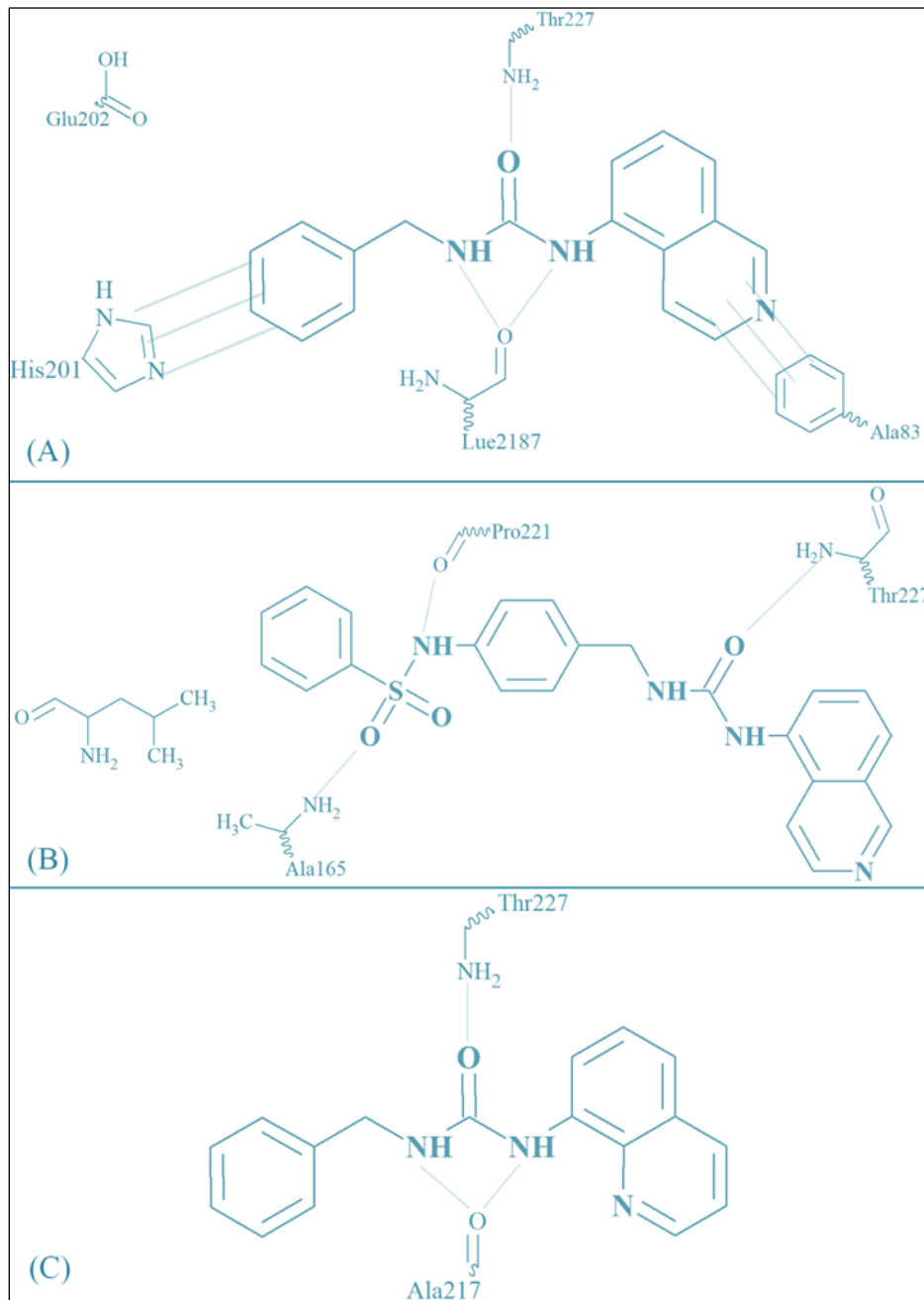
Figure 5.30: Docking of several non-zinc binding inhibitors (NZI) of MMP-2. This result is drawn from the docking of the best active compound as shown in (A) and upon docking with a larger compound as in (B) reveals more plausible interaction with this ligand and the subsequent derivative with the MMP-2 enzyme. The ligand shown in (C) shows a similar structure with the first compound but took lesser time (10 ns) to stabilize when compared to the first compounds (100 ns)

Green tea is known to be a good source of natural products with various bioactivities. Catechins are phenolic derivatives with antioxidant properties. Chowdhury *et al* (2017) explored some

Modeling techniques used in designing MMP-2 inhibitor

catechins (Epicatechin, epigallocatechin, epicatechin 3-gallate and epigallocatechin 3-gallate) for their MMP-2 inhibitory properties through a number of biological and computation approaches by experimenting on both proMMP-2 and active MMP-2. A thorough docking process was performed using the GOLD package of Discovery Studio 2.5. The quality of the docking was determined using CHEMPLP score of the docked ligand. The docking interaction showed all of the molecules had interaction (van der Waals, polar and stacking) with a significant number of amino acid residues. It was quite intriguing to observe that all of these natural products showed non-zinc binding inhibition of MMP-2. **Table 5.1** shows the interaction between phytochemicals found in green tea with MMP-2.

Table 5.1: Interaction of green tree derived catechin and its analogue with MMP-2 enzyme

Ligand	Residues with van der Waal interactions	Residues with polar interactions	π stacking interaction
Epicatechin	Met 421, Lue190, His193	Pro423, Val400, His413, His403, Glu404, Ala194, Ala192	-
Epigallocatechin	Leu190, His193, His413	Pro423, Val400, His403, Glu404, Ala194, Ala192, Leu191	-
Epicatechin 3-Galleate	Leu191, Leu190,Pro423, Trp213, Leu397, Leu399	Glu404, Ala192, Ala194, His403, Ser205	Phe207
Epigallocatechin 3-Galleate	Tyr425, Phe207, Gly189, Trp213, His193	Leu190, Glu404, His403, Leu399, Val400, Ala401, Ser205	Met421

Modeling techniques used in designing MMP-2 inhibitor

As the terminal hydroxyl group of a hydroxamate derivative works as a Zinc Binding Group, the polar opposite vertex is designed in such a way that it can be fitted into the deep S1' pocket which is exclusive to the MMP-2 protein. Para-substitution in the terminal phenyl ring was the main target of Gao and his coworkers (2016) for increasing the S1' binding affinity. A docking interaction was reported after designing, synthesis and assay of these compounds. It was found that due to the p-substitution at the aryl ring the compound was able to fit inside the S1' pocket of the enzyme. Moreover, as the residues near the neck region of the pocket are smaller in size (both being threonines) compared to MMP-9 (one Arginine) and MMP-14 (One Glycine and one methionine), better selectivity has been observed. 2D interaction is shown in **Figure 5.31**.
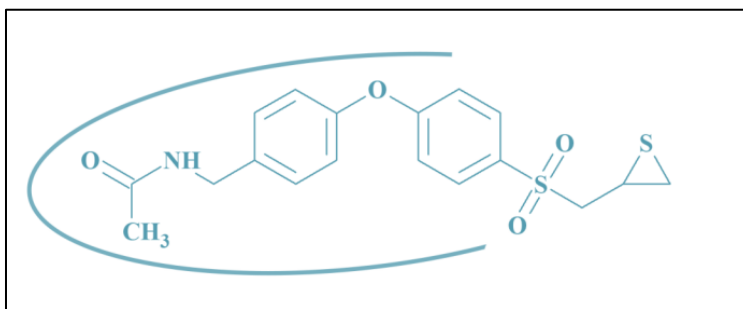
Figure 5.31: Interaction of the synthesized compound with MMP-2 enzyme active site. The ligand shows an S1' pocket binding (represented as the curve) which is also selective in nature compared to MMP-9 and MMP-14 binding.

Instead of directly inhibiting the catalytic region of the enzyme, the group of Jha and co-workers (2016) went onto inhibiting the fibronectin region. It was observed through docking studies of epigallocatechin galleate (EGCG) with an MMP-2 crystal structure that the inhibitor binds to Asn245 and Glu243 in repeat 1 portion, and Cys349, Ser338, Ser362, Glu361and Pro391 in repeat 3 (**Figure 5.32**). Such binding action can hinder the gelatin-binding activity of MMP-2. The mean distance between the active catalytic site of the enzyme and the glycin - leucine bonding increased to 7.0 Å from the usual 3.4 Å due to the presence of EGCG. Though this

Modeling techniques used in designing MMP-2 inhibitor

method of inhibition is beyond the scope of this discussion as we try to unearth theoretical properties of small molecular S1' pocket bound inhibitor, the potency of such action observed form a natural product derivative is quite appreciable and can be explored farther.
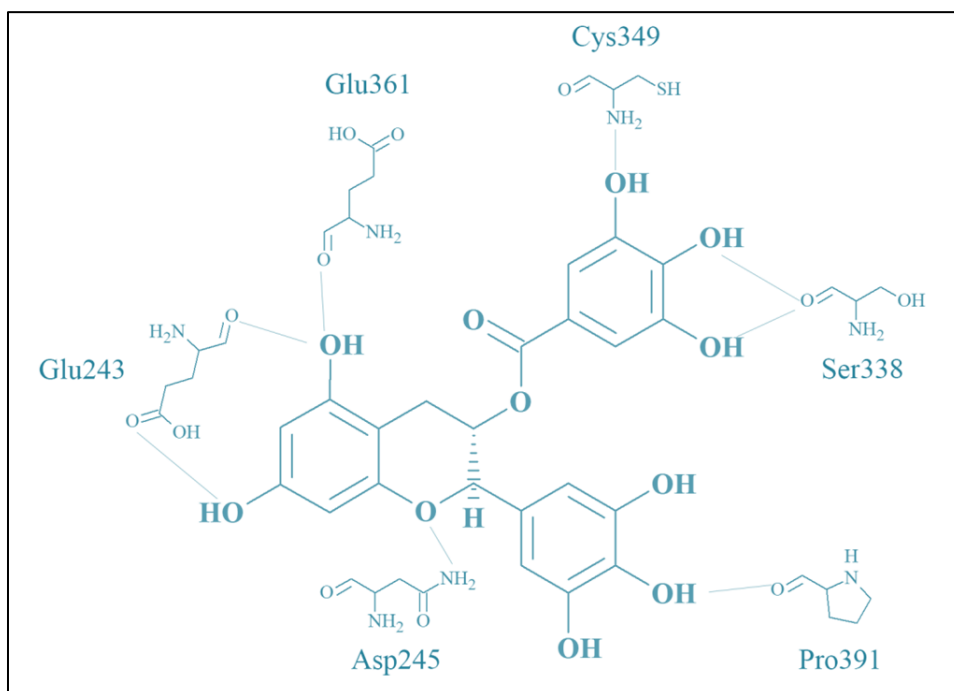
Figure 5.32: Interaction of Epigallocatechin gallate with the fibronectin region of MMP-2 (PDB: 1CK7).

Several Ilomastat analogues were synthesized by Song *et al* (2016) through benzamide group modification. The concept was taken from chidamide, an important histone deacetylase (HDAC) enzyme that would use a benzamide moiety instead of a hydroxamate moiety to bind with zinc ion of HDAC which also happens to be a Zn dependent protein. The best active compound expressed respectable selectivity towards MMP-2 compared to MMP-9. A docking interaction was calculated to fortify the SAR of the compound. The observation was interesting as the compound did not bind to the Zn^{2+} ion in the catalytic region. It was found that van der Waals interaction was evident for Tyr142, Leu116, Thr143, Leu137, Ala139, Ile141 and Ala136 pocket

of S1'pocket. Also, indole group was well placed into the S2 cavity and the phenyl ring His124 involved into π-π interaction force and also it was into van der Waals interaction with Ala84, His85 and Ala86 (**Figure 5.33**).
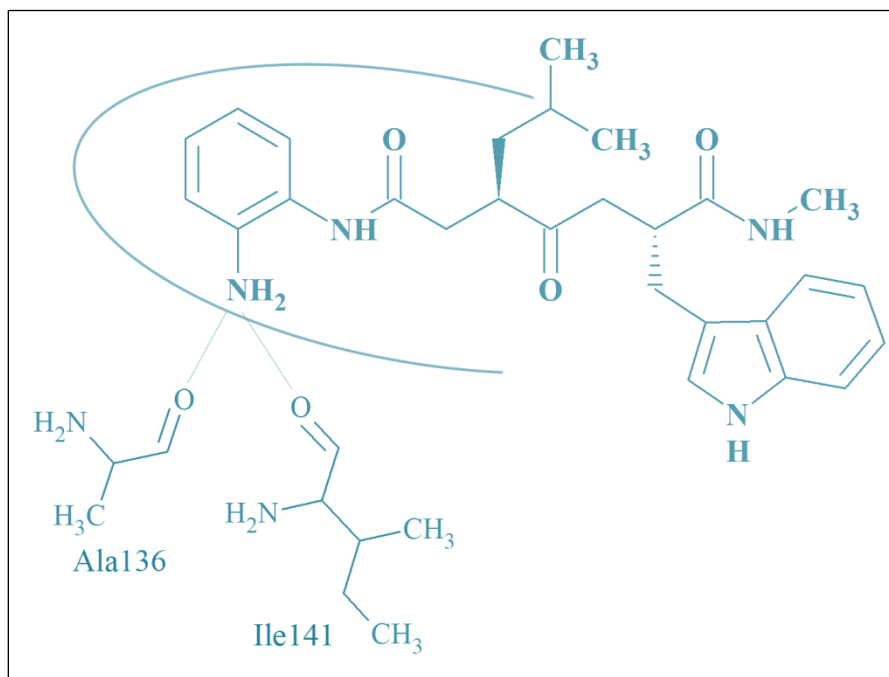
Figure 5.33: Interaction of Ilomastat derivative with MMP-2 (PDB: 1HOV). The ligand is non zinc binding with mostly van der Waal's attraction in action, however hydrogen bonds can be found in the deep S1' pocket (depicted as green curved line).

Lima and coworkers (2018) from Universidade Federal de Viçosa, Brazil synthesized several cinnamic acid derivatives with triazole ring. The SAR of the best active compound was established with docking studies. Here, the preferred tool was docking was AutoDock Vinna (1.1.2) and the 1QIB structure was used. The best active compound showed ionic-π stacking interaction Zn^{2+}, 2 of the nitrogen atoms of the triazole ring interacts with the Glu202 and Glu184 (2.5 Å and 9.1 Å respectively) whereas, the phenyl ring interacted with Glu150 amino acid of the S1' pocket (**Figure 5.34**).

Modeling techniques used in designing MMP-2 inhibitor
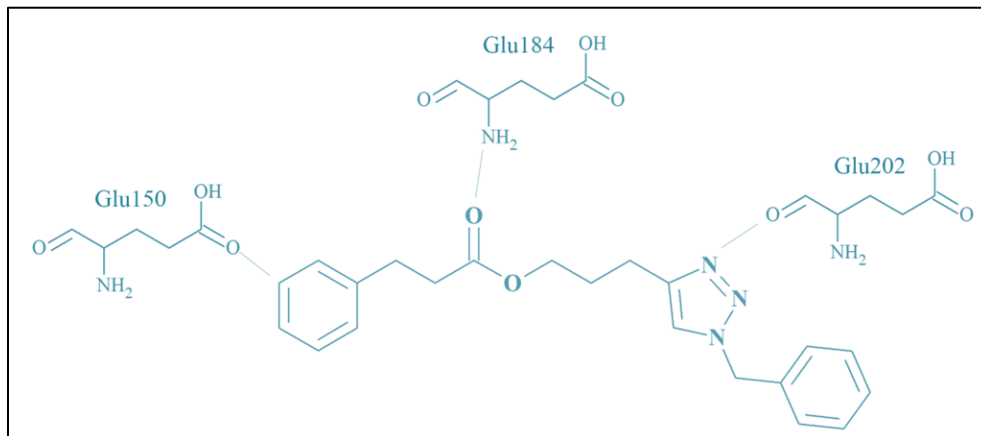
Figure 5.34: Triazole based cinnamic acid derivative and its interaction with MMP-2 protein (PDB: 1QIB). This inhibitor is a non zinc-binding inhibitor with the ability to bind at the selectivity pocket

Aswagandha (*Withania somnifera*) is known as the 'Indian Ginseng' as its similarities with the famous Chinese nerve fortifying herb, ginseng. Kumar et al conducted a detailed molecular docking study on 36 WS phytochemicals along with standard MMP-2 inhibitors like doxycycline, minocyclin and quencetin (Kumar et al, 2018). All the structures of the natural products were collected form reliable online database and using the OpenBabel software, all the structures were converted to 3D structure. A 0.375 Å grid box of 80×80×80 size was constructed as part of basic methodology of performing docking operation through AutoDock tools. Lamarckian Genetic algorithm (LGA) was performed for 30 runs and 150 max population sizes. The program was left to run for maximum 27,000 generation and 2,500,000 max energy evaluations were performed. The final visualization was performed using LigPlot+ tool. The study showed that 28 out of all the docked constituent of WS was showing better binding energy (and a better inhibition of MMP-2) than all three of doxycycline, minocyclin and quencetin. Among the steroids, Withanolide G had the least binding energy with (-11.10 kcal/mol) and the least K_i value (11.10 nM). For non steroids, the somniferine alkaloid showed acceptable binding energy (-10.58 kcal/mol). The phytocehmicals showed a nearly similar binding activity as

Modeling techniques used in designing MMP-2 inhibitor

hydroxamates and they remained bound to the S1' pocket juxtaposed to MMP-2 catalytic region.

Both the binding interactions are shown in **figure 5.35**.
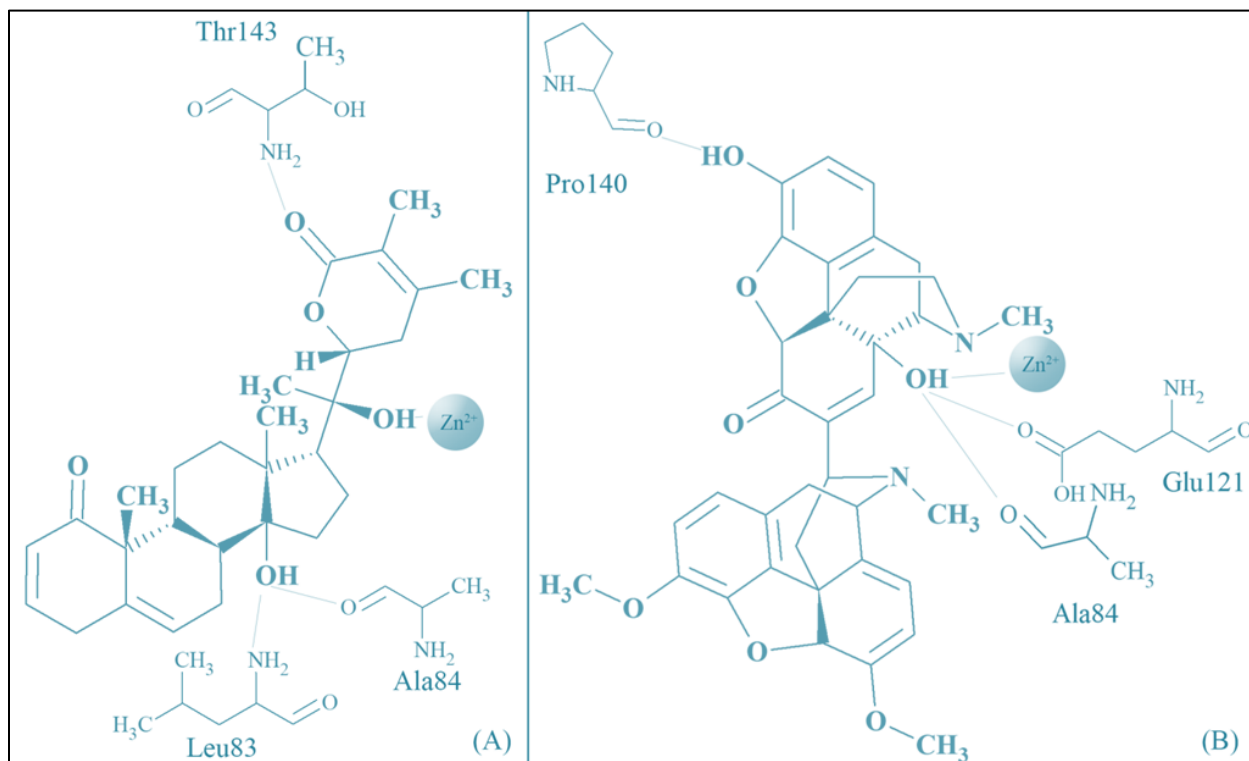
Figure 5.35: Interaction of some of the active constituents of Awsaganda (*Withania somnifera*). (A) Shows interaction between the Withanolide G and MMP-2 protein structure. The compound in Zinc binding and forms hyfrogen bond with three more amino acid residues as seen in the figure. (B) Same interaction model mapped for the compound Somniferine.

Virtual screening is a high through-put method of identifying a series of novel lead compounds. As the long-lasting research on unearthing a dependable and selective MMP-2I failed to ameliorate the current situation constant lead finding using both pharmacophore based and structure based approach is still in practice and widely accepted. Benscik *et al* went on finding a new probable lead from the Albany Medical Research inc. (AMRI) library of bio active compounds. A pharmacophore was built based on some of the known and important inhibitors of MMP-2. Computational sub structural search were carried out by using InstatJChem computational tool for identifying similarities and form a screening library of similar

Modeling techniques used in designing MMP-2 inhibitor

compounds. Finally the screened compounds were docked using genetically optimized ligand docking (GOLD) molecular docking software to quantify the binding interaction between the ligands and the crystal structure (1QIB) and NMR structure (1HOV) of the MMP-2 enzyme.
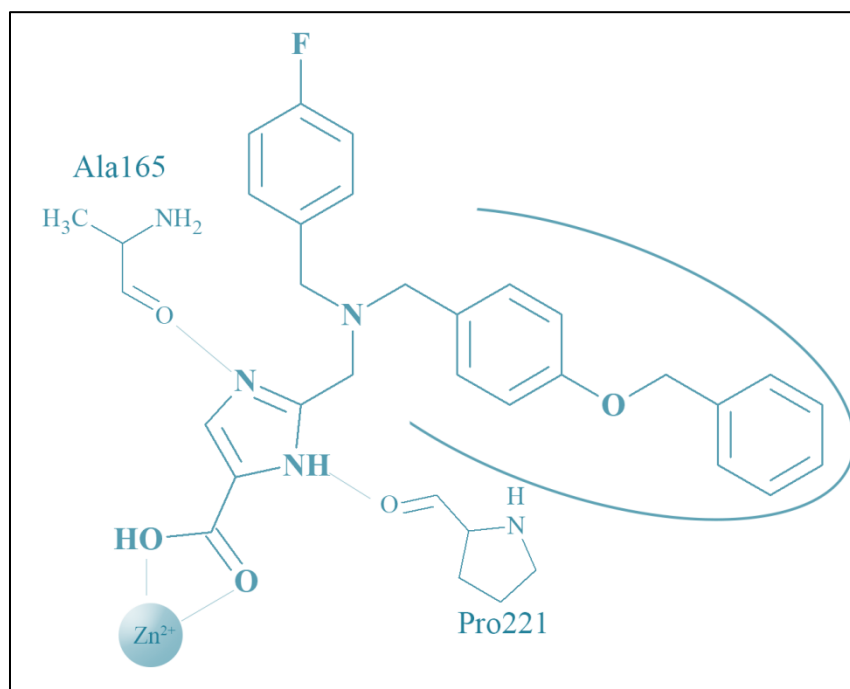
Figure 5.36: Hydrogen bond interaction of a compound as a result of virtual screening of a number of compounds. For ease of view only the hydrogen bond interaction and zinc binding ability has been shown in the diagram. The compound also shows good S1' pocket binding ability (depicted as green curved-line).

The lead compound thus found showed significant interaction energies with the crystal structure of the enzyme. It can be observed that, the molecule showed hydrogen bond (both weak and strong) with Ala165, Ile222, Tyr223, Leu21, leu191, His201, Val198. The carboxylic acid had bi-dentate chelation with the Zn^{2+} ion which is fortified in the presence of electron donating atom near this group. Special hydrophobic adherence was observed between pyrazole and Ala165 and Pro221 residues of the enzymes. Another interesting observation was the presence of pyridine moiety instead of the terminal phenyl group will increase the selectivity and at the same time

Modeling techniques used in designing MMP-2 inhibitor

improve the clogP and in turn allowing better cellular permeation. A 2D interaction is shown in

figure 5.36.

6. The rationale behind the work:

The rationale behind the work:

The previous study has shown us to what extent the work on a selctive MMP-2 inhibitor is being in progress. Inspite of major works in a few previous decades pharmaceutical companies and academia has not been able to put forward a target specific MMP-2 inhibitor in the market. Both zinc binding and non zinc binding inhibitors are under the keen observation of the researchers. Here the zinc binding type of inhibitors will be considered with a concentration towards the S1' pocket binding which is mainly because of the moderately sized pocket exclusive to MMP-2 only. This pocket is solvent exposed and possible π-π stacking interaction may provide a good interaction with the hydrophobic region. At the same time, hydroxamates are probably one of the most dependable zinc binding groups, even better than carboxylic acids. Therefore, a thorough study on a series of aryl sulphonyl group of compound can be beneficial for designing target specific molecule with MMP-2 inhibitory property in the near future.

For this study, we have focused on the arylsulphonamido hydroxamate compounds. For a primilinary study we have taken a small amount of compounds with some fixed structural feature and from a single group of researchers. After chemometric study of these compounds we selected larger dataset obtained from the binding data base has been used to explore common structural properties of a larger amount of compound. Finally we culminate the whole concept to discuss the probable future perspective of the work.

7. Materials and methods:

Materials and methods:

7.1 Molecule/ Dataset preparation:

For a preliminary study, several compounds were selected for QSAR performance from different literatures (Rossello et al, 2004; Nuti et al 2007; Nuti et al 2009; Nuti et al, 2011; Nuti et al 2013; Nuti et al, 2015; Nuti et al, 2015; Sjoli et al 2016). The molecules were selected in a way that pharmacologically it had MMP-2 inhibitory activity and at the same time structurally it was hydroxamates of aryl sulphonamides. The molecules were first drawn in ChemDraw 2 5.0 and were thoroughly checked multiple times for any structural error (**Figure 7.1**). On the next stage, it was converted to a corresponding three-dimensional structure and was viewed and modified to nullify any stereochemical error. The structures were then prepared using *Prepare Ligand* protocol in the discovery studio (www.3ds.com) and a few duplicate files were removed and all the structures were catenated into one **.sdf* file instead of the individual **.mol* file.
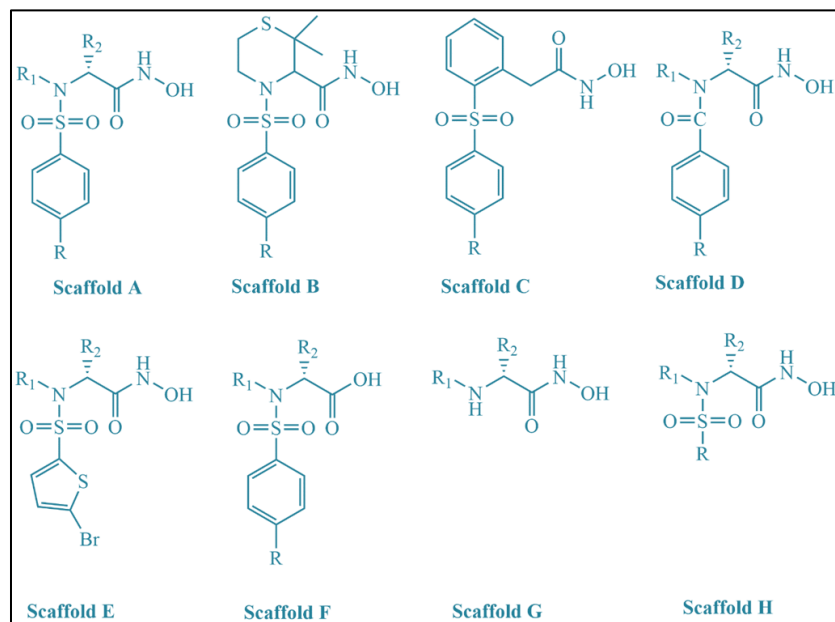
Figure 7.1: Common scaffold of the molecules subjected for the preliminary study

Materials and methods

No[a]	Scaffold	R	R$_1$	R$_2$	Activity[c] (IC50) nM	Activity[c] (pIC_{50})
1[b]	A	OMe	*i*-But	H	6.9	8.161
2	A	OMe	O*i*-Pr	H	3600	5.444
3	A	OMe	EtO	H	9600	5.018
4	A	OMe	*p*-Bn-O-Bn-O-	H	2100	5.678
5[b]	A	Ph	*i*-But	H	13	7.886
6	A	Ph	*i*-Pr-O	H	12	7.921
7[b]	A	Ph	OCH$_2$CHCH$_2$	H	33	7.481
8	A	Ph	*p*-Bn-O-Bn-O-	H	41	7.387
9	A	OPh	Pyridyl-3-methyl	*i*-Pr	20	7.699
10	B	OPh	--	--	0.083	10.081
11[b]	A	OMe	Benyloxy	H	1060	5.975
12	A	OMe	4 chlorobenzyloxy	H	1500	5.824
13	A	OMe	*n*-Tetradecyl oxy	H	30000	4.523
14	A	OMe	Adamentyl oxy	H	1980	5.703
15[b]	A	OMe	O-*i*-Pr	*i*-Pr	660	6.180
16	A	OMe	Sec-Butyloxy	*i*-Pr	730	6.137
17	A	Ph	O-*i*-Pr	Benzamidoethyl	1.5	8.824
18[b]	A	Ph	O-*i*-Pr	Methylsulfonamido ethyl	1	9.000
19	A	Ph	O-*i*-Pr	Acetamidoethyl	0.33	9.481
20	A	Ph	O-*i*-Pr	Phenylacetamidoethyl	1.4	8.854
21	A	MeOPh	O-*i*-Pr	Acetamidoethyl	0.13	9.886
22[b]	A	MeOPh	O-*i*-Pr	Phenylacetamidoethyl	0.37	9.432
23	A	H	O-*i*-Pr	Acetamidoethyl	590	6.229
24	A	Ph	O-*i*-Pr	Pthalimidoethyl	0.67	9.174
25	A	Ph	O-*i*-Pr	*i*-Pr	0.81	9.092
26	A	OMe	sec-But	*i*-Pr	29	7.538
27	A	Ph	sec-But	*i*-Pr	3	8.523
28[b]	A	Ph	H	H	2400	5.620
29	C	OMe	--	--	230	6.638
30[b]	D	PhO	H	Pthalylethyl	9000	5.046
31	E	--	O-i-Pr	H	1100	5.959
32[b]	F	--	BnO	n- But	34000	4.469
33[b]	A	Benyloxy	H	H	3.8	8.420
34	A	*p*-Bromobenyloxy	H	H	1.2	8.921
35	A	*p*-Fluorobenzyloxy	H	H	3	8.523
36	A	*p*-trifluoromethoxybenzyloxy	H	H	7.5	8.125
37	A	*p*-biphenylmethoxy	H	H	48	7.319
38	A	*m*-Bromo-benzyloxy	H	H	15	7.824
39	A	*m*-Fluorobenzyloxy	H	H	7.8	8.108
40	A	*m*-	H	H	20	7.699

		trifluoromethoxyben zyloxy				
41[b]	A	Biphenyl 3-methoxy	H	H	41	7.387
42	A	Isoquinolinyl 8-methoxy	H	H	670	6.174
43	A	*m,m*-difluorobenzyloxy	H	H	18	7.745
44[b]	A	*m,m*-dibromobenzyloxy	H	H	1750	5.757
45[b]	A	2-propynyl 3-oxy	H	H	110	6.959
46	A	*m,m*-dibromobenzyloxy	Me	H	1870	5.728
47	G	*m,m*-dibromobenzyloxy	H	H	125000	3.903
48	A	*m,m*-(methoxypenyl) bromobenzyloxy	H	H	3800	5.420
49	A	*m,m*-(pyridyl) bromobenzyloxy	H	H	1045	5.981
50[b]	H	4□-methylthio biphenyl 4- yl	O-*i*-Pr	H	1.5	8.824
51	H	4□-methoxy biphenyl 4- yl	O-*i*-Pr	H	4.2	8.377
52	H	4□-butoxy biphenyl 4- yl	O-*i*-Pr	H	5.5	8.260
53	H	4□-(p-chlorobenzyloxy) biphenyl	O-*i*-Pr	H	40	7.398
54	H	4□ (4-biphenyl) vinyl phenyl	O-*i*-Pr	H	74	7.131
55	H	4□ (phenyloxy) phenyl 4- yl	O-*i*-Pr	H	26	7.585
56	H	2-((4-butylphenyl)ethynyl) thiophen – 5 - yl	O-*i*-Pr	H	360	6.444
57[b]	H	2-(4-fluorophenyl)thiophe n -5- yl	O-*i*-Pr	H	310	6.509
58	H	2-(4-ethylphenyl) thiophen -5- yl	O-*i*-Pr	H	56	7.252
59	H	2- (4-methoxyphenyl) thiophen – 5-yl	O-*i*-Pr	H	17	7.770
60	A	OMe	H	H	32	7.495
61	A	Br	H	H	40	7.398
62	A	Ph	H	Acetamidoethyl	0.7	9.155
63	A	2- Trifluoromethyl phenyl	H	Pthalylethyl	390	6.409
64	A	Br	H	Acetamidoethyl	190	6.721
65	A	Br	H	Pthalylethyl	1.5	8.824
66	A	But2yn 2-yl oxy	H	Bezylcarba	17	7.770
67	A	Butyn 2-yl oxy	H	Pthalylethyl	4.2	8.377

68[b]	A	Butyn 2-yl oxy	*i*-Butyl	Pthalylethyl	21	7.678
69[b]	A	(But 2-yn) –yl oxy	H	$(C_6H_5CH_2O)CONHC_2H_4$	17	7.770
70	A	$(3,5\ C_6H_5Br_2)CH_2O$	H	$(CH_2)_3NHCONH_2$	240	6.620
71	A	$(3,5\ C_6H_5Br_2)CH_2O$	H	$(CH_2)_3NHCbz$	4100	5.387
72	A	$(3,5\ C_6H_5Br_2)CH_2O$	H	$(CH_2)_2NHCbz$	1440	5.842

On the next step, a database search was performed. The targeting ligand was MMP-2 from the binding database. Initially, 1,600 were found and first, they were filtered using SMILES filter of OpenBabel [citation] and only hydroxamates were selected. The duplicates were all removed using the previously discussed method. A classification QSAR model was then developed subsequently.

2 important filters were used in the conjuncture of the previous filters. The Veber proposed and Lipinski proposed methods of drugability were considered and several molecules that failed to comply with these rules were removed (Lipinski, 2004)

On a third step, a farther step was included and phenyl sulphonamide hydroxamate was selected through OpenBabel tool and Classification QSAR studies were performed.

For this 2nd phase of the study, the molecules were selected in an 'as is' condition, i.e. no in-depth cross-checking was performed to modify any structural feature or the biological activity of the

7.2 Descriptor calculation:

72 molecules thus formed were then processed through PaDEL tool at the next phase. PaDEL tool is a free chemometric tool created by CW Yap (2011) and his coworkers and remains one of the major descriptor calculation tools for QSAR studies. PaDEL has The ability to

Materials and methods

calculate 1875 descriptors (among which, 1444 are one- or two-dimensional descriptors and 431 are 3D descriptors) and 12 types of fingerprints (total 16092 bits). For this study, we select the 1444 descriptors to the 2D category. The biological activity (IC_{50}) was in the nanomole or micromole unit. All of them were converted to the molar range and the negative logarithm of IC_{50} (*pIC_{50}*) for a more linear prediction of models.

7.3 Training and test set division:

The dataset used for the preliminary studies was divided into training and test sets to get a cross-validated QSAR model. A y-based ranking method was used by sorting the biological activity data in ascending order and by selecting every fourth data from the pool. The best and least active molecule of the set was kept out of the training set.

A random selection of test and training set was performed for the classification QSAR model using the same dataset. 75% of the molecules were considered as the training set of the predictor model and rest were taken as the test sets. This training and test sets were used for both the Bayesian classification and recursive partitioning process.

7.4 Descriptor reduction:

For the first dataset, the dataset was reduced using the V-WSP tool (Ambure et al, 2015) of DTC lab with correlation cut-off set to 0.80. The 1,444 number of descriptors were reduced to 393. The variables were farther reduced to 15 using a genetic function approximation method of variable reduction. Finally the best subset of variables was excavated from the 15 descriptors and finally, 8 descriptors were selected. A forward stepwise multiple linear regressions was performed to find out the final regressive model for the prediction of biological activity.

104

Materials and methods

7.5 Model Validation:

Quality of any prediction model is subjected to several metrics calculation. Table 5 shows the metrics calculated for the acceptability of the model along with the working formula.

Table 7.2: Statistical metrics to validate 2D QSAR technique

Name of the metric	Working formula
R^2	$R^2 = 1 - \dfrac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - Y_{train})^2}$
Q^2	$Q^2 = 1 - \dfrac{\sum(Y_{obs(training)} - Y_{pred(training)})^2}{\sum(Y_{obs(training)} - Y_{train})^2}$
$Q^2_{(F1)}$	$Q^2_{(F1)} = 1 - \dfrac{\sum(Y_{obs(test)} - Y_{pred(training)})^2}{\sum(Y_{obs(training)} - Y_{training})^2}$
$r^2_{m(test)}$	$r^2_{m(test)} = r^2(1 - \sqrt{r^2 - r_0^2})$
Variable Inflation Factor (VIF)	$VIF = \dfrac{1}{1 - r^2}$
Standard error of estimation (SEE)	$SEE = \sqrt{\dfrac{\sum(Y_{Obs} - Y_{Calc})^2}{n - p - 1}}$
Euclidean distance	$d_{ij} = \sqrt{\sum\limits_{k=1}^{m}(x_{ik} - x_{jk})^2}$
Mean distance of the equation	$\bar{d}_i = \sum\limits_{j=1}^{m}\dfrac{d_{ij}}{n} - 1$
A squared mean correlation of the randomised model	$^C R_p^2 = R \times \sqrt{R^2 - R_r^2}$

Materials and methods

7.6 Molecular orbital visualization:

Molecular orbitals play a vast role in drug designing technique. These determine molecular reactability and at the same time the electron donating and withdrawing properties. HOMO or the Highest Occupied Molecular orbital shows the tendency of the nucleophilicity of the compound and on the other hand, the LUMO or Lowest Unoccupied Molecular Orbital demotes the ability to have an electrophilic interaction of the molecule. The molecular orbital for the best active molecule and the least active molecule from the data set 1 was used for the preparation of HOMO and LUMO generation using Gaussian 03W software. The Density Functional Theory (DFT) was used with a B3LYP (+d,p) basis set. A "pop=full" notation was added to create a visible molecular orbital. MASK or molecular modelling and simulation kit was used for the molecular visualization.

7.7 Bayesian classification:

Bayesian classification is a widely used statistical model for prediction method, and the concept of Naïve Bayesian classification is used in molecular modelling as well. Here the working formula used is

$$P(y|x_i) = \frac{\prod_{i=1}^{n}(x_i|y)}{\prod_{i=1}^{n}(x_i)}$$

In the above formula, the *y* is the dependent variable / biological activity reproduced as binary (active or inactive; represented as 1 and 0 respectively) the *x* on the other hand, is the independent or the intrinsic property of the molecule. This classification process has been done using Discovery Studio 3.0. For the calculation, several 2-D molecular properties were

Materials and methods

calculated along with the FCFP_6 molecular fingerprint descriptors and the biological activity was converted into binary considering any activity more than the mean activity to be 1 and less being 0 (Roy et al 2015).

7.8 Recursive partitioning method:

Recursive partitioning method is another method of classification QSAR. However, unlike the Bayesian classification model, a logical decision scheme is used instead of direct probabilistic function. Several molecular properties and fingerprint (FCFP_6) were used to split the root node into the final leaf nodes and compare the decision of calculated molecular property (binary) with the true value to get the final values of the biological activity of the molecules.

Tough decision tree method is easy to explain the overall process; the method has the chance of overfitting the data predicted data lacking the ability to extrapolate the model into a larger data set. To remove this issue, a random forest is recommended to get a more refined model. Random forest, as the name itself suggests, is a collection of decision trees. the slipping characteristics of the trees are selected randomly and then from the collection of the decision trees, the forest puts forward a cross-verified model. A receiver operating characteristic value proves the fitness of the classification of the model (Roy et al 2015).

8. Discussion

Discussion

8.1 Inter-correlated descriptor of can be real troublesome for generation of a reliable model. After the initial pretreatment of the dataset, the preliminary 15 descriptors were considered for the "*best subset selection*" method (Ambure et al 2015) as high number of variables are not only tough to explain but also the chance of overfitting the data remains high. A total of 6,435 models were generated through multiple linear regression (MLR) method when a set of 8 descriptors ($^{15}C_8$) were selected. Among these models, only 825 models were fulfilled the statistical criteria proposed by Golbraik and Tropsha (2002). Finally, the Q^2_{LOO} value was used for finalizing the best 2D-QSAR model. Ultimately, the MLR model is as follows:

pIC_{50} = - 53.133 (±12.117) + 17.840 (±3.138) *SpMax1_Bhv* – 10.4197 (±1.518) *MATS3m* - 1.423 (±0.161) *SsssCH* – 0.037 (±0.005) *AATS4m* – 0.132(±0.029) *VE3_Dze* + 0.127 (± 0.039) *VE3_DzZ* – 3.194 (±0.925) *GATS4p* – 1.776 (±0.641) *SpMax4_Bhv* **Eq. 1**

n_{train}= 54; R = 0.893; R^2= 0.797; R^2_A= 0.761; $F(8, 45)$ = 22.138; SEE= 0.689; $Q^2_{(LOO)}$= 0.725; *Avg.* $r_m^2{}_{(LOO)}$= 0.622; $\Delta r_m^2{}_{(LOO)}$= 0.162; n_{test}= 18; R^2_{pred} =0.643; *Avg.* $r_m^2{}_{(test)}$ = 0.550; $\Delta r_m^2{}_{(test)}$= 0.071; R^2_r = 0.166; Q^2_r = -0.222; cR^2_p = 0.721.

The QSAR model represented by this equation (eq. 1) could explain 76.1% variance and predicted 72.5% variance of biological activity. The external validation parameters were justified by R^2_{pred} value of 0.643 (>0.5).

A PLS regression study using three component and the same variables as the s-MLR method yielded a different equation as follows,

pIC_{50} = -53.643 + 17.651 *SpMax1_Bhv* – 9.775 *MATS3m* – 1.441 *SsssCH* – 0.029 *AATS4m* – 0.148 *VE3_Dze* + 0.152 *VE3_DzZ* – 2.153 *GATS4p* – 1.907 *SpMax4_Bhv* **Eq. 2**

n_{train}= 54; R= 0.883; R^2= 0.780; R^2_A= 0.766, SEE= 0.468; $Q^2_{(LOO)}$= 0.6852; *Avg.* $r_m^2{}_{(LOO)}$= 0.567; $\Delta r_m^2{}_{(LOO)}$=0.218; n_{test} = 18; R^2_{pred} = 0.666; *Avg.* $r_m^2{}_{(test)}$ = 0.567; $\Delta r_m^2{}_{(test)}$= 0.066;

Discussion

Here in eq. 2, it can be explored with this PLS model, we can explain 76.6% of the variance of the biological activity as well as can predict it with an accuracy of 68.5%. By the use of the test set the model was validated again and it was found that externally, the model has 66.6% predictability. Definitions of different descriptors are shown in **Table 8.1.**

Table 8.1: Description of the different descriptors

Sl No	Descriptor name	Description	Contribution
1	*SpMax1_Bhv*	Largest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative van der Waals volumes	Positive
2	*MATS3m*	Moran autocorrelation - lag 3 / weighted by mass	Negative
3	*SsssCH*	Sum of atom-type E-State: >CH-	Negative
4	*AATS4m*	Average Broto-Moreau autocorrelation - lag 4 / weighted by mass	Negative
5	*VE3_Dze*	Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by Sanderson electronegativities	Negative
6	*VE3_Dzz*	Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by atomic number	Positive
7	*GATS4P*	Geary autocorrelation - lag 4 / weighted by polarizabilities	Negative
8	*SpMax4_Bhv*	Largest absolute eigenvalue of Burden modified matrix - n 4 / weighted by relative van der Waals volumes	Negative

The *t*-values, *p*-level of the model are shown in the **Table 8.2.**

Table 8.2: VIF, t-Value and p-Value for Eq. 1

Parameters	*t*-value	*p*-level	VIF

Intercept	-4.385	0.000	--
SpMax1_Bhv	5.685	0.000	1.477
MATS3m	-6.865	0.000	1.923
SsssCH	-8.852	0.000	1.762
AATS4m	-7.394	0.000	1.879
VE3_Dze	-4.491	0.000	1.311
VE3_DzZ	3.177	0.003	1.467
GATS4p	-3.455	0.001	1.928
SpMax4_Bhv	-2.770	0.008	1.436

Moreover, chance using variable inflation factor (VIF) and the correlation matrix was also plotted for the verification of the said descriptors in the final model (**Table 8.3**). The p-value of the regression coefficient were found to be less than 0.05 and thus it remains in an acceptable criteria.

Table 3.3: Correlation matrix among biological activity and selected parameters of Eq. 1

	SpMax1_Bhv	*MATS3m*	*SsssCH*	*AATS4m*	*VE3_Dze*	*VE3_DzZ*	*GATS4p*	*SpMax4_Bhv*
SpMax1_Bhv	1.00	0.45	-0.33	-0.25	-0.03	-0.17	0.28	0.37
MATS3m		1.00	-0.58	-0.23	0.05	0.05	0.19	0.39
SsssCH			1.00	0.02	0.08	0.10	-0.17	-0.38
AATS4m				1.00	-0.07	-0.07	-0.63	-0.13
VE3_Dze					1.00	0.48	0.08	-0.04
VE3_DzZ						1.00	0.13	-0.22
GATS4p							1.00	-0.02
SpMax4_Bhv								1.00

Discussion

nterestingly, all of the other statistical modelling parameters (*i.e.*, R^2, Q^2, R^2_{pred}, scaled *Avg.* $r^2_{m~(LOO)}$, $\Delta r^2_{m~(LOO)}$, *Avg.* $r^2_{m~test}$, $\Delta r^2_{m~test}$) were within the acceptable limit as mentioned by Golbraikh and Tropsha. The following table can be considered as a validation (**Table 8.4**).

Table 8.4: Golbraikh and Tropsha acceptable model criteria of eq. 1

Parameter	Threshold value	eq. (10)
Q^2	$Q^2 > 0.5$	0.725
r^2	$r^2 > 0.6$	0.661
$\lvert r^2_0 - r^{2'}_0 \rvert$	$\lvert r^2_0 - r^{2'}_0 \rvert < 0.3$	0.037
k	$0.85 < k < 1.15$	0.989
k'	$0.85 < k' < 1.15$	0.998
$(r^2 - r^2_0) / r^2$	$(r^2 - r^2_0) / r^2 < 0.1$	0.028
$(r^2 - r^{2'}_0) / r^2$	$(r^2 - r^{2'}_0) / r^2 < 0.1$	0.083

In the following figure a comparison between the observed (Obs) and LOO-predicted (Pred) activities of the training and the test sets MMP-2 inhibitors has been visualized in **Figure 8.1 (A)**. Also the applicability domain (Euclidean distance method) can be found in **Figure 8.1 (B).**
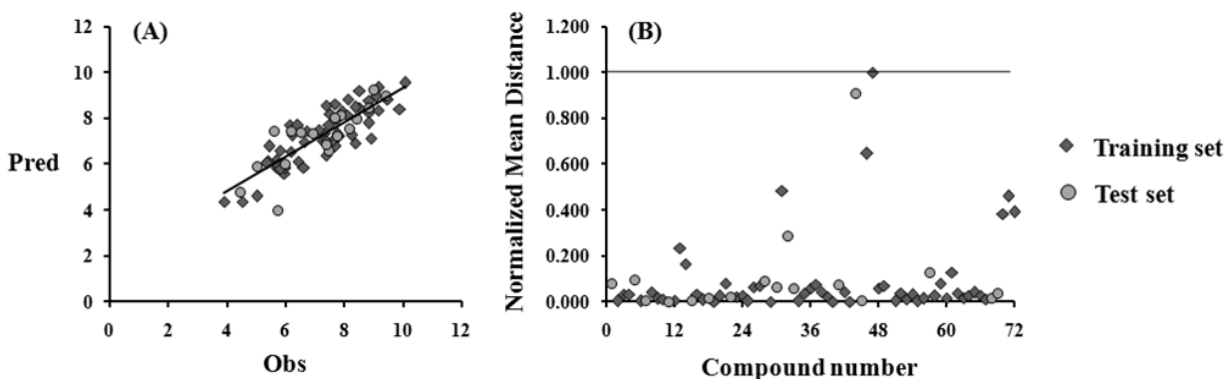
Figure 8.1 : (A) Observed (Obs) and LOO-predicted (Pred) activity of the training and test set inhibitors; (B) Test of applicability domain of eq. 1 by the Euclidean distance approach.

Discussion

The radar plots can be good comparative data visualization tool. **Figure 8.2** and **8.3** shows the comparative observed and predicted values obtained from the MLR model and the PLS model respectively.
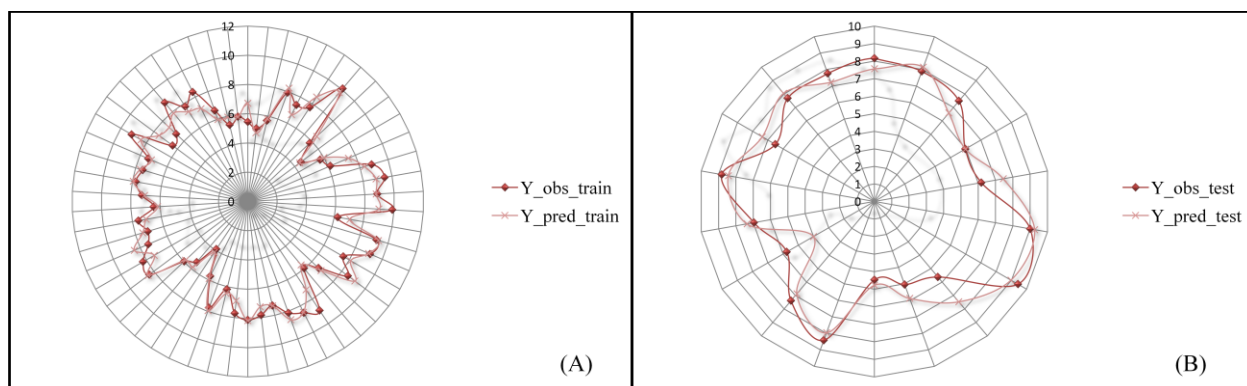
Figure 8.2: A radar plot showing the comparative view of Observed (Obs) and calculated (Calc) activity of (A) the train set and (B) the test set using the MLR method.

(A)   (B)
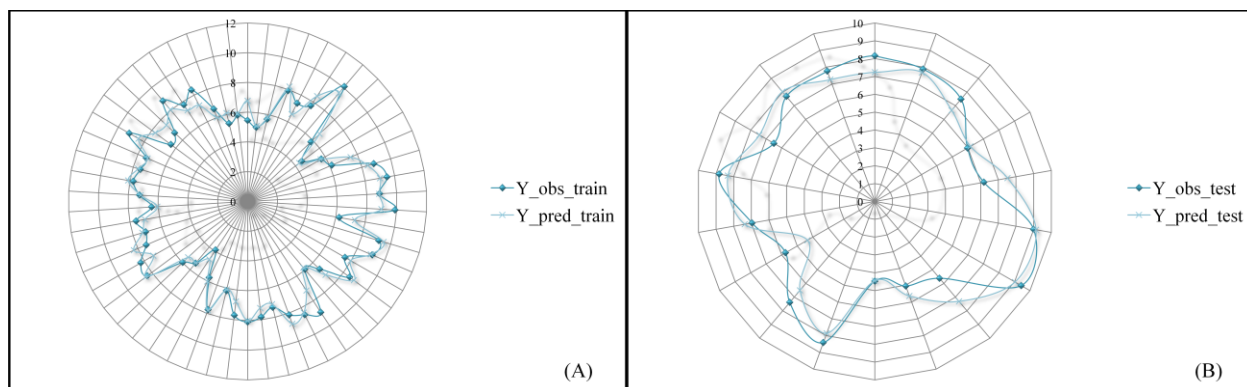
Figure 8.3: A radar plot showing the comparative view of Observed (Obs) and calculated (Calc) activity (A) the train set and (B) the test set using the PLS method.

Discussion

8.2 Mechanistic interpretation of the descriptors in the 2D-QSAR model

Both the MLR and the PLS models has consensually agreed on 8 descriptors. Among these, 2 descriptors, i.e., *SpMax1_Bhv, VE3_DzZ* contributed positively and *MATS3m, SsssCH, AATS4m, VE3_Dze, GATS4p, SpMax4_Bhv* were found to be negatively contributing in prediction of the biological activities of the molecules. A detailed description of these descriptors and their corresponding contribution is explained as follows.

8.2.3 SpMax1_Bhv:

The largest Eigen-value of a modified connectivity matrix weighted by van der Waals volume (n = 1) or Burden matrix (Roy et al, 2015) is the calculated value for the descriptor. The Eq. 1 shows that this descriptor contributes positively. This descriptor is related to the bond order as well as van der Waals volume. Compounds having the higher molecular weight due to bigger substitutions at R, R_1 and R_2 positions (cpds **17-22, 24-25, 27, 50-52, 62**) exhibit higher MMP-2 inhibitory activity. Comparing the structures of these compounds, it is observed that bulky unsaturated aryl substitution such as phenyl or *p*-methoxyphenyl, *p*-butoxyphenyl or *p*-methylthiophenyl substitution at R position may increase the van der Waals volume which, in turn, may increase the steric interaction at the S1′ hydrophobic pocket and thus increase the activity. Nevertheless, if the R_1 position is substituted with *i*-propoxy or *sec*-butyl groups which directs towards the S1 pocket and, on the other hand, the R_2 position is substituted with some bulky groups (e.g. acetamidoethyl, benzamidoethyl, methylsulfonamidoethyl or phenylacetamidoethyl) that are directed towards the S2 pocket may have some steric influence with the respective amino acids at these subpockets. Therefore, this feature exhibit positive

Discussion

impact on MMP-2 inhibitory activity. The cpds **30** and **32** have least values of this descriptor and also falls under the least active compounds (1.41 percentile and 5.63 percentile respectively) and these compounds have 36,000 nM and 9,000 nM activity. The cpds **21**, **22**, **24** shows a high value of the descriptor (>4.0) and these are among the top active compound (94 percentile) (sub-nanomolar activity of 0.13 nM, 0.38nM and 0.67nM respectively). However, compound **58** has the highest *SpMax1_Bhv* value but shows a moderate activity of 56 nM.

8.2.3 MATS3m

MATS3m is called the Moran's Coefficient derived from an autocorrelation function which takes geometry matrix in consideration weighted by the means of atomic mass and the lag time of the Moran's equation is considered to be 3. In the model, this descriptor contributes negatively to biological activity (Todeschini and Consonni, 2009).

Compounds with higher bulky substitution such as *n*-tetradecyloxy (cpd **13**), acetamidoethyl (cpd **23**), phthaloyl ethyl (cpd **30**), benzyloxy (cpd **32**) are lower active MMP-2 inhibitors. More lengthy and bulky substitutions may not enter properly into the respective pockets and therefore, are less active inhibitors. The most active compound (cpd **10**, IC_{50} = 0.087 nM), shows the least value of the descriptor and the compound with the highest value of the descriptor (No. 13**;** 2.81 percentile rank) has a very low biological value of 30000 nM. Likewise, two other molecules cpds **30** and **32** which have activities in the micromolar range show the bottom 4th (1.41 percentile) and 5th (5.63 percentile) rank of the descriptor. The high molecular mass of the side chain can result in such relation. For example, compound **13** with an *n*-tetradecyl chain at the R_1 position of the scaffold A has the largest aliphatic chain in the entire data set. Cpd **10**, the most

Discussion

active compound has the R_1 and R_2 positioned fused with a cyclic group. Interestingly, another cpd **32**, which also has a ring structure, lacks a nitrogen atom converting the ring into a planar orientation with lesser mass and spatial occupancy than the cpd **10** and thus both the biological activity and the descriptor *MATS3m* are average.

8.2.4 SsssCH

This descriptor, *SsssCH,* is an E-state descriptor that calculates the electrical as well as the topological parameters of a compound. It is the sum of the electrotopolological values of >CH= carbon atoms present in a molecule (Roy et al, 2015) The descriptor is a negative contributor to the biological activity, i.e. the compounds with lower values of this parameter will have higher inhibitory activity (cpds **17**-**25**) compared to compounds with comparatively higher values of this parameter (cpd **1**, **5**, **14**, **30**-**32**, **53**-**59**). Interestingly, the compound having an *i*-propoxy group at R_1 position is favourable. On the other hand, the compounds having other substituents at the same position (cpd **15** > cpd **16** > cpd **14**; cpd **6** > cpd **5**; cpd **25** > cpd **27**) have lesser activity. Therefore, it may be preferable as it may properly fit into the S1 pocket. In other words, smaller branched alkyl substitutions perfectly fit into the smaller S1 pocket for executing better MMP-2 inhibition. The cpd **24** has the least value of the descriptors and has an activity of 0.67 nM which is highly aceptable. However, the cpd **15** shows lesser activity than the other even after less value of the descriptor which might be the explained through other descriptors.

8.2.5 AATS3m

AATS3m is another 3D autocorrelation factor proposed by Moreau & Broto (1980) based on molecular graph with mass as the atomic property and a lag time of 3 is considered. The

116

Discussion

descriptor is negatively contributing. Compounds having higher molecular weight has shown to be lesser active (cpds **31**, **44**, **46-47**, **57**, **61**, **70-72**). Compounds having higher bulky substitution at R position such as 5-bromo-thiophene-2-yl (cpd **31**), dibromobenzyloxy (cpds **44**, **47**), 2-(4-fluorophenyl)-thiophene-5-yl (cpd **57**), (3, 5-dibromophenyl)CH_2O (cpds **70-72**) are lower active MMP-2 inhibitors. Probably, these compounds with bulky bromo substituents do not properly fit into the S1′ pocket and therefore, exerts lower efficacy.

8.2.6 VE3_Dze

VE3_Dze considers the logarithmic coefficient of the sum of the Barysz distance matrix the last eigenvector of using the Sanderson electronegativity as the weightage point (Todeschini and Consonni, 2009). This descriptor has a negative correlation with the biological activity of the molecules present in the preliminary dataset. Compounds **36**, **21**, **24**, **59** and **65** having some of the very high *VE3_Dze* value showed a moderate to high activity with nanomolar range. The compound with highest value of this descriptor, cpd **13** was found to be unfavourable for the activity (IC_{50} = 30,000 nM).

8.2.7 VE3_Dzz

VE3_Dzz is similar to the previous descriptor with a replacement of the electronegativity with atomic number. The descriptor has a positive coefficient which shows that higher value of the descriptor shows favourability towards the activity. This states that the presence of atoms with more electronegativity and Geary's auto-correlation is another spatial auto correlation factor that is operated on the 3D molecular graph (Todeschini and Consonni, 2009) .

Discussion

8.2.8 GATS4P

For the descriptor *GATS4P*, the considered molecular property is the polarizability of the molecule and the lag value is taken as k = 4. This descriptor precedes a negative coefficient and thus it can be said that the higher value of the descriptor would lead to detrimental MMP-2 inhibitory activity. Compounds with lower polarizability are therefore, better active MMP-2 inhibitors (cpds **1**, **5**, **10**, **18**, **25**-**27**, **34**-**35**, **38**, **59**) (Todeschini and Consonni, 2009).

8.2.9 SpMax4_Bhv

Similar to the other variable discussed earlier, the *SpMax1_Bhv* uses the Burden's Matrix and the 4th largest (n=4) eigenvalue of the matrix is considered. The descriptor *SpMax4_Bhv* has a negative impact on MMP-2 inhibitory activity. It is observed that compounds with more bulky substitution at R_1 and R_2 positions may impart lower inhibitory efficacy. It is observed that *p*-Bn-O-Bn (cpds **4**, **8**), *n*-tetradecyloxy (cpd **13**) at R_1 position as well as phthalylethyl (cpd **30**), trimethyleneaminocarboxybenzyl (cpds **71**, **72**), trimethyleneaminocarbamate (cpd **70**), $(C_6H_5CH_2O)CONHC_2H_4$ (cpd **69**) at R_2 position disfavours the MMP-2 inhibitory activity (Todeschini and Consonni, 2009).

8.3 Molecular orbital of the highest acting Compound 10 and the least active Compound 47:

According to the molecular orbital theory, the reacting orbital or the frontier orbital takes part in any chemical reaction. The highest occupied molecular orbital (HOMO) represents the strong nucleophilic region of the compound whereas, lowest unoccupied molecular orbital (LUMO)

Discussion

represents a highly electrophilic region. Figure shows the HOMO of Cpd 1 which is the most active compound. It is well visible that the orbital condensation is near the linker group of the compound which in turn proves the presence of high nucleophilic region near the sulphonyl thiane region. Through the 2D QSAR model, it is already established that the presence of a fused ring might be beneficial for the activity. This study farther establishes it.
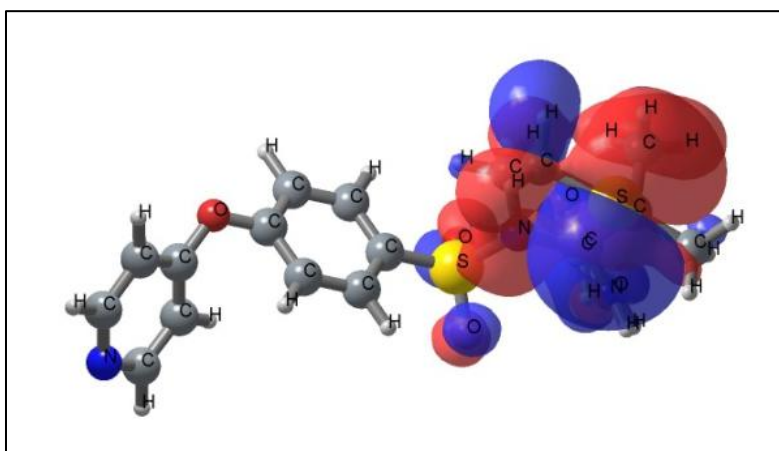
Figure 8.4: HOMO orbitals of Cpd 10

The electrophilic LUMO stays near the 4-phenoxy 4-pyridine structures shows that this particular region which possibly inserts into the S1' group will be favourable with a good electrophilic molecular profile.

Discussion
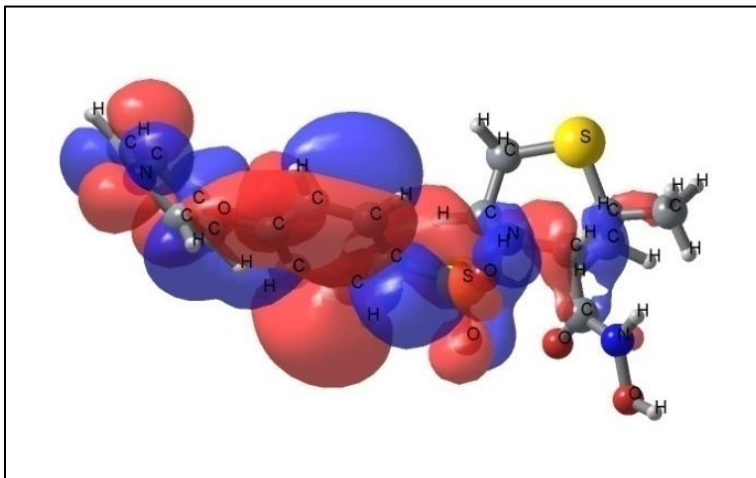
Figure 8.5: LUMO orbitals of CPD 10

When compared with the least active compound the HOMO is entirely different from that of the CPD 10. Both the bonding and anti-bonding orbitals are distributed around the m,m dibromo benzyl phenyl group instead of the linker group adjacent to the sulphonyl group.
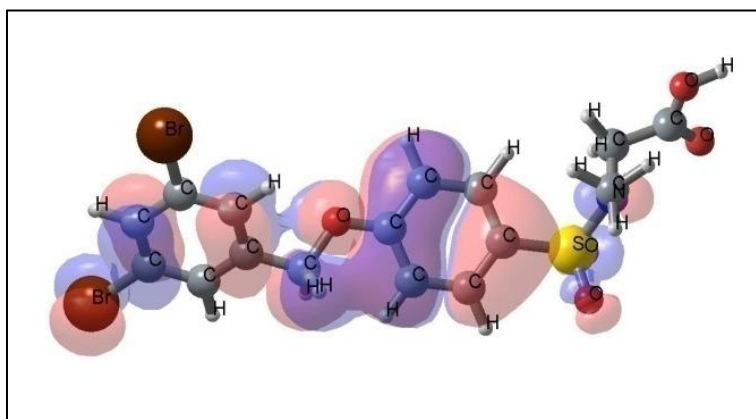
Figure 8.6: HOMO orbitals of CPD 47

The LUMO is distributed differently as well. The electrophilic LUMO is mainly distributed around the suphonylamdio hydroxamate group instead of the P1' group. This hinders the positively charged Zn^{2+} ion to be bound to the hydroxamate group resulting in a very low inhibitory activity.

Discussion
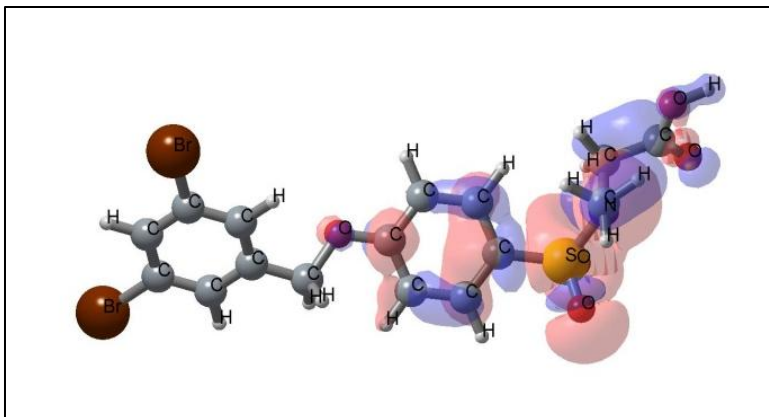
Figure 8.7: LUMO of CPD 47

Discussion

8.4 Bayesian classification of Dataset #2:

The Bayesian classification model which was constructed using the 2nd dataset shows some interesting and important responses especially in using the fingerprint descriptors as some important sub-structural features has been explored. ALogP, Molecular_Weight, Num_H_Donors, Num_H_Acceptors, Num_RotatableBonds, Num_Rings, Num_AromaticRings, Molecular_FractionalPolarSurfaceArea, ECFP_6 were the descriptor used for model generation.

The receiver operating characteristic or ROC score was good with a value of 0.817. The sensitivity (True positive prediction rate) and specificity value (False positive prediction rate) were in the acceptable range (0.652 and 0.956 respectively). All the detailed relevant data are provided in the following table 10.

Table 8.7: Cross validation result of the Bayesian classification model

ROC Score	ROC Rating	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	Concordance
0.817	Good	75	40	4	87	0.652	0.956	0.786

A ROC curve depicting the interrelationship between sensitivity and specificity of the training set is depicted as in figure 60.

Discussion
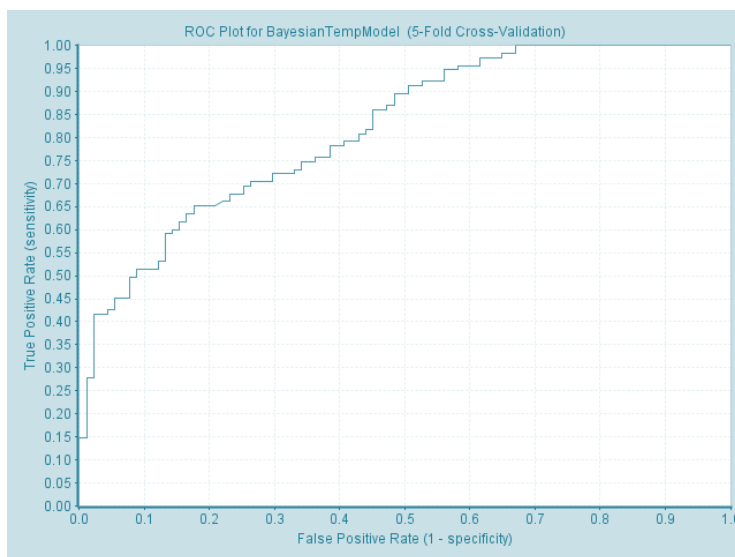
Figure 8.8: ROC curve of the Bayesian Classification model (Training set)

For the external validation the 25% of the dataset were used and the ROC value was found to be 0.848 and the specificity of the model was 0.968 whereas the sensitivity value was 0.649. A detailed data is presented in table YY.

Table 8.8: External validation parameter using the test set molecules

ROC Score	ROC Rating	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	Concordance
0.848	Good	24	13	1	30	0.649	0.968	0.794

The ROC diagram of the test set will be as the following figure 61.
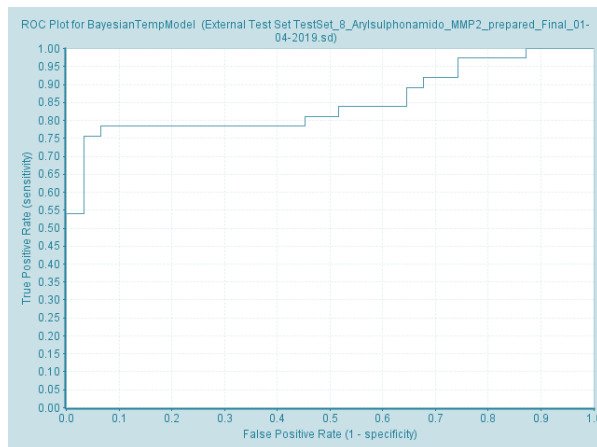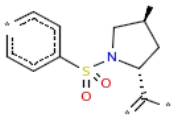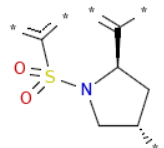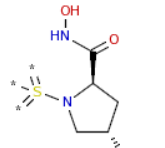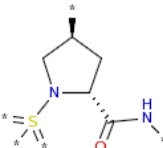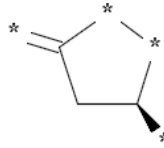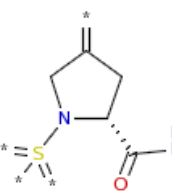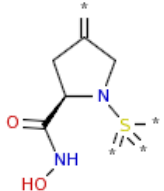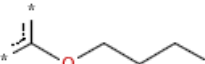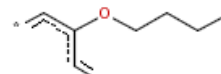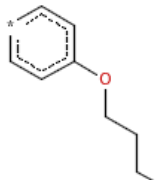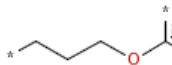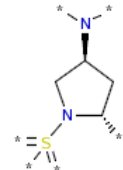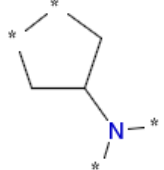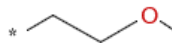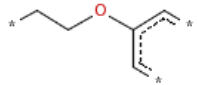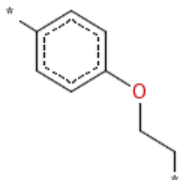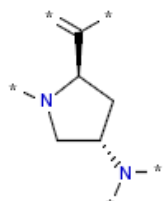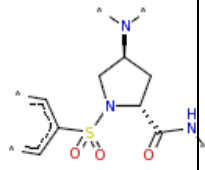
Figure 61: ROC curve of the valedictory test set

123

Discussion

The most important aspect of Bayesian classification is to include the 'Good' and 'Bad' fingerprints of the molecule. The list of good fingerprints that contributes positively is shown in **Table 8.9**.
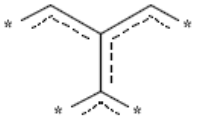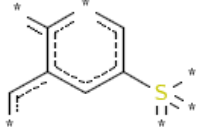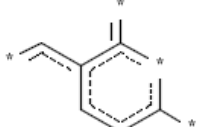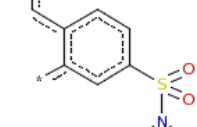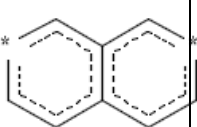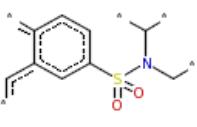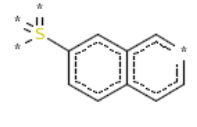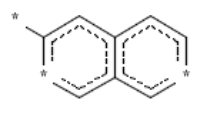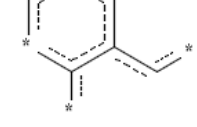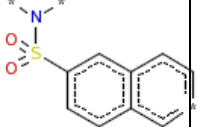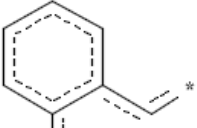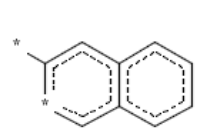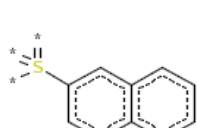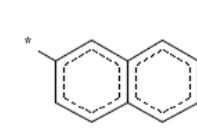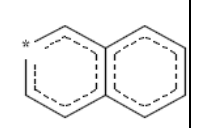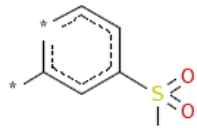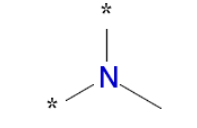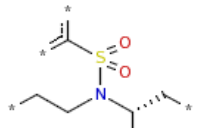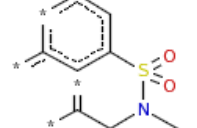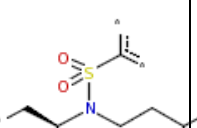
G1: -1026428813 18 out of 18 good Bayesian Score: 0.519	G2: 1785154134 18 out of 18 good Bayesian Score: 0.519	G3: -1763969961 17 out of 17 good Bayesian Score: 0.517	G4: 226405343 17 out of 17 good Bayesian Score: 0.517	G5: 53207596 16 out of 16 good Bayesian Score: 0.515
G6: -662736727 14 out of 14 good Bayesian Score: 0.509	G7: 1641501950 14 out of 14 good Bayesian Score: 0.509	G8: -313708941 10 out of 10 good Bayesian Score: 0.492	G9: -1244654081 10 out of 10 good Bayesian Score: 0.492	G10: 1776735893 10 out of 10 good Bayesian Score: 0.492
G11: 919059369 10 out of 10 good Bayesian Score: 0.492	G12: 822277761 10 out of 10 good Bayesian Score: 0.492	G13: -86640970 10 out of 10 good Bayesian Score: 0.492	G14: 1752998756 9 out of 9 good Bayesian Score: 0.485	G15: -1869628272 9 out of 9 good Bayesian Score: 0.485
G16: -1790412586 22 out of 23 good Bayesian Score: 0.484	G17: 1706085096 21 out of 22 good Bayesian Score: 0.481	G18: 1563344559 21 out of 22 good Bayesian Score: 0.481	G19: -707559679 8 out of 8 good Bayesian Score: 0.478	G20: 726376392 8 out of 8 good Bayesian Score: 0.478

Discussion

A list of all the bad fingerprints are shown in Table 13.

 B1: -178525456 0 out of 29 good Bayesian Score: -2.868	 B2: -1150899835 0 out of 28 good Bayesian Score: -2.835	 B3: 710652510 0 out of 27 good Bayesian Score: -2.801	 B4: 2146857511 0 out of 25 good Bayesian Score: -2.729	 B5: -755462605 0 out of 24 good Bayesian Score: -2.691
 B6: 1225196391 0 out of 24 good Bayesian Score: -2.691	 B7: 1929265201 0 out of 24 good Bayesian Score: -2.691	 B8: -2019199918 0 out of 24 good Bayesian Score: -2.691	 B9: 717474525 0 out of 23 good Bayesian Score: -2.651	 B10: -1364251345 0 out of 22 good Bayesian Score: -2.610
 B11: -81428579 0 out of 22 good Bayesian Score: -2.610	 B12: 1233434266 0 out of 21 good Bayesian Score: -2.567	 B13: -1173882748 0 out of 21 good Bayesian Score: -2.567	 B14: 1637591468 0 out of 21 good Bayesian Score: -2.567	 B15: 2006518499 0 out of 21 good Bayesian Score: -2.567
 B16: 1684711603 2 out of 29 good Bayesian Score: -1.770	 B17: 865379614 0 out of 7 good Bayesian Score: -1.611	 B18: 1120131233 0 out of 5 good Bayesian Score: -1.352	 B19: 1143581291 0 out of 5 good Bayesian Score: -1.352	 B20: -862200360 0 out of 5 good Bayesian Score: -1.352

In the above study we can easily visualize the presence of some of the important parameter of the substructural features present in the molecules of the dataset. For the positively contributing fingerprint, the substituted phenylsulphonyl group is a indeed a favourable group as it can be

Discussion

found in **G1**, **G2**, **G3**, **G4**, **G5**, **G6**, **G14**, **G20** fingerprints. However, the same group acts differently when a fused ring (**B7, B10 and B13**) or a secondary or tertiary amine group (**B4, B16, B19, B20 and B18**) is juxtaposed with it. The sulphonyl group is a strong electron withdrawing group favourable for protein binding properties.

On the other hand, the presence of electron donating *sec-* and *tert-* amines may compensate the presence of sulphonyl group by alloweing the electron to drift towards the sulfur atom. Interestingly, same tertiary amine group can have a huge difference if it not an open chain but fused into a cyclic azolidine group (**G1- G4, G6- G7 and G20**). From the preliminary study we have seen that the presence of this sulfonyl- azolidine group is a favourable factor which may be due to its probable insertion into the very shallow S2' pocket.

The focus of this study is hydroxamates. It is the Zn^{2+} binding group and therefore will be ubiquitous to all these molecules. The good fingerprints **G4, G6, G7** and **G20** shows that the presence of the hydroxamate group in the 2 position of substituted N-sulphonyl azolidine group will be acceptable for biological activity of the selected molecules.

Apart from these, another important parameter is the presence of the alkoxy group (**G8 – G13**) shows favourable activities. This may be due to flexibility as well the possible hydrophobicity of the group that tends to bind with the S1 pocket of the enzyme.

In the negative group the major common factor for nearly most of the fingerprints is the presence of the naphthalene or fuse ring structure. This group is bulky and planar in nature. Being close to the sulphonyl group, this is a S1' directing group. The bulky planar structure of this group might not be acceptable for enzyme binding.
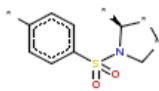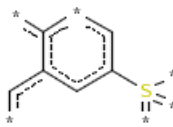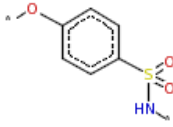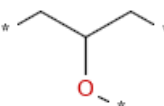
8.5 Recursive partitioning of Dataset #2:

Discussion

8.5.1 Decision tree model:

As discussed earlier the classification of the recursive partitioning method rely on the division of

the training data to multiple nodes unless all of the comounds are classified properly. The

The majorly important fragments are shown in Table 14.

Table 8.11: FCFP _6 fingerprints used for tree division

| FCFP_6 1275657259 | FCFP_6 -1272798659 | FCFP_6 -1317581692 | FCFP_6 -1085821960 | FCFP_6 -1043310069 |

Among three different trees generated, the statistical parameter of the first tree were in the

acceptable range with an ROC score of 0.884 and a cross-validated ROC of 0.756. For the model

development set, the number of true positive and true negative is found to be optimal (105 TP

and 66 FN) as well as 25 false positive and 10 false negative were calculated. Detailed report of

the model is provided in Table 15.

Table 8.12: Summarization of the recursive partitioning report using the decision tree method.

Tree Reports						
Info	ConfusionMatrix		YName	ROCScore	BestTree	ROCScoreCV
**** Best Tree **** Tree 1: 9 leaves Error Rate (training data): 37.253 Min alpha: 0	Actual\Pred. 1 0 / 1 105 10 / 0 25 66		Binary	0.88428	true	0.75566
Tree 2: 5 leaves Error Rate (training data): 46.771 Min alpha: 2.655	Actual\Pred. 1 0 / 1 83 32 / 0 16 75		Binary	0.80081	-	0.72785
Tree 3: 2 leaves	Actual\Pred. 1 0		Binary	0.73106	-	0.74935

| Error Rate (training data): 55.402 Min alpha: 3.541 | | 1 | 86 | 29 | | | | | |
| | | 0 | 26 | 65 | | | | | |

In case of external validation, the first tree showed a fair ROC value of 0.715 with 16 true negative and 28 true positive values of biological activities of the compound. 15 compounds were predicted as false positive and 9 of the test set compounds were false negative. The confusion matrix and the detailed statistical parameters are provided in Table 16.

Table 8.13: Summarization of the test set prediction result using the decision tree method.

Test Results					
ConfusionMatrix			TreeID	YName	ROCScore	ROCRating
Actual\Pred.	0	1	1	Binary	0.71534	Accuracy 0.715: Fair
0	16	15				
1	9	28				
Actual\Pred.	0	1	2	Binary	0.65824	Accuracy 0.658: Poor
0	20	11				
1	15	22				
Actual\Pred.	0	1	3	Binary	0.59852	Accuracy 0.599: Fail
0	17	14				
1	13	24				

One of the interesting observations of the decision tree is that the molecules have been, at the very beginning of the model development based on the fragment no1275657259 which is actually the derived form of N- (phenyl sulphonyl) azolidine group. This is coinciding with all the previous discussion with the 2D QSAR study and the Bayesian classification model
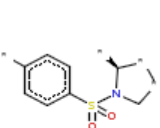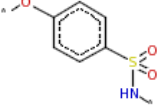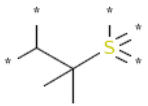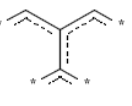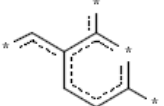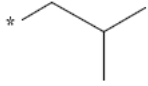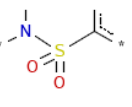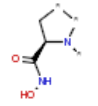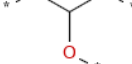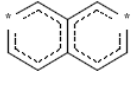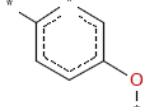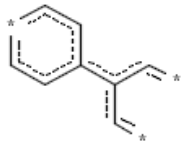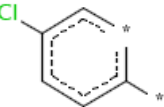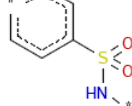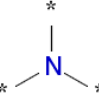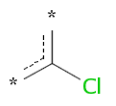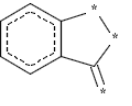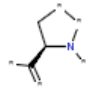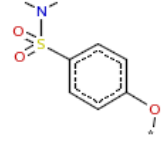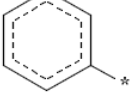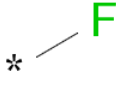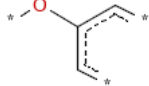
8.5.2 The random forest model:
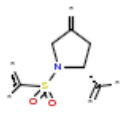
The random forest model was also created to cross verify the presence of any overfitting error. It was found that this method is also in consensus with the previous decision tree based model. Several moleucular fragments (mentioned in Table 17) were considered to make this model. The ROC values (0.955 in bag, 0.789 out of bag) shows acceptable result. The important fingerprints

Discussion

(FCFP_6) is shown in **Table 8.14**. The statistical data and confusion matrix are shown in **Table 8.15.**

FCFP_6 1275657259	FCFP_6 -1085821960	FCFP_6 -415245925	FCFP_6 307419094	FCFP_6 -1320007763
FCFP_6 -1272798659	FCFP_6 -1272768868	FCFP_6 -1043339860	FCFP_6 -1096517202	FCFP_6 2387394
FCFP_6 -1043310069	FCFP_6 -105186863	FCFP_6 1674451008	FCFP_6 1764617907	FCFP_6 551850122
FCFP_6 -580192194	FCFP_6 9	FCFP_6 71476542	FCFP_6 -1698724694	FCFP_6 -1946918893
FCFP_6 -165009748	FCFP_6 -2093839777	FCFP_6 32	FCFP_6 332760439	FCFP_6 941223536
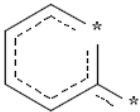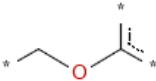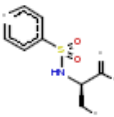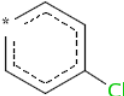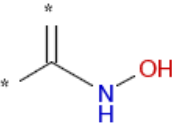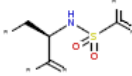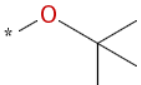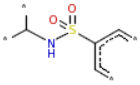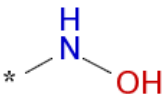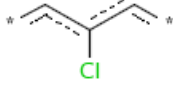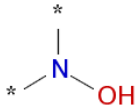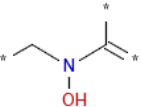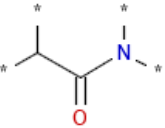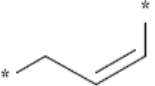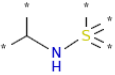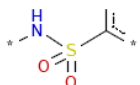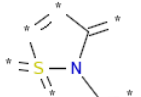
FCFP_6 991735244	FCFP_6 1036089772	FCFP_6 -1255089061	FCFP_6 -745491832	FCFP_6 1666606459
FCFP_6 -239082946	FCFP_6 34028999	FCFP_6 -415216134	FCFP_6 873872329	FCFP_6 -548542844
FCFP_6 367998008	FCFP_6 -548602426	FCFP_6 -1682956756	FCFP_6 565998553	FCFP_6 451847724
FCFP_6 -885550502	FCFP_6 -1096219292	FCFP_6 -1553874037		

Table 8.15: Statistical parameter of the Random Forest method of recursive partitioning

Model Information	Y Property	Confusion Matrix			ROC Score
In-bag training data results for 10 trees	Binary		1	0	0.9557
		1	101	14	
		0	15	76	
Out-of-bag training data results for 10 trees	Binary		1	0	0.7893
		1	85	29	
		0	31	60	

The test set statistical validation along with the confusion matrix is explained below **Table 8.16**.

Table 8.16: The confusion matrix of test set prediction with random forest recursive partitioning model:

ConfusionMatrix			YName	ROCScore	ROCRating
Actual\Pred.	0	1			
0	22	9	Binary	0.78901	Accuracy 0.789: Fair
1	11	26			

9. Conclusion and future perspective

Conclusion and future perspective

MMP-2 has been a tried and tested target of anti metastatic drugs. However, the pan MMP inhibitors which were explored in the late 20th century had a major setback due to their adverse drug effect and got rejected into many phases of clinical trial. However, due to the chief mechanism of action of MMP-2 inhibitors, especially their role in inhibition of the VEGF, the possible drug can be an important choice for prevention of metastasis of cancer.

The structure of MMP-2 has been a point of interest of the structural biologist circa 1999. The protein, several subdomains, especially the catalytic sites have been elucidated through X-Ray crystallography and NMR crystallography. Several structure-based drug discovery techniques has been important to understand the possible interaction of the enzyme with the small molecules, both synthetic and natural product alike. On the other hand, irrespective of the availability of the structure, a fruitful inhibitor fails to arise. The intervention of ligand based drug designing may prove to be equally important in this way.

The purpose of this study is to pave a preliminary study of designing a target specific MMP-2 inhibitor and the first strategy was to utilize the most common and effective drug designing strategy of QSAR. This study has used a small number of molecules (72) and with 2D QSAR method some important molecular properties has been unearthed. The presence of a hydroxamic acid group and at the same time a cyclic 5 member ring proved to be important as a fused ring structure juxtaposed to the ring system attached to the amide group has proven to be important as well.

In the next step we have done a preliminary quantum chemical study to identify whether any future endeavors are possible. Interestingly, it was found that the position of the HOMO and

Conclusion and future perspective

LUMO can make a great difference in the activity for the best active and the least active compound. A detailed study with quantum chemical descriptors such as HOMO energy, LUMO energy, HOMO-LUMO energy gap, Fukui function, molecular hardness etc. might help in a better model development and at the same time a better mechanistic approach to explain the SAR.

In farther steps, the larger set proved useful for classification study. Though a 2D QSAR study was performed, the methodology failed to construct a faithful model. The online database or the Binding Database might be an important source of molecules with their protein/ enzyme binding activity, the source does not seem to be well maintained as the very same molecules shows dubious activity in different phases making it harder to map using a model. However, the classification studies proved to be effective and several molecular fingerprints were found to be useful. The presence of the medium length chain, biphenyl group, ring structure at the amine group can be proven to be effective and at the same time, the presence of a naphthalene group can be deplorable.

In summary, it can be clearly said that this study holds a premium importance as a preliminary study on finding newer molecular characteristics of selective MMP-2 inhibitors. A further study with a more focused dataset and advanced methods of 2D QSAR, 3D QSAR, quantum QSAR, Molecular Docking and simulation can be useful for future work.

10. Reference

- Abbasi, M., Ramezani, F., Elyasi, M., Sadeghi-Aliabadi, H., & Amanlou, M. (2015). A study on quantitative structure–activity relationship and molecular docking of metalloproteinase inhibitors based on L-tyrosine scaffold. DARU Journal of Pharmaceutical Sciences, 23(1). https://doi.org/10.1186/s40199-015-0111-z

- Adhikari, N., Amin, S. A., Saha, A., & Jha, T. (2017). Exploring in house glutamate inhibitors of matrix metalloproteinase-2 through validated robust chemico-biological quantitative approaches. Structural Chemistry, 29(1), 285–297. https://doi.org/10.1007/s11224-017-1028-6

- Adhikari, N., Amin, S. A., Saha, A., & Jha, T. (2017). Understanding Chemico-Biological Interactions of Glutamate MMP-2 Inhibitors through Rigorous Alignment-Dependent 3D-QSAR Analyses. ChemistrySelect, 2(26), 7888–7898. https://doi.org/10.1002/slct.201701330

- Adhikari, N., Amin, S. A., Saha, A., & Jha, T. (2018). Structural exploration for the refinement of anticancer matrix metalloproteinase-2 inhibitor designing approaches through robust validated multi-QSARs. Journal of Molecular Structure, 1156, 501–515. https://doi.org/10.1016/j.molstruc.2017.12.005

- Adhikari, N., Halder, A. K., Mallick, S., Saha, A., Saha, K. D., & Jha, T. (2016). Robust design of some selective matrix metalloproteinase-2 inhibitors over matrix metalloproteinase-9 through in silico/fragment-based lead identification and de novo lead modification: Syntheses and biological assays. Bioorganic & Medicinal Chemistry, 24(18), 4291–4309. https://doi.org/10.1016/j.bmc.2016.07.023

- Adhikari, N., Mukherjee, A., Saha, A., & Jha, T. (2017). Arylsulfonamides and selectivity of matrix metalloproteinase-2: An overview. European Journal of Medicinal Chemistry, 129, 72–109. https://doi.org/10.1016/j.ejmech.2017.02.014

- Adhikari, N., Mukherjee, A., Saha, A., & Jha, T. (2017). Arylsulfonamides and selectivity of matrix metalloproteinase-2: An overview. European Journal of Medicinal Chemistry, 129, 72–109. https://doi.org/10.1016/j.ejmech.2017.02.014

- Ahmad A, Sayed A, Ginnebaugh KR, Sharma V, Suri A, Saraph et al (2015) Molecular docking and inhibition of matrix metalloproteinase-2 by novel difluorinatedbenzylidene curcumin analog. Am J Transl Res 7: 298-308.

- Ajmani, S., Jadhav, K., & Kulkarni, S. A. (2009). Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR & Combinatorial Science, 28(1), 36–51. https://doi.org/10.1002/qsar.200810063

- Altruda, F., Poli, V., Restagno, G., Argos, P., Cortese, R., & Silengo, L. (1985). The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal, repeating homology. Nucleic Acids Research, 13(11), 3841–3859. https://doi.org/10.1093/nar/13.11.3841

- Ambure, P., Aher, R. B., Gajewicz, A., Puzyn, T., & Roy, K. (2015). "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. Chemometrics and Intelligent Laboratory Systems, 147, 1–13. https://doi.org/10.1016/j.chemolab.2015.07.007

- Ammazzalorso, A., De Filippis, B., Campestre, C., Laghezza, A., Marrone, A., Amoroso, R., … Agamennone, M. (2016). Seeking for Non-Zinc-Binding MMP-2 Inhibitors: Synthesis, Biological Evaluation and Molecular Modelling Studies. International Journal of Molecular Sciences, 17(10), 1768. https://doi.org/10.3390/ijms17101768

- Barksby, H. E., Milner, J. M., Patterson, A. M., Peake, N. J., Hui, W., Robson, T., … Rowan, A. D. (2006). Matrix metalloproteinase 10 promotion of collagenolysis via procollagenase activation: Implications for cartilage degradation in arthritis. Arthritis & Rheumatism, 54(10), 3244–3253. https://doi.org/10.1002/art.22167

- Bar-Or, A. (2003). Analyses of all matrix metalloproteinase members in leukocytes emphasize monocytes as major inflammatory mediators in multiple sclerosis. Brain, 126(12), 2738–2749. https://doi.org/10.1093/brain/awg285

- Bergmann, U., Tuuttila, A., Tryggvason, K., & Morgunova, E. (2002). Catalytic domain of proMMP-2 E404Q mutant. Protein Data Bank, Rutgers University. https://doi.org/10.2210/pdb1eak/pdb

- Bode, W. (1995). A helping hand for collagenases: the haemopexin-like domain. Structure, 3(6), 527–530. https://doi.org/10.1016/s0969-2126(01)00185-x

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394–424. https://doi.org/10.3322/caac.21492

- Briknarová, K., Gehrmann, M., Bányai, L., Tordai, H., Patthy, L., & Llinás, M. (2001). Gelatin-binding Region of Human Matrix Metalloproteinase-2. Journal of Biological Chemistry, 276(29), 27613–27621. https://doi.org/10.1074/jbc.m101105200

- Briknarová, K., Grishaev, A., Bányai, L., Tordai, H., Patthy, L., & Llinás, M. (1999). The second type II module from human matrix metalloproteinase 2: structure, function and dynamics. Structure, 7(10), 1235-S2. https://doi.org/10.1016/s0969-2126(00)80057-x

- Cai, J., Li, C., Liu, Z., Du, J., Ye, J., Gu, Q., & Xu, J. (2017). Predicting DPP-IV inhibitors with machine learning approaches. Journal of Computer-Aided Molecular Design, 31(4), 393–402. https://doi.org/10.1007/s10822-017-0009-6

- de Melo, E. B. (2012). A QSAR Study of Matrix Metalloproteinases Type 2 (MMP-2) Inhibitors with Cinnamoyl Pyrrolidine Derivatives. Scientia Pharmaceutica, 80(2), 265–281. https://doi.org/10.3797/scipharm.1112-21

- Dhanaraj, V., Williams, M. G., Ye, Q.-Z., Molina, F., Johnson, L. L., Ortwine, D. F., … Blundell, T. L. (1999). CRYSTAL STRUCTURE OF GELATINASE A CATALYTIC DOMAIN. Protein Data Bank, Rutgers University. https://doi.org/10.2210/pdb1qib/pdb

- Di Pizio, A., Agamennone, M., & Aschi, M. (2012). An Integrated Computational Approach to Rationalize the Activity of Non-Zinc-Binding MMP-2 Inhibitors. PLoS ONE, 7(11), e47774. https://doi.org/10.1371/journal.pone.0047774

- Di Pizio, A., Laghezza, A., Tortorella, P., & Agamennone, M. (2013). Probing the S1′ Site for the Identification of Non- Zinc- Binding MMP- 2 Inhibitors. ChemMedChem, 8(9), 1475–1482. https://doi.org/10.1002/cmdc.201300186

- Durrant, J. D., de Oliveira, C. A. F., & McCammon, J. A. (2009). Including receptor flexibility and induced fit effects into the design of MMP-2 inhibitors. Journal of Molecular Recognition, n/a-n/a. https://doi.org/10.1002/jmr.989

- Durrant, J. D., de Oliveira, C. A. F., & McCammon, J. A. (2011). Pyrone-Based Inhibitors of Metalloproteinase Types 2 and 3 May Work as Conformation-Selective Inhibitors. Chemical Biology & Drug Design, 78(2), 191–198. https://doi.org/10.1111/j.1747-0285.2011.01148.x

- European Bioinformatics InstituteProtein Information Resource SIB Swiss Institute of Bioinformatics. (2019, April 10). 72 kDa type IV collagenase. Retrieved from https://www.uniprot.org/uniprot/P08253

- Fabre, B., Filipiak, K., Zapico, J. M., Díaz, N., Carbajo, R. J., Schott, A. K., … de Pascual-Teresa, B. (2013). Progress towards water-soluble triazole-based selective MMP-2 inhibitors. Organic & Biomolecular Chemistry, 11(38), 6623. https://doi.org/10.1039/c3ob41046c

- Fanjul-Fernández, M., Folgueras, A. R., Cabrera, S., & López-Otín, C. (2010). Matrix metalloproteinases: Evolution, gene regulation and functional analysis in mouse models. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1803(1), 3–19. https://doi.org/10.1016/j.bbamcr.2009.07.004

- Feng, Y., Likos, J. J., Zhu, L., Woodward, H., Munie, G., McDonald, J. J., … Stallings, W. C. (2002). Solution structure and backbone dynamics of the catalytic domain of matrix metalloproteinase-2 complexed with a hydroxamic acid inhibitor. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1598(1–2), 10–23. https://doi.org/10.1016/s0167-4838(02)00307-2

- Fernández, M., Caballero, J., & Tundidor-Camba, A. (2006). Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. Bioorganic & Medicinal Chemistry, 14(12), 4137–4150. https://doi.org/10.1016/j.bmc.2006.01.072

- G. Moreau and C. Turpin, Use of similarity analysis to reduce large molecular libraries to smaller sets of representative molecules, Analusis, 24 (1996), pp. M17–M21.

- G. Moreau and P. Broto, Autocorrelation of molecular structures. Application to SAR studies, Nouv. J. Chim. 4 (1980), pp. 757–764.

- Gao, M., Zhang, H., Trivedi, A., Mahasenan, K. V., Schroeder, V. A., Wolter, W. R., … Chang, M. (2016). Selective Inhibition of MMP-2 Does Not Alter Neurological Recovery after Spinal Cord Injury. ACS Chemical Neuroscience, 7(11), 1482–1487. https://doi.org/10.1021/acschemneuro.6b00217

- Gapstur, S. M., Drope, J. M., Jacobs, E. J., Teras, L. R., McCullough, M. L., Douglas, C. E., … Brawley, O. W. (2018). A blueprint for the primary prevention of cancer: Targeting established, modifiable risk factors. CA: A Cancer Journal for Clinicians, 68(6), 446–470. https://doi.org/10.3322/caac.21496

- Gearing, A. J. H., Beckett, P., Christodoulou, M., Churchill, M., Clements, J., Davidson, A. H., … Woolley, K. (1994). Processing of tumour necrosis factor-α precursor by metalloproteinases. Nature, 370(6490), 555–557. https://doi.org/10.1038/370555a0

- Gehrmann, M., Briknarová, K., Bányai, L., Patthy, L., & Llinás, M. (2002). The Col-1 Module of Human Matrix Metalloproteinase-2 (MMP-2): Structural/Functional Relatedness between Gelatin-Binding Fibronectin Type II Modules and Lysine-Binding Kringle Domains. Biological Chemistry, 383(1). https://doi.org/10.1515/bc.2002.014

- Geurts, N., Martens, E., Van Aelst, I., Proost, P., Opdenakker, G., & Van den Steen, P. E. (2008). β-Hematin Interaction with the Hemopexin Domain of Gelatinase B/MMP-9 Provokes Autocatalytic Processing of the Propeptide, Thereby Priming Activation by MMP-3†. Biochemistry, 47(8), 2689–2699. https://doi.org/10.1021/bi702260q

- Gohlke, U., & Bode, W. (1996). C-TERMINAL DOMAIN (HAEMOPEXIN-LIKE DOMAIN) OF HUMAN MATRIX METALLOPROTEINASE-2. Protein Data Bank, Rutgers University. https://doi.org/10.2210/pdb1rtg/pdb

- Golbraikh, A., & Tropsha, A. (2002). Journal of Computer-Aided Molecular Design, 16(5/6), 357–369. https://doi.org/10.1023/a:1020869118689

- Gupta, S. P., Kumar, D., & Kumaran, S. (2003). A quantitative structure–activity relationship study of hydroxamate matrix metalloproteinase inhibitors derived from funtionalized 4-

aminoprolines. Bioorganic & Medicinal Chemistry, 11(9), 1975–1981. https://doi.org/10.1016/s0968-0896(03)00069-5

- Halder, A. K., Saha, A., & Jha, T. (2013). Exploring QSAR and pharmacophore mapping of structurally diverse selective matrix metalloproteinase-2 inhibitors. Journal of Pharmacy and Pharmacology, 65(10), 1541–1554. https://doi.org/10.1111/jphp.12133

- Hashimoto, H., Takeuchi, T., Komatsu, K., Miyazaki, K., Sato, M., & Higashi, S. (2011). Crystal structure of MMP-2 active site mutant in complex with APP-drived decapeptide inhibitor. Protein Data Bank, Rutgers University. https://doi.org/10.2210/pdb3ayu/pdb

- Jamloki, A., Karthikeyan, C., Hari Narayana Moorthy, N. S., & Trivedi, P. (2006). QSAR analysis of some 5-amino-2-mercapto-1,3,4-thiadiazole based inhibitors of matrix metalloproteinases and bacterial collagenase. Bioorganic & Medicinal Chemistry Letters, 16(14), 3847–3854. https://doi.org/10.1016/j.bmcl.2006.04.014

- Jenne D., Stanley K.K. (1987) Nucleotide sequence and organization of the human S-protein gene: repeating peptide motifs in the "pexin" family and a model for their evolution. Biochemistry.;26(21):6735-42.

- Jha, S., Kanaujia, S. P., & Limaye, A. M. (2016). Direct inhibition of matrix metalloproteinase-2 (MMP-2) by (−)-epigallocatechin-3-gallate: A possible role for the fibronectin type II repeats. Gene, 593(1), 126–130. https://doi.org/10.1016/j.gene.2016.07.061

- Jha, T., Adhikari, N., Saha, A., & Amin, S. A. (2017). Multiple molecular modelling studies on some derivatives and analogues of glutamic acid as matrix metalloproteinase-2 inhibitors. SAR and QSAR in Environmental Research, 29(1), 43–68. https://doi.org/10.1080/1062936x.2017.1406986

- Jha, T., Adhikari, N., Saha, A., & Amin, S. A. (2017). Multiple molecular modelling studies on some derivatives and analogues of glutamic acid as matrix metalloproteinase-2 inhibitors. SAR and QSAR in Environmental Research, 29(1), 43–68. https://doi.org/10.1080/1062936x.2017.1406986

- Jozic, D., Bourenkov, G., Lim, N.-H., Visse, R., Nagase, H., Bode, W., & Maskos, K. (2004). X-ray Structure of Human proMMP-1: NEW INSIGHTS INTO PROCOLLAGENASE ACTIVATION AND COLLAGEN BINDING. Journal of Biological Chemistry, 280(10), 9578–9585. https://doi.org/10.1074/jbc.m411084200

- Kolb, C., Mauch, S., Peter, H.-H., Krawinkel, U., & Sedlacek, R. (1997). The matrix metalloproteinase RASI-1 is expressed in synovial blood vessels of a rheumatoid arthritis patient. Immunology Letters, 57(1–3), 83–88. https://doi.org/10.1016/s0165-2478(97)00057-6

- Kumar, D., & Gupta, S. P. (2003). A quantitative structure–activity relationship study on some matrix metalloproteinase and collagenase inhibitors. Bioorganic & Medicinal Chemistry, 11(3), 421–426. https://doi.org/10.1016/s0968-0896(02)00438-8

- Kumaran, S., & Gupta, S. P. (2007). A quantitative structure-activity relationship study on matrix metalloproteinase inhibitors: Piperidine sulfonamide aryl hydroxamic acid analogs. Journal of Enzyme Inhibition and Medicinal Chemistry, 22(1), 23–27. https://doi.org/10.1080/14756360600956655

- Li, Y.-X., Li, Y., Je, J.-Y., & Kim, S.-K. (2015). Dieckol as a novel anti-proliferative and anti-angiogenic agent and computational anti-angiogenic activity evaluation. Environmental Toxicology and Pharmacology, 39(1), 259–270. https://doi.org/10.1016/j.etap.2014.11.027

- Libson, A. M., Gittis, A. G., Collier, I. E., Marmer, B. L., Goldberg, G. I., & Lattman, E. E. (1995). Crystal structure of the haemopexin-like C-terminal domain of gelatinase A. Nature Structural & Molecular Biology, 2(11), 938–942. https://doi.org/10.1038/nsb1195-938

- Lima, G. D. de A., Rodrigues, M. P., Mendes, T. A. de O., Moreira, G. A., Siqueira, R. P., da Silva, A. M., … Teixeira, R. R. (2018). Synthesis and antimetastatic activity evaluation of cinnamic acid derivatives containing 1,2,3-triazolic portions. Toxicology in Vitro, 53, 1–9. https://doi.org/10.1016/j.tiv.2018.07.015

- Lu, Y., Papagerakis, P., Yamakoshi, Y., Hu, J. C.-C., Bartlett, J. D., & Simmer, J. P. (2008). Functions of KLK4 and MMP-20 in dental enamel formation. Biological Chemistry, 389(6). https://doi.org/10.1515/bc.2008.080

- Medeiros Turra, K., Pineda Rivelli, D., Berlanga de Moraes Barros, S., & Fernanda Mesquita Pasqualoto, K. (2014). Predicting Novel Antitumor Agents: 3D-Pharmacophore Mapping of β-N-biaryl Ether Sulfonamide-Based Hydroxamates as Potentially MMP-2 Inhibitors. Molecular Informatics, 33(9), 573–587. https://doi.org/10.1002/minf.201400073

- Morgunova, E. (1999). Structure of Human Pro-Matrix Metalloproteinase-2: Activation Mechanism Revealed. Science, 284(5420), 1667–1670. https://doi.org/10.1126/science.284.5420.1667

- Morgunova, E., Tuuttila, A., Bergmann, U., & Tryggvason, K. (2002). Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. Proceedings of the National Academy of Sciences, 99(11), 7414–7419. https://doi.org/10.1073/pnas.102185399

- Moroy, G., Denhez, C., El Mourabit, H., Toribio, A., Dassonville, A., Decarme, M., … Bourguet, E. (2007). Simultaneous presence of unsaturation and long alkyl chain at P1′ of Ilomastat confers selectivity for gelatinase A (MMP-2) over gelatinase B (MMP-9) inhibition as shown by

molecular modelling studies. Bioorganic & Medicinal Chemistry, 15(14), 4753–4766. https://doi.org/10.1016/j.bmc.2007.05.001

- Motrescu, E. R., Blaise, S., Etique, N., Messaddeq, N., Chenard, M.-P., Stoll, I., … Rio, M.-C. (2008). Matrix metalloproteinase-11/stromelysin-3 exhibits collagenolytic function against collagen VI under normal and malignant conditions. Oncogene, 27(49), 6347–6355. https://doi.org/10.1038/onc.2008.218

- Murphy, G., & Knäuper, V. (1997). Relating matrix metalloproteinase structure to function: Why the "hemopexin" domain? Matrix Biology, 15(8–9), 511–518. https://doi.org/10.1016/s0945-053x(97)90025-1

- NAGASE, H., VISSE, R., & MURPHY, G. (2006). Structure and function of matrix metalloproteinases and TIMPs. Cardiovascular Research, 69(3), 562–573. https://doi.org/10.1016/j.cardiores.2005.12.002

- Nanjan, P., Nambiar, J., Nair, B. G., & Banerji, A. (2015). Synthesis and discovery of (I-3,II-3)-biacacetin as a novel non-zinc binding inhibitor of MMP-2 and MMP-9. Bioorganic & Medicinal Chemistry, 23(13), 3781–3787. https://doi.org/10.1016/j.bmc.2015.03.084

- Nicolotti, O., Catto, M., Giangreco, I., Barletta, M., Leonetti, F., Stefanachi, A., … Carotti, A. (2012). Design, synthesis and biological evaluation of 5-hydroxy, 5-substituted-pyrimidine-2,4,6-triones as potent inhibitors of gelatinases MMP-2 and MMP-9. European Journal of Medicinal Chemistry, 58, 368–376. https://doi.org/10.1016/j.ejmech.2012.09.036

- Nuti, E., Cantelmo, A. R., Gallo, C., Bruno, A., Bassani, B., Camodeca, C., … Rossello, A. (2015). N-O-Isopropyl Sulfonamido-Based Hydroxamates as Matrix Metalloproteinase Inhibitors: Hit Selection and in Vivo Antiangiogenic Activity. Journal of Medicinal Chemistry, 58(18), 7224–7240. https://doi.org/10.1021/acs.jmedchem.5b00367

- Nuti, E., Casalini, F., Avramova, S. I., Santamaria, S., Cercignani, G., Marinelli, L., … Rossello, A. (2009). N-O-Isopropyl Sulfonamido-Based Hydroxamates: Design, Synthesis and Biological Evaluation of Selective Matrix Metalloproteinase-13 Inhibitors as Potential Therapeutic Agents for Osteoarthritis. Journal of Medicinal Chemistry, 52(15), 4757–4773. https://doi.org/10.1021/jm900261f

- Nuti, E., Casalini, F., Santamaria, S., Fabbi, M., Carbotti, G., Ferrini, S., … Rossello, A. (2013). Selective Arylsulfonamide Inhibitors of ADAM-17: Hit Optimization and Activity in Ovarian Cancer Cell Models. Journal of Medicinal Chemistry, 56(20), 8089–8103. https://doi.org/10.1021/jm4011753

- Nuti, E., Casalini, F., Santamaria, S., Gabelloni, P., Bendinelli, S., Da Pozzo, E., … Rossello, A. (2011). Synthesis and biological evaluation in U87MG glioma cells of

(ethynylthiophene)sulfonamido-based hydroxamates as matrix metalloproteinase inhibitors. European Journal of Medicinal Chemistry, 46(7), 2617–2629. https://doi.org/10.1016/j.ejmech.2011.03.033

- Overall, C. M. (2002). Molecular Determinants of Metalloproteinase Substrate Specificity: Matrix Metalloproteinase Substrate Binding Domains, Modules, and Exosites. Molecular Biotechnology, 22(1), 051–086. https://doi.org/10.1385/mb:22:1:051

- P.A.P. Moran, Notes on continuous stochastic phenomena, Biometrika 37 (1950), pp. 17–23.

- Pandey, A., Patnaik, R., Bhattacharya, P., Shukla, S., & Paul, S. (2015). Resveratrol inhibits matrix metalloproteinases to attenuate neuronal damage in cerebral ischemia: a molecular docking study exploring possible neuroprotection. Neural Regeneration Research, 10(4), 568. https://doi.org/10.4103/1673-5374.155429

- Qiu, H.-Y., Wang, Z.-C., Wang, P.-F., Yan, X.-Q., Wang, X.-M., Yang, Y.-H., & Zhu, H.-L. (2014). Design, synthesis, evaluation and 3D-QSAR analysis of benzosulfonamide benzenesulfonates as potent and selective inhibitors of MMP-2. RSC Advances, 4(74), 39214. https://doi.org/10.1039/c4ra06438k

- Qiu, H.-Y., Wang, Z.-C., Wang, P.-F., Yan, X.-Q., Wang, X.-M., Yang, Y.-H., & Zhu, H.-L. (2014). Design, synthesis, evaluation and 3D-QSAR analysis of benzosulfonamide benzenesulfonates as potent and selective inhibitors of MMP-2. RSC Advances, 4(74), 39214. https://doi.org/10.1039/c4ra06438k

- Rio, M. C. (2005). From a unique cell to metastasis is a long way to go: clues to stromelysin-3 participation. Biochimie, 87(3–4), 299–306. https://doi.org/10.1016/j.biochi.2004.11.016

- Roy, K., Kar, S. & Das, R. (2015). Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Amsterdam Boston: Academic Press, an imprint of Elsevier.

- Sanyal, S., Amin, S. A., Adhikari, N., & Jha, T. (2019). QSAR modelling on a series of arylsulfonamide-based hydroxamates as potent MMP-2 inhibitors. SAR and QSAR in Environmental Research, 30(4), 247–263. https://doi.org/10.1080/1062936x.2019.1588159

- Sjøli, S., Nuti, E., Camodeca, C., Bilto, I., Rossello, A., Winberg, J.-O., … Adekoya, O. A. (2016). Synthesis, experimental evaluation and molecular modelling of hydroxamate derivatives as zinc metalloproteinase inhibitors. European Journal of Medicinal Chemistry, 108, 141–153. https://doi.org/10.1016/j.ejmech.2015.11.019

- Song, J., Peng, P., Chang, J., Liu, M.-M., Yu, J.-M., Zhou, L., & Sun, X. (2016). Selective non-zinc binding MMP-2 inhibitors: Novel benzamide Ilomastat analogs with anti-tumor metastasis.

Bioorganic & Medicinal Chemistry Letters, 26(9), 2174–2178. https://doi.org/10.1016/j.bmcl.2016.03.064

- Sounni, N. E., & Noel, A. (2005). Membrane type-matrix metalloproteinases and tumor progression. Biochimie, 87(3–4), 329–342. https://doi.org/10.1016/j.biochi.2004.07.012

- SYBYL-X 2.0 Software, Tripos Inc., St. Louis. MO, USA (2012) http://www.certara.com

- Tallant, C., Marrero, A., & Gomis-Rüth, F. X. (2010). Matrix metalloproteinases: Fold and function of their catalytic domains. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1803(1), 20–28. https://doi.org/10.1016/j.bbamcr.2009.04.003

- Tauro, M., Laghezza, A., Loiodice, F., Agamennone, M., Campestre, C., & Tortorella, P. (2013). Arylamino methylene bisphosphonate derivatives as bone seeking matrix metalloproteinase inhibitors. Bioorganic & Medicinal Chemistry, 21(21), 6456–6465. https://doi.org/10.1016/j.bmc.2013.08.054

- Tester, A. M., Cox, J. H., Connor, A. R., Starr, A. E., Dean, R. A., Puente, X. S., … Overall, C. M. (2007). LPS Responsiveness and Neutrophil Chemotaxis In Vivo Require PMN MMP-8 Activity. PLoS ONE, 2(3), e312. https://doi.org/10.1371/journal.pone.0000312

- Todeschini R, Consonni V, Gramatica P. (2009). Chemometrics in QSAR. In: Comprehensive chemometrics (Volume 4). Brown S, Tauler R, Walczak R, (Ed.) Oxford, Elsevier, UK, 129- 172

- Tosco, P., & Balle, T. (2010). Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. Journal of Molecular Modeling, 17(1), 201–208. https://doi.org/10.1007/s00894-010-0684-x

- Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure and Applied Chemistry, 70(5), 1129–1143. https://doi.org/10.1351/pac199870051129

- Werner, S. R., Dotzlaf, J. E., & Smith, R. C. (2008). MMP-28 as a regulator of myelination. BMC Neuroscience, 9(1). https://doi.org/10.1186/1471-2202-9-83

- Yan, X.-Q., Wang, Z.-C., Li, Z., Wang, P.-F., Qiu, H.-Y., Chen, L.-W., … Zhu, H.-L. (2015). Sulfonamide derivatives containing dihydropyrazole moieties selectively and potently inhibit MMP-2/MMP-9: Design, synthesis, inhibitory activity and 3D-QSAR analysis. Bioorganic & Medicinal Chemistry Letters, 25(20), 4664–4671. https://doi.org/10.1016/j.bmcl.2015.08.026

- Yap, C. W. (2010). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry, 32(7), 1466–1474. https://doi.org/10.1002/jcc.21707

- Yasunori Okada (2017) in Kelley and Firestein's textbook of rheumatology. Edited by Firestein, G., Gabriel, S., McInnes, I. & Dell, J. (2017). Philadelphia, PA: Elsevier.

- Zheng, J., Wen, R., & Guillaume, D. (2008). Three-dimensional quantitative structure-activity relationship (CoMFA and CoMSIA) studies on galardin derivatives as gelatinase A (matrix metalloproteinase 2) inhibitors. Journal of Enzyme Inhibition and Medicinal Chemistry, 23(4), 445–453. https://doi.org/10.1080/14756360701632221

- Zhu, H., Fang, H., Cheng, X., Wang, Q., Zhang, L., Feng, J., & Xu, W. (2009). 3D-QSAR study of pyrrolidine derivatives as matrix metalloproteinase-2 inhibitors. Medicinal Chemistry Research, 18(8), 683–701. https://doi.org/10.1007/s00044-008-9160-x

- Zucker, S., Pei, D., Cao, J., & Lopez-Otin, C. (2003). Membrane type-matrix metalloproteinases (MT-MMP). In Current Topics in Developmental Biology (pp. 1–74). Elsevier. https://doi.org/10.1016/s0070-2153(03)54004-2

- Lipinski, C. A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. Drug Discovery Today: Technologies, 1(4), 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007