INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior Agrária

Natural products as collagenase inhibitors: a molecular modelling approach

Eduardo Miguel da Costa Carvalho

Dissertation submitted to Escola Superior Agrária de Bragança
to obtain the Degree of Master in Biotechnological Engineering

Supervised by

Rui Miguel Vaz de Abreu

Cristiano Gabi dos Santos Mateus

Bragança
2025

Abstract

Skin photoaging is a process that involves the degradation of extracellular matrix proteins, including collagen and elastin. These are essential components of the skin, providing sturdiness and elasticity; their degradation leads to the loss of tensile strength, flexibility, and the development of wrinkles. This process is mediated by matrix metalloproteinases (MMPs), including Collagenase (MMP-1). Excessive production of MMP-1 accelerates extrinsic skin ageing, which may be mitigated by applying anti-ageing compounds with anti-collagenase activities. Therefore, there's interest in discovering better MMP-1 inhibitors, with emphasis on natural compounds, since there's been a significant shift in market interest towards the usage of natural products in cosmeceutical development. One way to quickly test a wide range of compounds with minimal costs is to rely on *in silico* methodologies to predict compound bioactivities.

This study aimed to explore different molecular modelling approaches to assess the inhibition potential of different natural products against MMP-1, namely molecular docking and quantitative structure-activity relationship (QSAR) modelling. For this purpose, a virtual library of 83 known MMP-1 inhibitors was developed and used to develop QSAR models. QSAR model 2 was selected for further use since it presented solid statistical parameters, with an R^2 value of 0.96 and a RMSE value of 0.191.
A library of natural compounds, with a structure similar to compounds used to develop the QSAR models, was constructed using the COCONUT database of natural compounds. A total of 715 compounds were gathered, and the QSAR model 2 was applied to it.

The molecular docking approach was also validated by performing a Re-Docking protocol, and the validated approach was applied in the same library of 715 compounds.

Finally, the data from the two methods were combined. Using this data, the tested compounds were ranked according to their combined inhibition probability (%) and the top 10 ranked compounds were analyzed in terms of their structure and binding conformations. In the future, it would be of interest to buy these compounds and test them experimentally to validate these *in silico* models and confirm the predicted inhibition activity. If this activity is confirmed, then these compounds could be used in cosmeceutical and drug design applications.

Keywords: Collagenase, MMP-1, virtual screening, QSAR, molecular docking, COCONUT, natural products, inhibition activity

Resumo

O fotoenvelhecimento da pele é um processo que envolve a degradação de proteínas da matriz extracelular, incluindo o colagénio e a elastina. Estes são componentes essenciais da pele, responsáveis pela sua firmeza e elasticidade; a sua degradação conduz à perda de resistência à tração, flexibilidade e ao desenvolvimento de rugas. Este processo é mediado pelas metaloproteinases (MMPs), incluindo a colagenase (MMP-1). A produção excessiva de MMP-1 acelera o envelhecimento cutâneo extrínseco, o que pode ser atenuado pela aplicação de compostos anti-envelhecimento com atividades anti-colagenase. Assim, existe um interesse crescente na descoberta de melhores inibidores da MMP-1, com ênfase em compostos naturais, uma vez que se tem verificado uma mudança significativa no interesse do mercado para a utilização de produtos naturais no desenvolvimento de cosméticos.

Uma forma de testar rapidamente uma grande variedade de compostos com custos reduzidos é recorrer a metodologias *in silico* para prever as bioatividades dos compostos.

O presente estudo teve como objetivo explorar diferentes abordagens de modelação molecular para avaliar o potencial inibitório de diversos produtos naturais contra a MMP-1, nomeadamente molecular docking e QSAR modelling. Para este fim, foi desenvolvida uma biblioteca virtual composta por 83 inibidores conhecidos da MMP-1, a qual foi utilizada para o desenvolvimento dos modelos QSAR. O modelo QSAR 2 foi selecionado para uso posterior, uma vez que apresentou parâmetros estatísticos sólidos, com um valor de R^2 de 0,96 e um valor de RMSE de 0,191.

Uma biblioteca de compostos naturais, com estrutura semelhante à dos compostos utilizados no desenvolvimento dos modelos QSAR, foi construída recorrendo à base de dados COCONUT de compostos naturais. No total, foram reunidos 715 compostos, aos quais foi aplicado o modelo QSAR 2.

A abordagem de molecular docking foi igualmente validada através da realização de um protocolo de Re-Docking, tendo a metodologia validada sido aplicada à mesma biblioteca de 715 compostos.

Por fim, os dados provenientes dos dois métodos foram combinados. Com base nestes dados, os compostos testados foram classificados de acordo com a sua probabilidade combinada de inibição (%) e os 10 compostos mais bem classificados foram analisados em termos da sua estrutura e conformações de ligação. No futuro, seria de interesse adquirir estes compostos e testá-los experimentalmente para validar os

modelos *in silico* e confirmar a atividade inibitória prevista. Caso esta atividade seja confirmada, estes compostos poderão ser utilizados em aplicações de design cosmética e farmacêutico.

Palavras-chave: Colagenase, MMP-1, triagem *in silico* de compostos, QSAR, molecular docking, COCONUT, produtos naturais, atividade inibitória

Table of Contents

Index of Tables

Index of Figures

List of Abbreviations

MMP-1 – Interstitial collagenase

QSAR – Quantitative structure-activity relationship

UV – Ultraviolet

ECM – Dermal extracellular matrix

PDB – Protein Data Bank

3D – Three-dimensional

OCHEM - Online Chemical Modelling Environment

IC_{50} - Half-Maximal Inhibitory Concentration

PLS - Partial Least Squares

MLR - Multiple Linear Regression

CV-LOO – Cross Validation Leave one-out

kNN – K-nearest neighbors

R^2 – Determination coefficient

ΔG – Gibbs free energy of binding

Ki – Inhibition constant

1. Introduction

1.1. Skin Photoaging

The skin is the largest organ in the human body and is the boundary between the body's internal and external environment. It serves a wide range of functions, including acting as a physical and immunologically protective barrier or maintaining the skin's homeostasis. However, beyond these primary functions, the skin also tremendously impacts our social relationships. Having healthy skin is a significant part of our appearance and is something that is highly strived for. Therefore, avoiding damage to the skin is crucial, both for health and aesthetic reasons [1].

On that note, skin ageing is a natural and inevitable process that progressively leads to a loss of structural integrity and physiological function of the skin. Intrinsic ageing is unavoidable and depends on genetically determined rates that can't be controlled. On the other side, the extrinsic factors are controllable and include a wide variety of environmental factors, such as exposure to sunlight or pollution, and health aspects, like diet or sleeping habits. Constant exposure to ultraviolet (UV) radiation and other environmental factors damages the dermal matrix, leading to premature skin ageing. As the matrix is damaged, the body will try to repair the skin, leading to imperfect reparations. These will have a cumulative effect on the skin, leading to the formation of wrinkles, sagging of the skin, or changes in pigmentation [2].

Many studies have associated premature ageing with the stimulated production of various enzymes that degrade the components of the dermal extracellular matrix (ECM), such as MMP-1. The strength and durability of the skin depend on the proper arrangement of collagen fibrils and elastin in the dermis, meaning that excessive production of enzymes that degrade these components will gradually contribute to accelerated extrinsic ageing. There has been an effort to find reliable anti-ageing compounds from natural sources to prevent this phenomenon. These include phenolic compounds, flavonoids and other compounds that exhibit anti-collagenase activities. For example, in a study by Kumud M. & Sanju N. (2018), safranal obtained from saffron flowers was identified as a reliable inhibitor of matrix metalloproteinases, showing significant potential. For this reason, saffron flowers are reliable agents that can be included in herbal anti-ageing formulations [3]. On that note, Chiocchio et al. (2018) tested a hundred plant extracts *in vitro* against tyrosinase and elastase to assess their potential value as anti-ageing

compounds and skin protectors, finding 17 different extracts with strong active inhibition on both enzymes and a selective elastase inhibitor – *Hypericum scruglii*. [4].

Historically, natural products like these have been extensively used in drug discovery and various therapeutic procedures. Despite having significant advantages over synthetic molecules, such as displaying higher safety and an optimized structure for various biological functions, the usage of these products has declined over the years. This can be attributed to various difficulties in its application, such as challenges in identifying the bioactive compounds of interest, lack of compatibility with traditional screening assays, difficulties in obtaining sufficient material for characterization purposes, and the impossibility of patenting certain compounds [5]. However, due to the rising need for novel drugs to tackle the surge of new diseases and increasing drug resistance, there has been a renewed focus on the usage of natural products. With the use of innovative computational, analytical and automation methods, it's possible to understand the profile of these complex products and quickly develop new drugs or structural analogues [6].

1.2. Matrix metalloproteinases (MMPs)

Metalloproteinases are endopeptidases that contain an active site ZN^{2+} and are divided into subfamilies depending on their catalytic domain structure and evolutionary connections. The metzincin subfamily of metalloproteases are distinguished by a 3-histidine zinc-binding motif and a conserved methionine turn after the active site. This family consists of various members such as reprolysins, serralysins, astacins and the matrixins (also known as MMPs) [7].

These MMPs are usually associated with the degradation of the ECM proteins, the main components behind the dermal connective tissue. They can be classified as collagenases, gelatinases, stromelysins and membrane type MMPs, depending on their substrate specificities and whether they are secreted or bound to the membrane. The MMP gene family consists of 25 different members, with 24 of them being expressed in mammals [8].

1.2.1. Interstitial Collagenase

Collagens are one of mammalians' main components of the ECM, being present in their skin, ligaments, and tendons. These tissues consist of different types of collagens, each with their own characteristics. The most abundant type is type I collagen, consisting of three α-polypeptides that weave around each other to form a triple helix with non-

helical moieties, telopeptides, on either end of it. This collagen type is highly resistant to common proteases and can only be cleaved by enzymes with collagenolytic activity, such as MMPs [9].

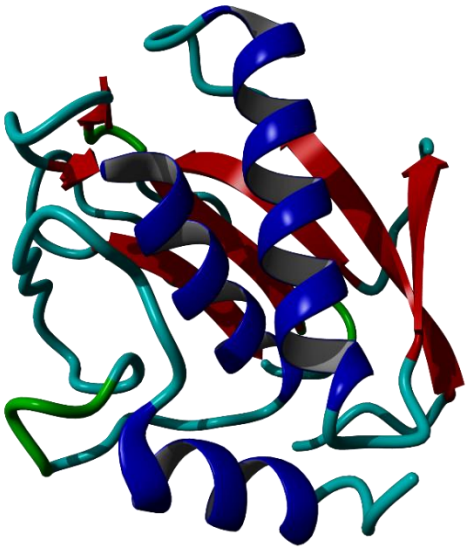


Figure 1 – Crystal structure of mature truncated MMP-1 (PDB ID: 1HFC) [10].

Collagenases are an important part of the MMP family, being able to cleave other molecules found within the cells. The MMP-1 (**Figure 1**), also known as fibroblast collagenase and interstitial collagenase, is synthesized as a proenzyme, and activated by proteolysis, through removal of the N-terminal residues. Its structure is like other MMPs, containing a catalytic domain and a carboxy terminal domain [11].

They cleave the X-gly bond of collagen (**Figure 2**) and any synthetic peptides that contain the sequence -Pro-X-Gly-Pro-, where X is an amino acid with the amino terminus blocked [12].

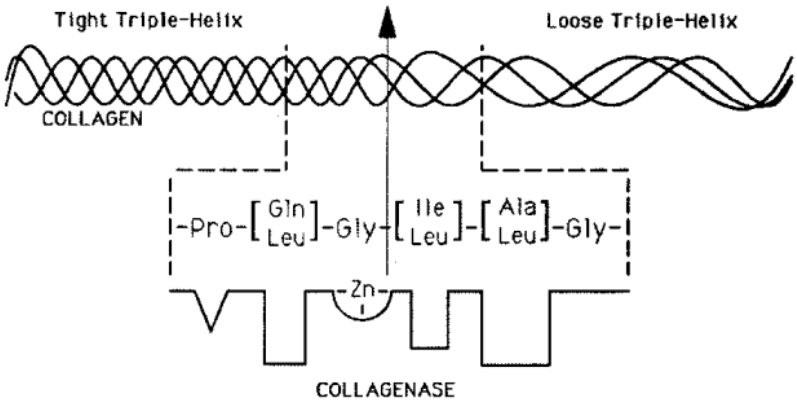


Figure 2 – Collagenase cleavage site in interstitial collagens. The cleaving is aimed at the border between a tight triple helix region and a loose triple helix. Adapted from: [13]

Considering that collagen is the main molecular component involved in the construction of human skin and the one responsible for its tensile strength, its degradation and hydrolysis will significantly impact the skin's health. The upregulation of MMP enzymes secreted by fibroblasts and other cells will promote collagen breakdown and decrease its synthesis, causing a cumulative breakdown of the connective tissues. This can be naturally caused by old age or excessive sun exposure: photoageing [14].

1.3. In Silico Approach: Molecular Modelling

Finding natural products with cosmetic or cosmeceutical viability is challenging, requiring extensive testing. This can be somewhat mitigated using *in silico* methods. These tools alleviate the time, cost, and scale issues, providing an effective and fast method of virtual screening large chemical libraries. These results will then serve as a basis for posterior *in vitro/in vivo* tests that experimentally confirm the computationally obtained predictions. Essentially, these tools can effectively remove inefficient chemical compounds from consideration and highlight the ones with the most potential for a specific purpose without any experimental testing [15].

These computational methods usually rely on previously obtained biological data, such as protein structural information, chemical properties, bioactivity information or toxicity. For this purpose, various databases are available online, each with different focuses and available catalogues, as represented in **Table 1**.

There are two main types of *in silico* virtual screening tools, each with different fundamentals and approaches [15]:

a. Ligand-based screening - relies on previously obtained data on active compounds and can predict new ones with similar effects (e.g. QSAR modelling, Pharmacophore modelling).
b. Structure-based screening – calculates the binding affinity between a ligand and a target protein using structure information (e.g. molecular docking, molecular dynamics simulation).

This work will focus on two main methods: molecular docking and QSAR modelling.

Table 1 – Main databases of chemical compounds available online.

Database	Number of compounds	URL	Description
PubChem	Over 115M	https://pubchem.ncbi.nlm.nih.gov/	Chemical database developed by the NIH (National Institute of Health)
ZINC15	Over 750M	https://zinc15.docking.org/	Database of commercially available compounds for virtual screening
COCONUT	Around 407k	https://coconut.naturalproducts.net/	Resource dedicated to storage, search, and analysis of Natural Products
PHAROS	No information	https://pharos.nih.gov	Interface containing a wide variety of data, including protein targets, small molecule activity and genomic behavior
ChEMBL	Around 2.4 M	https://www.ebi.ac.uk/chembl/	Database of bioactive molecules with drug-like properties

1.3.1. Molecular docking

Molecular docking is a tool that aims to predict the experimental binding mode and affinity of a small molecule within the binding site of a particular receptor target. It's considered a standard computational tool for drug design and provides an effective method of optimizing compounds and finding new biologically active molecules.

The main objective is to understand and predict molecular recognition structurally by finding different possible binding modes and energetically by predicting binding affinity. These results are obtained by predicting the ligand conformation, including its position and orientation, within the protein binding site and assessing the quality of the pose using a scoring function. This score represents the potentiality of binding, allowing the user to rank the different poses according to their binding affinity [16].

As seen in **Figure 3**, the first step in using a molecular docking tool is obtaining the 3D structures of the input molecules: both the target protein and the ligand(s) to be tested. These structures are usually obtained by experimental techniques such as X-ray crystallography or nuclear magnetic resonance and can be easily accessed and downloaded from online databases. Each experimental 3D structure has different properties, which may affect its quality and reliability for docking tests. The principal property is the crystal structure resolution parameter, which indicates the quality of the data collected on the crystal containing the protein. The resolution parameter indicates the level of detail present in the diffraction pattern and the level of detail seen when the

electron density map is calculated. Higher resolution structures (equal or lower than 2 Å) are highly ordered and provide a more extensive insight into every atom of the structure, making it easier to pinpoint their locations [17].

After obtaining the necessary structures, the selected protein must be optimized and modified to be used appropriately by the molecular docking tool. This step includes the addition of any missing extra atoms or residues and the usage of an energy minimization experiment to get rid of any steric interference. Additionally, the protonation states of the ionizable residues are established, water molecules and extraneous ligands are eliminated from the structure and force field parameters are supplied to the structure to adequately reflect the protein's behavior during docking [18]. Once the structure is ready, the protein target binding site must be identified to set the grid box correctly for the docking tool. The docking experiment will be confined in the defined box, meaning that it's essential that the binding site is set within its limits.

Finally, the docking tool may be used to obtain multiple ligand poses in the binding site defined within the prepared grid box. The software's scoring algorithm will rank the obtained poses by predicting the Gibbs free energy of binding (**ΔG**) of that interaction: lower values will imply higher likelihoods of that binding interaction occurring. After the process is complete, the results will be stored on a file, allowing the user to choose the most promising results according to their needs [19].

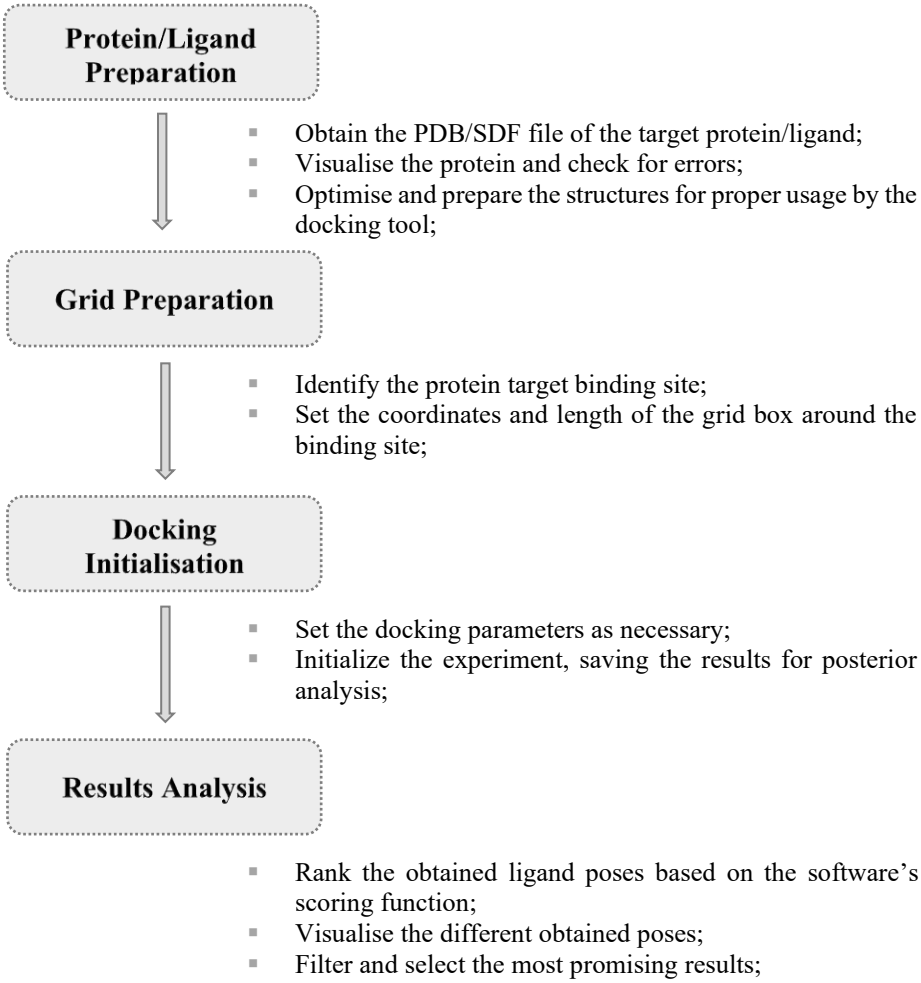


Figure 3 - Diagram depicting the step-by-step process of a generic docking experiment. Adapted from: [19].

1.3.2. QSAR Modelling

The development of a QSAR model relies on the creation of a mathematical correlation between a chemical or biological molecular response and a set of quantitative chemical attributes that define the properties of the analyzed molecule. The main principle behind this method is that variations in structural properties cause different biological activities. By establishing a predictive correlation model and using a complex mathematical algorithm, this method aims to explore and predict different properties of untested chemical compounds, such as toxicological responses, physicochemical properties, and biological activity. The extracted information can then be utilized for various purposes, such as modifying chemical structures to maximize the efficiency of a specific property and biological response or to develop new chemicals for use in a pharmaceutical context [20]. Such utility is extremely valuable for different scientific areas and industries, including drug design, material science and predictive toxicology.

The application of QSAR modelling depends on its capacity to prioritize many chemicals quickly and effectively in their desired biological activities, significantly reducing the number of future candidates to test with in vivo experiments. Essentially, it's a fast and reliable *in silico* method that quickens the process of chemical testing [21].

Its development process is complex and requires several steps, as represented in **Figure 4**.

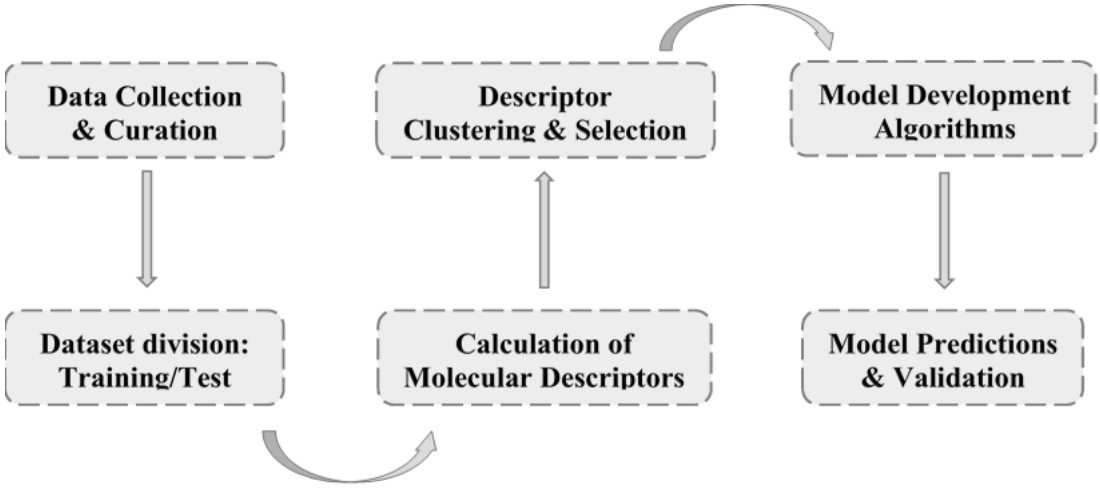


Figure 4 – Diagram describing the steps required to develop a QSAR model.

For the model to be reliable and robust, it's necessary to define and collect the desired chemical attributes carefully. These are usually provided as precise quantitative information that was either obtained using experimental analysis or a suitable theoretical algorithm. This information will be vital in establishing the interpretative relationship that addresses a given response. Additionally, the obtained data should be curated to avoid any inaccuracies which would cause the model to provide false or unreliable predictions. This can be done by double-checking the data sets for any errors and analyzing the sources of the respective data.

Another critical step in developing a QSAR model is the calculation of molecular descriptors. These descriptors are vectors or scalars that act as surrogates for the actual molecular properties. They can represent different characteristics of the target molecules, such as molecular size, volume, lipophilicity, hydrogen bounding counts, shape, and electronic distribution [22]. Most of the mathematical equations used to develop QSAR models are extremely sensitive to the number of descriptors used, leading to the usage of regression tools, such as partial least squares (PLS) analysis or multiple linear regression (MLR), to better assess the contribution of the variables added to the model.

Most molecular descriptors used in QSAR modelling can be divided into five main classifications [23]:

1. Topological: used for modelling biological, physicochemical, and pharmacokinetic properties through representation of the connectivity of atoms in molecules.
2. Geometrical: calculated from the three-dimensional (3D) coordinates of the atoms in the molecules, it captures the various information regarding different aspects of molecular structures, such as molecular size and shape.
3. Thermodynamic: used to relate chemical structures to observed chemical behavior.
4. Electronic: applied to describe electronic aspects of the molecule or atom bonds;

Table 2 – Summary of available software packages for descriptor calculation. Adapted from: [24].

Name	Type of descriptors	Input format	Software type	Web address
CDK	Topological, constitutional, geometric, electronic, hybrid descriptors	SMILES, SDF, InChl, MOL2, CML, and others	Java Library	http://cdk.github.io
MOE	2D, 3D descriptors, topological, physical properties, structural keys	PDB, SDF, SMILES, and others	Software	www.chemcomp.com
ChemDes	Molecular descriptors	SMILES, SDF, SMI-file	Web-Based Platform	www.scbdd.com/chemdes
CODESSA	Constitutional, topological, geometrical, charge-related, semi-empirical, thermodynamical	n.a.	Software	www.codessa-pro.com
DRAGON	Constitutional, topological 2d-autocorrelations, geometrical, WHIM, GETAWAY, RDF, functional groups, etc.	MDL, Sybyl, SMILES, CML, and others	Software	www.talete.mi.it
PADEL	1D, 2D, 3D descriptors, molecular fingerprints	n.a.	n.a.	www.padel.nus.edu.sg
PreADMET	Constitutional, topological, geometrical, physicochemical, etc.	n.a.	Webserver	http://preadmet.bmdrc.org
RDKit	2D, 3D descriptors, fingerprints	SDF, MOL2, PDB, and others	Toolkit	https://www.rdkit.org

5. Constitutional: simple and common descriptors that reflect the chemical information of a molecule without any additional information on atom connectivity.

A wide range of computer software (**Table 2**) is available online to calculate molecular descriptors, such as PADEL, MOE, DRAGON, and CODESSA, facilitating the process of defining descriptors for use in QSAR modelling.

2. Material and methods

2.1. Visualization & Preparation of the Target Structures

For the application of molecular modelling procedures, it is necessary to choose, obtain and modify the 3D structures of the target proteins that are going to be tested. The selection of the MMP-1 3D structure was realized through analysis of the database "Protein Data Bank" (https://www.rcsb.org/). The available MMP-1 structures were analyzed regarding their resolution, chains, mutations, ligands, and the presence/absence of inhibitors.

The following characteristics were favored: *Homo sapiens* origin, no mutations, obtained from X-Ray crystallography, good resolution (below or equal to 2.40 Å), existence of ligands and the availability of its respective experimental inhibition value. The most promising 4 MMP-1 structures were downloaded from the database in the ".pdb" format.

2.2. Re-docking of co-crystallized ligands against MMP-1 3D structures

The choice of the best 3D structure is an essential step in virtual screening studies, significantly influencing the process and the obtained results. One of the most used methods to discern this is the usage of re-docking experiments, allowing the user to establish which available 3D structure has the best potential for usage within the scope of a virtual screening experiment [25]. This co-crystalized inhibitor present in the 3D structure is separated from its structure and then used against that same structure in a docking experiment, allowing the evaluation of how efficient the used software is.

The following figure (**Figure 5)** shows each necessary step for these methodologies, using the platform that was chosen for this work: SAMSON and its in-built extension of AutoDock Vina.

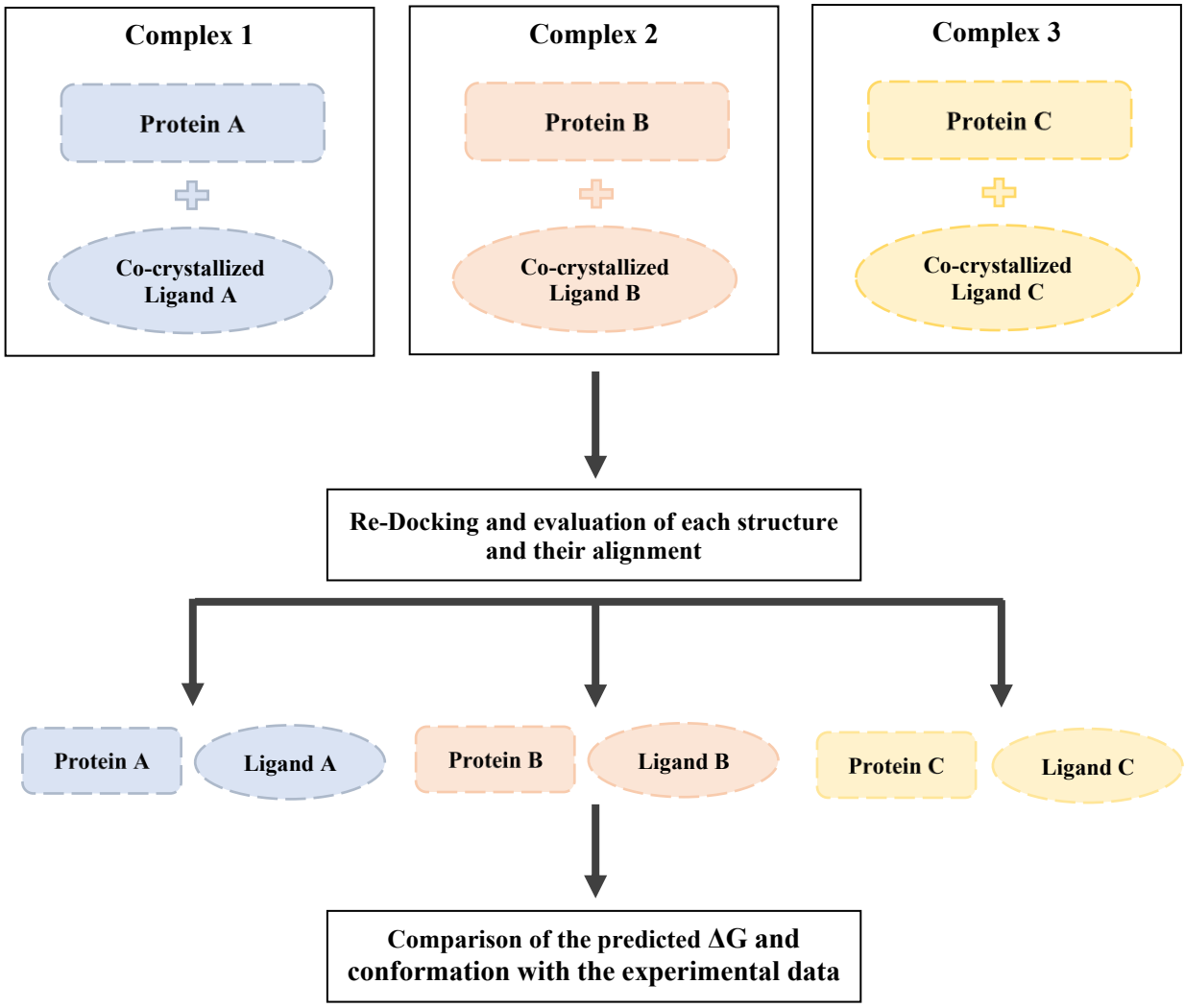


Figure 5 - Re-Docking methodology step-by-step breakdown.

2.3. MMP-1 Ligand library construction

The MMP-1 ligand library was obtained from the platform "Collection of Open Natural Products" (COCONUT) 2.0. COCONUT 2.0 is a comprehensive and vast dataset of natural products, all sourced from open databases, offering an easy-to-use, user-friendly web-based interface that facilitates the searching and downloading of natural product compounds. Since the scope of this experiment was aimed at finding promising natural compounds with the capacity of inhibiting MMP-1, this database was ideal for this purpose.

2.4. Molecular Docking

2.4.1. Preparation of the target receptor, ligands and docking settings

The molecular docking of the developed library was conducted by the Autodock Vina extension within the SAMSON platform. SAMSON (*Software for Adaptive Modeling and Simulation Of Nanosystems)* is a software platform developed by OneAngstrom for integrated molecular design. It's an open platform that can be used in a variety of fields including drug design, chemistry, nanoscience and physics. Due to its open architecture and the provided Software Development Kit (SDK), anyone can develop apps and services to use within the platform. Essentially, it can integrate a variety of computational methods all into one unified environment that will facilitate and speed up the user workflow. A lot of these apps can be obtained from the website SAMSON Connect, in the form of extensions that can be installed onto the base program.

One of these extensions is an integration of the commonly used Autodock Vina, a program for molecular docking and virtual screening that can quickly predict the binding of small molecules to proteins and grade each obtained conformation according to a scoring function [26], [27]. The version used for this work was version 1.2.7.

This module was used not only for the molecular docking procedure itself but also for the preparation of the receptor and ligand structures:

1. The structure was fetched from the PDB database using the built-in function within SAMSON. After loading the chosen structure within the software, its 3D structure was analyzed to make sure there weren't any discrepancies within its composition. Additionally, the ligand binding site was identified and its coordinates saved, with the goal of using it to set up the docking procedure later. The structure was then optimized for docking purposes by using the "Prepare system" option provided within the SAMSON platform. This feature modifies the target receptor by removing unnecessary components within the structure such as alternate locations, ligands, waters and monatomic ions and, finally, adding hydrogens.
2. Using the coordinates obtained previously, a simulation cell/grid was established in SAMSON, encapsulating the receptor's binding site and surrounding areas. This will allow the experiment to solely focus on that defined area, increasing the docking tool's efficiency. The grid's center was

placed on the active site of the structure with a size of 20 Å in terms of width, length and depth.

3. The preparation of ligands was also realized within SAMSON's Autodock Vina extension. This includes minimization experiments, the addition of missing hydrogens and the conversion of each ligand into ".pdbqt" files, which will then serve as the main ligand input for Vina.
4. Finally, the extension also provides a built-in UI where the user can further customize the docking experiment, ranging from the used scoring function, the exhaustiveness of the search, the maximum number of generated binding modes and the type of verbosity used. The output of results can also be customized, allowing the user to just obtain the ligand scores and/or the protein-ligand 3D conformations.

2.4.2. Initialization of the docking procedure using SAMSON

With the receptor and ligands loaded and the parameters set, the docking procedure was initiated using SAMSON's Vina extension. Just like the standalone Vina, the process can be accompanied by opening the ".log" file that contains various details from the docking procedure, such as the progress of the current ligand being docked or the affinities of each test conformation of any of the finished ligands.

Due to the large size of the tested library and the limited specifications of the hardware used, the procedure can take long periods of time.

2.4.3. Scoring & Ranking of the Obtained Ligand Poses

The results were loaded onto the AutoDock Vina extension, providing a table with the affinity of every obtained ligand mode of binding. This table was then exported in the ".csv" format to couple the obtained results with the QSAR model ones that were obtained afterwards. The data was then treated, analyzed and discussed accordingly.

2.5. QSAR Model development

2.5.1. Library Division: Training & Test Sets

To develop a QSAR model, it's necessary to have a training library with compounds that have been subjected to *in vitro* enzymatic assays against the target receptor. Additionally, the IC_{50} value for each of these compounds needs to be present since this value and the respective SMILE are the main inputs behind the development process of a QSAR model. A library of 83 compounds fitting these criteria was prepared and is available in its entirety on the **Appendix A**. These compounds were obtained from

PHAROS, a user interface to the Knowledge Management Center (KMC) that contains a comprehensive database of chemical compounds, their effects on different target proteins including the IC_{50} or Ki value and the research article from which these values hail from.

2.5.2. Molecular Descriptors Calculation

As stated previously, the calculation of molecular descriptors is one of the main steps of building a QSAR model. The proper choice of descriptors is essential to build a robust predictive model, meaning that it's extremely important to choose a reliable software or platform for such purpose [28]. For this work, the OCHEM platform was chosen for its availability and efficacy that was showcased in numerous other studies.

OCHEM (Online Chemical Modelling Environment) is a web-based platform that automates and simplifies the operations necessary to construct a QSAR model. It's based on a wiki principle and focuses mainly on the quality and verification of the available data [29]. The database is integrated with the modelling structure and supports all the steps necessary to build a predictive model:

- Data search.
- Calculation and selection of molecular descriptors.
- Application of various machine learning methods.
- Validation and analysis of models.
- Applicability domain assessment.

Despite being capable of these features, it is essential to note that for this work, OCHEM was only used for calculating the molecular descriptors of the provided compounds, and other methodologies were used to perform the model itself.

2.5.3. CrescentSilico's ChemMaster

ChemMaster is a drug design software developed by CrescentSilico, with a focus on QSAR modelling and predictive machine learning. It offers an extensive toolkit that covers every main step of the QSAR workflow, from dataset preparation and descriptors computation to model evaluation, application (virtual screening) and management.

In this experiment, this software was used as the main platform for the QSAR model development with most of the main steps being executed using its functionalities.

2.5.4. Data Input and Molecular Descriptors Selection

ChemMaster utilizes the experimental values associated with the provided molecules and the molecular descriptors generated by the descriptor calculation tool as the main input. In addition to generating a set of basic and commonly used molecular

descriptors, it is also possible to import descriptors from external sources (in ".csv" format), such as those previously developed on the OCHEM platform. These descriptors were calculated and loaded together with the previously prepared training set.

2.5.5. QSAR Model Validation and Visualization

Beyond building and preparing QSAR models, ChemMaster also provides tools for validation and result viewing. One such validation method is the cross-validation leave-one-out (CV-LOO).

Cross-validation is a technique that relies on data splitting to evaluate the predictive performance of statistical models. It splits the available data into a training and test set, fitting the developed model to the training data and evaluating it by predicting the test data. By repeating this process, the method can estimate the model's average predictive performance [30]. Cross-validation can be used with a variety of different data-splitting schemes. In this work, the CV-LOO method is used, in which a single observation is used for validation, while the remaining data form the training set. This process is repeated further, such that each observation in the whole dataset is used only once for validation. This way, LOOCV provides nearly unbiased error estimates, but it is time-consuming, especially when applied to large datasets, and may lead to high variance [31]. This validation will check the model's accuracy in predicting the desired biological activity of the tested substances and offers statistical evidence to support that claim.

3. Results and Discussion

3.1. QSAR Model Selection

For this dissertation, various QSAR models were developed using the previously gathered and described compound library. These models were prepared using a variety of parameters in ChemMaster, including different machine learning methods, and sets of molecular descriptors (**Table 3**).

Table 3 – Statistical results of 14 developed QSAR models. R^2 – determination coefficient. RMSE – square root of the mean error. R^2CV – determination coefficient after use of the cross-validation method. Q^2CV – average value outputted by the machine learning method after the use of the cross-validation method. N.D. Not determined.

Model Descriptors	Model Number	Machine Learning Method	Training Set				Test Set		
			R^2	RMSE	R^2CV	Q^2CV	R^2	RMSE	R^2CV
ChemMaster Descriptors	1	MLR	0.44	0.717	0.597	0.238	0.431	0.729	0.619
	2	KNN (k=5)	0.96	0.191	0.098	0.264	0.492	0.689	0.56
	3	KNN (k=8)	0.641	0.574	0.386	0.243	0.501	0.682	0.531
	4	PLS (n=1)	0.367	0.762	0.63	0.248	0.465	0.707	0.554
	5	SVR	0.353	0.77	0.521	0.143	0.394	0.753	0.603
	6	KNN (k=3)	0.964	0.183	0.096	0.253	0.221	0.853	0.71
CDK Descriptors	7	KNN (k=5)	1	0	0	0.183	0.405	0.745	0.572
	8	PLS (n=1)	0.318	0.791	0.59	0.205	0.382	0.76	0.607
Dragon Descriptors	9	PLS (n=1)	0.522	0.662	0.48	0.462	0.577	0.629	0.527
	10	KNN (k=3)	0.99	0.097	0.027	0.456	N.D.	0.995	0.805
RDKIT Descriptors	11	PLS (n=1)	0.591	0.612	0.478	0.526	0.413	0.741	0.593
CDK + Dragon + RDKIT Descriptors	12	PLS (n=1)	0.479	0.691	0.502	0.377	0.292	0.813	0.618
	13	KNN (k=5)	0.844	0.075	0.022	0.632	N.D.	0.998	0.784
	14	MLR	0.481	0.69	0.487	0.379	0.1	0.917	0.749

The models described above were prepared using different sets of molecular descriptors present in the ChemMaster software and in the OCHEM platform. Essentially, the models were prepared using the software's own descriptors (a total of 29 descriptors), the CDK package from OCHEM (a total of 256 descriptors), the Dragon package (a total of 5270 descriptors), and the RDKIT package (a total of 3579 descriptors). This allowed testing the different descriptors and the software's parameters to optimize the development of a statistically competent model.

Several of these models showed strong statistical validation, with the top-ranked models being 2, 6, 10, and 13. These models all shared the same machine learning

method, kNN – k-Nearest Neighbors regression. The kNN algorithm is a simple machine-learning method that classifies an instance by a majority vote of its neighbors; each test instance is assumed to belong to the class most common amongst its k closest neighbors, with k a positive integer. These neighbors are identified based on distance in the feature space, such as the Euclidean distance [32].

Although these four QSAR models were the best, a decision was made to use only one for the remainder of this work. It was decided to use model 2, since it showed strong statistical performance in both the training set ($R^2 > 0.9$; RMSE < 0.2) and the test set ($R^2 > 0.4$).

3.2. Detailed analysis of QSAR model

Understanding the statistical data provided by the developed QSAR model is an essential step in virtual screening studies. Analyzing these parameters will help determine whether the model is appropriate for predicting the biological activity being targeted – MMP-1 inhibition. For this type of QSAR model, there is no statistical reference or threshold beyond which a QSAR model is considered to have good predictive capacity. However, several QSAR models published in the literature consider a QSAR model to have good predictive power if the determination coefficient (R^2) is greater than 0.750 and the root mean square error (RMSE) is less than 0.300. All these statistical parameters provide useful information about the model's reliability. For QSAR model 2, all statistical parameters relating to the training set (compounds used to build the model) fall within the values indicated: R^2 (0.96), well above 0.750, and RMSE (0.191), well below 0.3.

Additionally, the analysis of the molecular descriptors selected by ChemMaster for this model is significant, as it will provide insights into the favorable characteristics of the compounds used. The descriptors selected for this model were:

1. Minimum Absolute Partial Charge – Reflects the lowest (most neutral) atomic partial charge in the molecule, indicating regions of minimal polarity or electron density variation.
2. Aliphatic COO – Counts or represents the presence of carboxylate (–COO–) groups attached to aliphatic (non-aromatic) carbon atoms, often related to acidity and solubility.
3. Aromatic NH – Identifies nitrogen–hydrogen (–NH) groups directly attached to an aromatic ring, commonly linked to hydrogen bonding and electronic effects in heterocycles.

4. HOCCN – Describes the occurrence of the HO–C–C–N substructure pattern, capturing specific hydrogen-bonding and connectivity features involving hydroxyl, carbon, and nitrogen atoms.
5. Ndealkylation 1 – Refers to the metabolic likelihood or accessibility of an N-dealkylation site, indicating how easily an alkyl group can be removed from nitrogen during metabolism.

In the context of human collagenase inhibition, the molecular descriptors selected in the QSAR model-building reflect key structural and electronic features that influence enzyme binding and activity. The minimum absolute partial charge reflects regions of high polarity within the molecule, which can facilitate hydrogen bonding or coordinate with the catalytic zinc ion in the MMP-1 active site, enhancing binding affinity [33]. The presence of aliphatic carboxyl groups (COO) is noteworthy, as carboxylate or related functional groups (e.g. sulfonamide group (–SO_2NH–)) are known to chelate the catalytic zinc ion in MMPs, anchoring the inhibitor within the active site and contributing to selectivity [34]. Aromatic NH groups can form hydrogen bonds with residues in the S1' pocket of the enzyme, stabilizing the inhibitor-enzyme complex and influencing specificity [35]. The HOCCN motif, comprising hydrogen-bond donors and acceptors, supports favorable interactions within the binding site, potentially enhancing affinity and selectivity. Finally, the N-dealkylation propensity reflects steric and electronic features around nitrogen atoms, which may influence binding geometry and metabolic stability, indirectly affecting the duration of inhibitory activity [36].

Collectively, these descriptors highlight the importance of electronic and steric factors in the binding between the inhibitor and MMP-1, underscoring fundamental characteristics for the development of inhibitors with greater efficacy and stability, crucial for therapeutic applications targeting this enzyme.

3.3. Molecular docking approach validation

To complement the QSAR model results, a molecular docking approach was implemented for assessing collagenase inhibition activity using the SAMSON platform and AutoDock Vina as the docking algorithm. We started by selecting an appropriate 3D structure of collagenase for docking and validated it by performing Redocking.

3.3.1. Selection of the MMP-1 3D structure

This step of the work began with an analysis of the experimental 3D structures of the MMP-1 protein available in the Protein Data Bank (PDB). The initial search returned 15 MMP-1 structures for *Homo sapiens* in the PDB database. From these 15 initial structures, 4 promising ones were selected based on the previously mentioned criteria: no mutations, experimentally obtained by X-ray crystallography, good resolution (< 2.40 Å), presence of ligands, and availability of their respective experimental inhibition values. The selected structures can be seen highlighted in **Table 4**:

Table 4 – Experimental structures of the MMP-1 protein available in PDB.

PDB Code	Resolution	Reference	Mutations	Inhibitor (PubChem CID)	KI (nM)
1HFC	1.50 Å	[37]	0	PLH (194777)	8.15 - log(M)
1CGF	2.10 Å	[38]	0	NO	X
1CGE	1.90 Å	[38]	0	NO	X
1CGL	2.40 Å	[39]	0	0ED (444824)	135 nM
1SU3	2.20 Å	[40]	0	EPE (23831)	X
2CLT	2.67 Å	[41]	1	NO	X
2J0T	2.54 Å	[42]	0	TIMP-1	X
2TCL	2.20 Å	[43]	0	RO4 (5487313)	40 nM
3SHI	2.20 Å	[44]	0	NO	X
4AUO	3.00 Å	[45]	1	NO	X
966C	1.90 Å	[46]	0	RS2 (4369)	23 nM
1AYK	NMR	[47]	0	NO	X
2AYK	NMR	[47]	0	NO	X
3AYK	NMR	[47]	0	CGS (446504)	X
4AYK	NMR	[47]	0	CGS (446504)	X

3.3.2. Re-Docking using AutoDock Vina

After selecting the structures that fit within the criteria, the active site coordinates from each one were obtained with the intention of using them for a re-docking experiment (**Table 5**). A grid was established with 20 Å of width, height and length and centered in the following coordinates.

Table 5 – Selected protein structures and their respective active site coordinates.

PDB Code	Coordinate X	Coordinate Y	Coordinate Z
1HFC	24.454	23.213	24.648
1CGL	28.009	39.685	-1.335
2TCL	74.443	8.149	7.673
966C	9.248	-10.602	36.826

In molecular docking assays, one of the results obtained is the predicted inhibition constant (Ki), which represents the estimated inhibitory value of a given compound against the target protein structure. For the selected MMP-1 structures, experimental inhibition constants (experimental Ki) of the co-crystallized ligands were gathered. Therefore, when performing Re-Docking assays, these experimental Ki values were compared with the obtained predicted Ki values. This is the main reason why PDB structures containing co-crystallized inhibitors with known experimental Ki values were prioritized and selected.

Using this method of comparison, it becomes possible to assess whether these values are comparable and correlated, thereby validating the most suitable 3D protein structures for virtual screening experiments (**Table 6**).

Table 6 – Re-dock experiment results of each selected structure. The predicted ΔG value obtained from re-docking was converted into pKi to compare it with the experimental values available on PDB and associated articles.

Structure/ Ligand	Predicted ΔG (kcal/mol)	Predicted Ki (nM)	Predicted pKi	Experimental Ki	Experimental pKI	pKI Difference
966C (RS2)	-9.4	135.21	6.87	23	7.64	0.77
1CGL (0ED)	-7.5	3368.00	5.47	135	6.87	1.40
1HFC (PLH)	-7.5	2952.55	5.53	6.03	8.22	2.69
2TCL (RO4)	-6.8	9819.71	5.01	40	7.40	2.39

Regarding the comparison between the predicted and experimental values, the structure/ligand with the most promise was 966C/RS2, with RS2 being an acetamide derivative. This compound presented the smallest difference between predicted and experimental values, with a pKi difference of 0.77, indicating a more accurate predicted docking conformation. To further solidify this result, the superimposition of both the experimental and predicted 3D conformations was performed in PyMOL (**Figure 6**). As shown in the figure, the predicted conformation was very similar to the experimental one, occupying the protein's active site in a similar manner.

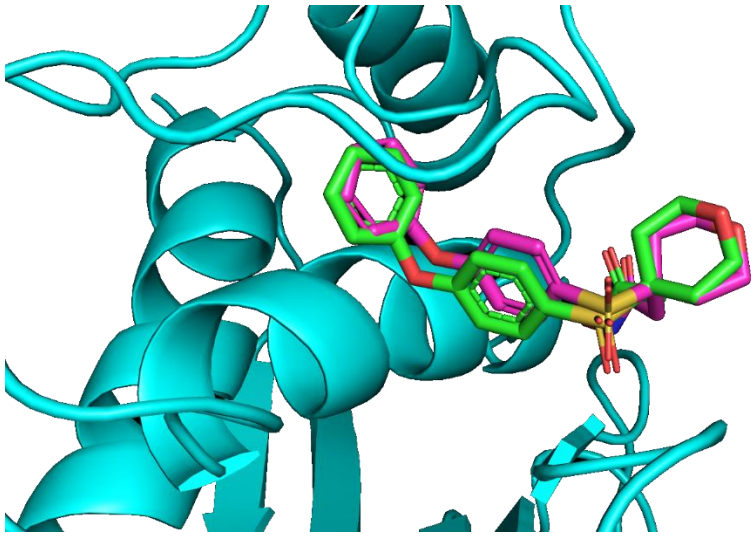


Figure 6 – Superimposition of the experimental and docked conformations for the acetamide derivative (RS2) inhibitor present in 966C PDB structure of MMP-1. Conformations are presented as stick models: docked conformation (green color) and experimental conformation (pink color). Analysis and image preparation was performed using Pymol software.

3.4. Library of natural compounds from the COCONUT database

At this point of the work, the best QSAR model was selected (model 2) and the best PDB structure was also selected using the Re-docking approach. The next step was to select a library of natural compounds to be tested by applying the QSAR model 2 and by performing docking using the selected PDB structure. This library was chosen using the COCONUT platform and its scaffold search tools.

The COCONUT database provides multiple built-in tools for searching for structures related to a user-specified scaffold. For instance, users can perform similarity searches by defining a percentage of structural resemblance (ranging from 1% to 100%) or by selecting a specific scaffold to locate compounds exhibiting varying degrees of similarity. Additionally, the platform supports scaffold-based searches using various search algorithms.

For this work, it was decided to use compounds that closely resembled the ones used in the training set to develop the QSAR models. Most of the training set exhibited the scaffold depicted in **Figure 7**. This scaffold was introduced in the SMILES chemical format, "CS(c1ccccc1)(=O)=O", and utilized in scaffold-based searching of the COCONUT database, resulting in the development of a library of 715 natural compounds. This library will be fully available in the supplementary material.

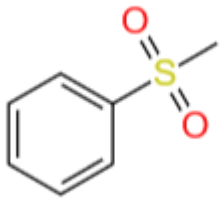


Figure 7 – Structure used as a scaffold for searching the COCONUT database

The full library with the compound names, COCONUT IDs and chemical structures in SMILES format can be accessed in this link.

3.5. Virtual screening of the library of natural compounds using a QSAR/Docking combined approach

The selected library of natural compounds was then virtually screened using a combined QSAR/Docking approach, where the MMP-1 inhibition activity of the library of 715 compounds was predicted by applying the selected QSAR model and by docking

all compounds against MMP-1, using the selected PDB structure 966C. The results from the 2 approaches were combined, as using both is expected to yield more reliable MMP-1 inhibition predictions. The complete predicted results for the 715 compounds from the library can be accessed using this link, the results of the 10 top-ranked compounds are presented in **Table 7**.

The selected QSAR model (model 2) was used to predict the MMP-1 inhibition activity of the library. The model predicted pIC_{50} values ranging from 10.39 to 6.16. Since the values obtained from these methods have significantly different ranges, they were fitted to a 0.0 to 1.0 range (Fitted Predicted pIC_{50} values presented in **Table 7**). For QSAR results, the lowest value of 6.16 was assigned a 0 value, and the highest value of 10.39 was assigned a 1 value.

After identifying the optimal 3D structure, molecular docking experiments were subsequently performed using the 715 compounds from the natural compound library. The same AutoDock Vina docking parameters and PDB: 966C structure, validated using the Re-docking approach, were used, and the process was automated using the AutoDock Vina extension in the SAMSON software. The primary objective was to predict the inhibitory potential of each natural compound and estimate its 3D binding conformation.

Table 7 – Virtual screening results of the selected library of the Coconut database: best 10 compounds, sorted by the inhibition probability, calculated by the average between the affinity (molecular docking) and the predicted pIC_{50} (QSAR model).

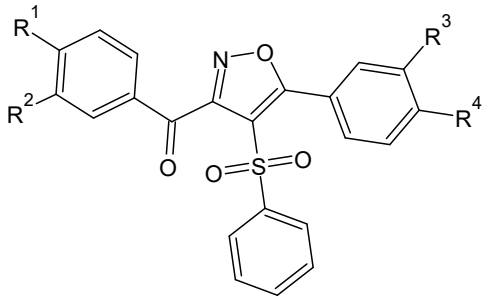
Top-Ranked	Coconut ID	Docking		QSAR		Combined Inhibition Probability (%)
		Affinity (kcal/mol)	Fitted Affinity	Predicted pIC_{50}	Fitted Predicted pIC_{50}	
1	CNP0587393.0	-10.483	0.821	10.390	1.000	86.5
2	CNP0059736.0	-11.686	1.000	6.934	0.183	79.6
3	CNP0521351.0	-9.755	0.712	10.390	1.000	78.4
4	CNP0591040.0	-9.570	0.685	10.390	1.000	76.3
5	CNP0585640.0	-9.567	0.684	10.390	1.000	76.3
6	CNP0522565.1	-11.142	0.919	7.350	0.281	76.0
7	CNP0497659.0	-9.457	0.668	10.390	1.000	75.1
8	CNP0048634.1	-11.091	0.911	7.295	0.268	75.05
9	CNP0585231.0	-9.364	0.654	10.390	1.000	74.0
10	CNP0564068.0	-9.345	0.651	10.390	1.000	73.8

Regarding the Docking predictions, since the values obtained have significantly different ranges, they were also fitted for a 0.0 to 1.0 range. The lowest value of -4.979 was given the 0 value, and the highest value of -11.686 was given the 1 value (Fitted Affinity presented in **Table 7.**

The compound ranking is sorted by the combined inhibition probability (%) obtained from the average fitted values obtained from both docking and QSAR. The average between the two methods was established with a weight of 75% (Docking) and 25% (QSAR). Initially, we calculated the combined inhibition probability (%) using a 50% (Docking) and 50% (QSAR) contribution; however, this 50%/50% ratio favored compounds that presented very similar structures, mainly because the QSAR model selected a specific chemical scaffold, resulting in very similar compounds on the 10 top-ranked compounds. This way, when the split was 50/50 between the methods, the top results were all extremely similar in terms of chemical structure. To have a more diverse selection of chemical compounds, we tested the 75% (Docking) and 25% (QSAR) ratio, and this led to a selection of more structurally diverse compounds in the 10 top-ranked compounds, with different scaffolds. In the next section the binding conformation of the 10 top-ranked compounds, according to the % (Docking) and 25% (QSAR) combined ratio, will be presented and analyzed.

3.6. Binding conformation analysis of the top-ranked compounds

In order to investigate the results of protein-ligand interactions in greater detail, the chemical structures (**Figure 8**) and predicted binding conformations (**Figures 9 & Figure 10**) of the 10 top-ranked compounds with the highest potential were examined.



1. R_1 = Phenyl; R_2 = Phenyl; R_3 = H; R_4 = NO_2, **2.** R_1 = Phenyl; R_2 = Phenyl; R_3 = Phenyl; R_4 = Phenyl, **4.** R_1 = NO_2; R_2 = H; R_3 = H; R_4 = NO_2, **5.** R_1 = NO_2; R_2 = H; R_3 = H; R_4 = CH_3, **9.** R_1 = NO_2 ; R_2 = H; R_3 = H; R_4 = CH_3, **10.** R_1 = CH_3; R_2 = H; R_3 = H; R_4 = NO_2,

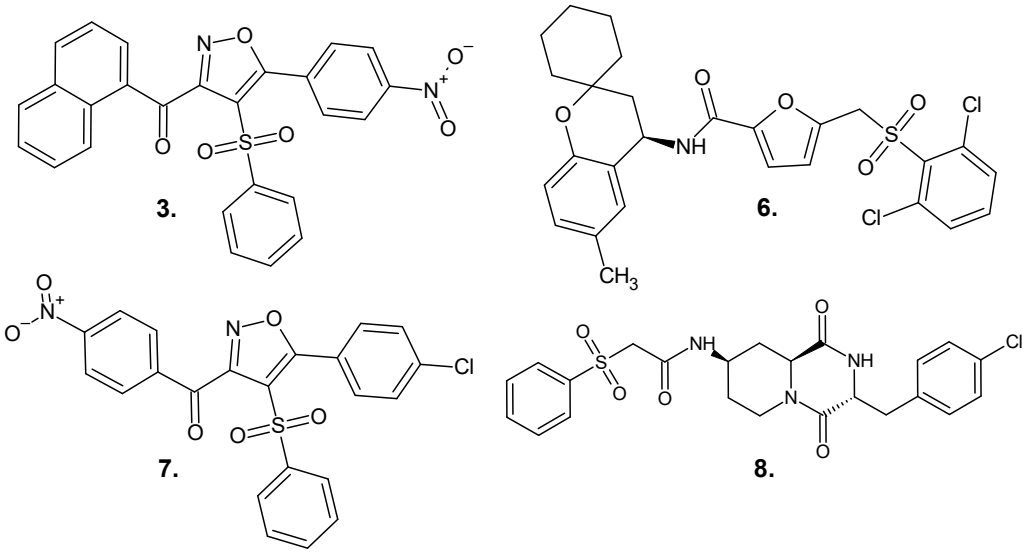


Figure 8 - Chemical 2D structures of the top 10 ranked compounds. Image prepared using Chemsketch.

Regarding the 3D ligand conformations, the key interactions between each of the top-ranked compounds and the protein structure are highlighted and the positioning of each compound within the active site of the protein is illustrated. In addition, the hydrogen bonds predicted by AutoDock Vina between the ligands and the MMP-1 structure, as well as the amino acid residues involved in these hydrogen bonds, are also represented.

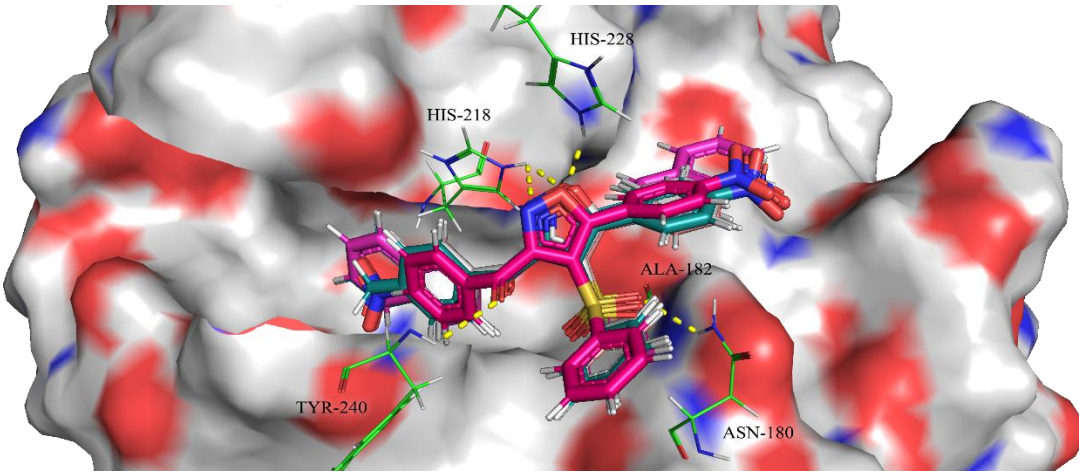


Figure 9 - Superimposition of the docked conformations against MMP-1 of the 6 top-ranked compounds, with similar scaffolds from the Docking/QSAR combined approach (1,2,4,5,9,10). Compounds are presented as sticks in various colors (1-cyan, 2-purple, 4-salmon, 5-grey, 9-blue and 10-pink). The 3D surface represents the MMP1 structure (PDB:966c). The H-bonds are shown in dotted yellow lines. Analysis and image preparation performed using Pymol software.

Although none of the top 10 compounds have been reported in the literature as explicit inhibitors of MMP-1, their structural features align well with molecular design principles for metalloproteinase inhibition. Many reported MMP inhibitors possess a zinc-binding moiety, typically a carboxylate or hydroxamate, which coordinates the catalytic Zn^{2+} in the MMP active site and thereby disrupts substrate hydrolysis [48]. For example, sulfonate (-SO_2-) and related functional groups are well-established as effective zinc-binding moieties in the catalytic site of MMPs, and compounds bearing such groups have been shown to inhibit MMP-1 in the micromolar range [49].

In parallel, the architecture of the MMP-1 active site is known to be comparatively shallow and constrained in size (for instance, due to the presence of Arg214/Tyr214) compared to other MMPs, making the shape and substituents of the ligand particularly relevant to affinity and selectivity. In our top 10 compounds, the presence of substituted aromatic rings (e.g., phenyl, nitrophenyl) and specially 1,2-oxazole moieties (compounds 1-5,7,9,10) suggest good potential for hydrophobic/aromatic occupancy of the main pocket of the catalytic site and adjacent subsites, while polar groups promote hydrogen bonding or dipole interactions with active-site residues or backbone carbonyls [50]. Moreover, descriptor terms such as the “minimum absolute partial charge”, that was one of the chosen descriptors in the model, correspond to the need for well-placed polar/charged regions that can contribute to coordinate or electrostatic interaction with the catalytic Zn^{2+} or surrounding residues [51].

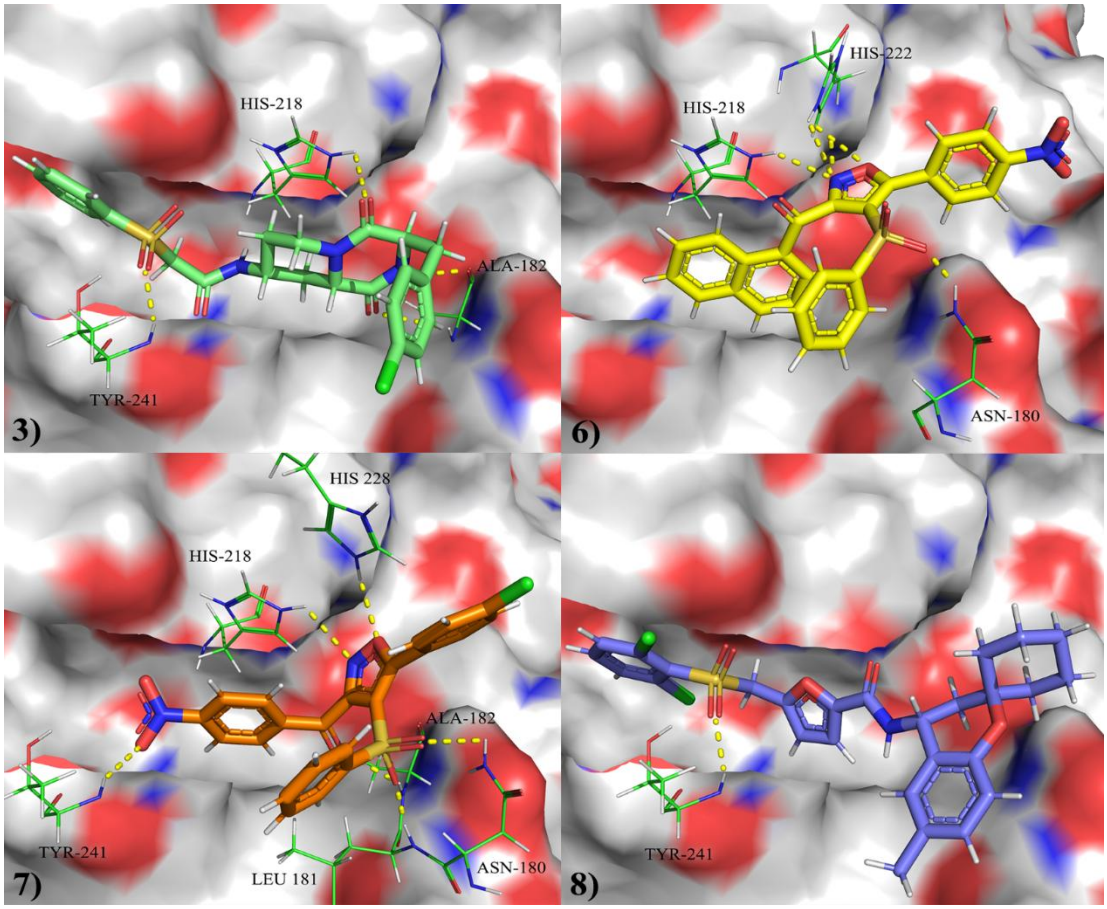


Figure 10 - Docked conformations of the other 4 top-ranked compounds against MMP-1, with alternative scaffolds. Compounds are presented as sticks in various colors (3-green, 6-yellow, 7-orange, 8-blue. The H-bonds are shown in dotted yellow lines. Analysis and image preparation performed using Pymol software.

Finally, although metabolic-related descriptors such as "N-dealkylation propensity" may not directly influence binding, they reflect steric and electronic features around nitrogen atoms that can impact binding geometry and perhaps the lifetime of active species [52].

Taken together, while these compounds were not experimentally tested against MMP-1, they display the core structural motifs (zinc-chelation, hydrophobic/aromatic pocket occupancy, hydrogen-bonding donor/acceptor groups) identified in successful MMP inhibitor design, which supports the plausibility of their predicted inhibitory activity.

4. Conclusion

A library of 83 compounds with MMP-1 inhibitory activity was prepared based on a database and literature search. From these compounds, it was possible to develop QSAR models to predict MMP-1 activity. After analyzing all the QSAR models' statistical data, QSAR model 2 was selected for further use. Additionally, a library of 715 natural products with a similar structure to the compounds used in the QSAR model's training set was prepared using search tools from the COCONUT database of natural compounds. The molecular docking approach was validated by performing a Re-Docking protocol and by performing a structural analysis of the MMP-1 structures from PDB.

Through the application of the developed QSAR model and molecular docking methodologies in the natural compounds library, it was possible to predict the MMP-1 inhibitory activity of these compounds. The compounds were sorted by their combined inhibition probability, calculated as the average of the results from both *in silico* methodologies. To have a more diverse selection of chemical compounds, the 75% (Docking) and 25% (QSAR) ratio for the combined average was tested, and this led to a selection of more structurally diverse compounds in the 10 top-ranked compounds, with different scaffolds. These were analyzed further in terms of their structure and binding conformations.

Although the 10 top-ranked compounds do not have studies that prove their inhibition and are new in the context of MMP-1 inhibition, they satisfy the main structural motifs (zinc chelation + H-bond donor/acceptor + aromatic/hydrophobic substituents) typical of inhibitors of other metalloproteinases, supporting the plausibility of their predicted activity and validating the robustness and accuracy of the QSAR model and Docking studies performed.

In the future, it would be of interest to verify which of these promising compounds are available for acquisition, and if so, buy them to test them experimentally against MMP-1. If these compounds show similar experimental inhibition to the predicted inhibition, it will thus validate the use of *in silico* tools as a predictive tool. Finally, these compounds could be considered for potential use in cosmeceutical applications regarding skin photoaging.

5. Bibliographic References

[1] J. C. Chamcheu, A. L. Walker, and F. K. Noubissi (2021) "Natural and synthetic bioactives for skin health, disease and management" *Nutrients*, 13, doi: 10.3390/nu13124383.

[2] M. A. Farage, K. W. Miller, P. Elsner, and H. I. Maibach (2008) "Intrinsic and extrinsic factors in skin ageing: A review" *Int J Cosmet Sci*, 30(2): 87-95, doi: 10.1111/j.1468-2494.2007.00415.x.

[3] K. Madan and S. Nanda (2018) "In-vitro evaluation of antioxidant, anti-elastase, anti-collagenase, anti-hyaluronidase activities of safranal and determination of its sun protection factor in skin photoaging" *Bioorg Chem*, 77: 159–167, doi: 10.1016/j.bioorg.2017.12.030.

[4] I. Chiocchio, M. Mandrone, C. Sanna, A. Maxia, M. Tacchini, and F. Poli (2018) "Screening of a hundred plant extracts as tyrosinase and elastase inhibitors, two enzymatic targets of cosmetic interest" *Ind Crops Prod*, 122: 498–505, doi: 10.1016/j.indcrop.2018.06.029.

[5] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, and C. T. Supuran (2021) "Natural products in drug discovery: advances and opportunities" *Nature Research*, 20:200-216, doi: 10.1038/s41573-020-00114-z.

[6] K. Dzobo (2022) "The Role of Natural Products as Sources of Therapeutic Agents for Innovative Drug Discovery" in *Comprehensive Pharmacology,* Elsevier Edition, 2: 408-422, doi: 10.1016/B978-0-12-820472-6.00041-4.

[7] W. Bode and K. Maskos (2003) "Structural basis of the matrix metalloproteinases and their physiological inhibitors, the tissue inhibitors of metalloproteinases" *Biol Chem*, 84(6): 863–72, doi: 10.1515/BC.2003.097.

[8] T. Quan, Z. Qin, W. Xia, Y. Shao, J. J. Voorhees, and G. J. Fisher (2009) "Matrix-degrading metalloproteinases in photoaging" *J Investig Dermatol Symp* Proc, 14(1):20–24, doi: 10.1038/jidsymp.2009.8.

[9] S. Wu, X. Zhou, Z. Jin, and H. Cheng (2023) "Collagenases and their inhibitors: a review" *Collagen Leather*, 5: no. 19, doi: 10.1186/s42825-023-00126-6.

[10] J. C. Spurlino, A. M. Smallwood, D. D. Carlton, T. M. Banks, and K. J. Vavra (1994) "1.56 Å structure of mature truncated human fibroblast collagenase" *Proteins:*

Structure, Function, and Bioinformatics, 19:98–109, doi: 10.1002/prot.340190203.

[11] A. Pardo and M. Selman (2005) "MMP-1: The elder of the family" *Int J Biochem Cell Biol*, 37(2):283-288, doi: 10.1016/j.biocel.2004.06.017.

[12] T. S. A. Thring, P. Hili, and D. P. Naughton (2009) "Anti-collagenase, anti-elastase and anti-oxidant activities of extracts from 21 plants" *BMC Complement Altern Med*, 9: no. 27, doi: 10.1186/1472-6882-9-27.

[13] J. L. Lauer-Fields, D. Juska, and G. B. Fields (2002) "Matrix metalloproteinases and collagen catabolism", *Pept. Sci.,* 66:19-32, doi: 10.1002/bip.10201.

[14] S. Jadoon, S. Karim, M. Asad, M. R. Akram, A. K. Khan, A. Malik, C. Chen and G. Murtaza (2015) "Anti-Aging Potential of Phytoextract Loaded-Pharmaceutical Creams for Human Skin Cell Longetivity", *Oxid Med Cell Longev, doi*: 10.1155/2015/709628.

[15] B. Shaker, S. Ahmad, J. Lee, C. Jung, and D. Na (2021) "In silico methods and tools for drug discovery" *Comput Biol,* 137, doi: 10.1016/j.compbiomed.2021.104851.

[16] F. Stanzione, I. Giangreco, and J. C. Cole (2021) "Use of molecular docking computational tools in drug discovery", *Prog Med Chem*, 60:273–343, doi: 10.1016/bs.pmch.2021.01.004.

[17] C. Zardecki, S. Dutta, D. S. Goodsell, R. Lowe, M. Voigt, and S. K. Burley (2022) "PDB -101: Educational resources supporting molecular explorations through biology and medicine" *Protein Science*, 31:129–140, doi: 10.1002/pro.4200.

[18] P. C. Agu, C. A. Afiukwa, O. Orji, and E. M. Ezeh (2023) "Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management" *Sci Rep*, 13, doi: 10.1038/s41598-023-40160-2.

[19] N. T. Issa, E. V. Badiavas, and S. Schürer (2019) "Research Techniques Made Simple: Molecular Docking in Dermatology - A Foray into In Silico Drug Discovery" *J Invest Dermatol*, 139:2400-2408, doi: 10.1016/j.jid.2019.06.129.

[20] K. Roy, S. Kar, R. Narayan Das, and N. Das Rudra (2015) "A Primer on QSAR/QSPR Modeling - Fundamental Concepts", Springer Edition, doi: 10.1007/979-3-319-17281-1.

[21] S. Kwon, H. Bae, J. Jo, and S. Yoon (2019) “Comprehensive ensemble in QSAR prediction for drug discovery” *BMC Bioinformatics*, 20:521 doi: 10.1186/s12859-019-3135-4.

[22] A. M. Davis (2017) “Quantitative Structure-Activity Relationships” in *Comprehensive Medicinal Chemistry*, Elsevier Edition, 3:379–392, doi: 10.1016/B978-0-12-409547-2.12348-0.

[23] Danishuddin and A. U. Khan (2016) “Descriptors and their selection methods in QSAR analysis: paradigm for drug design” *Drug Discov Today*, 60:273-343, doi: 10.1016/j.drudis.2016.06.013.

[24] J. Emonts and J. F. Buyel (2023) “An overview of descriptors to capture protein properties – Tools and perspectives in the context of QSAR modeling” *Comput Struct Biotechnol J*, 21:3234-3247, doi: 10.1016/j.csbj.2023.05.022.

[25] J. Shamsara (2016) “CrossDocker: a tool for performing cross-docking using Autodock Vina” *Springerplus*, 5: no. 1, doi: 10.1186/s40064-016-1972-4.

[26] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli (2021) “AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings” *J Chem Inf Model*, 61:3891–3898, doi: 10.1021/acs.jcim.1c00203.

[27] O. Trott and A. J. Olson (2010) “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading” *J Comput Chem*, 31:455–461, doi: 10.1002/jcc.21334.

[28] S. S. Young, F. Yuan, and M. Zhu (2012) “Chemical Descriptors Are More Important Than Learning Algorithms for Modelling” *Mol Inform*, 31:707–10, doi: 10.1002/minf.201200031.

[29] I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko and I. V. Tetko (2011) “Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information” *J Comput Aided Mol Des*, 25:533–554, doi: 10.1007/s10822-011-9440-2.

[30] L. A. Yates, Z. Aandahl, S. A. Richards, and B. W. Brook (2023) “Cross validation for model selection: A review with examples from ecology” *Ecol Monogr,* 93(1), doi: 10.1002/ecm.1557.

[31] V. Lumumba, D. Kiprotich, M. Mpaine, N. Makena, and M. Kavita (2024) “Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models” *Am J Theor Appl Bus*, 13:127–137, doi: 10.11648/j.ajtas.20241305.13.

[32] J. B. O. Mitchell (2014) “Machine learning methods in chemoinformatics” *Wiley Interdiscip Rev Comput Mol Sci.,* 4(5):468-481, doi: 10.1002/wcms.1183.

[33] M. Fernández and J. Caballero (2007) “QSAR modeling of matrix metalloproteinase inhibition by N-hydroxy-α-phenylsulfonylacetamide derivatives” *Bioorg Med Chem*, 15:6298–6310, doi: 10.1016/j.bmc.2007.06.014.

[34] K. Roy, D. K. Pal, A. U. De, and C. Sengupta (2001) “QSAR of matrix metalloproteinase inhibitor N-[(substituted phenyl)sulfonyl]-N-4-nitrobenzylglycine hydroxamates using LFER model” *Drug Des Discov*, 17:315–323.

[35] E. B. de Melo (2012) “A QSAR Study of Matrix Metalloproteinases Type 2 (MMP-2) Inhibitors with Cinnamoyl Pyrrolidine Derivatives” *Sci Pharm*, 80:265–81, doi: 10.3797/scipharm.1112-21.

[36] B. K. Sharma, P. Singh and Y. S. Prabhakar (2013) “QSAR Rationale of Matrix Metalloproteinase Inhibition Activity in a Class of Carboxylic Acid Based Compounds” *Br J Pharm Res*, 3:697–721, doi: 10.9734/BJPR/2013/3903.

[37] J. C. Spurlino, A.M. Smallwood, D. D. Carlton, T. M. Banks, K. J. Vavra, J. S. Johnson, E. R. Cook, J. Falvo, R. C. Wahl, T. A. Pulvino, J. J. Wendoloski, and D. L. Smith (1994) “A Structure of Mature Truncated Human Fibroblast Collagenase” *Proteins*, 19:98-109, doi: 10.1002/prot.340190203.

[38] B. Lovejoy, A. M. Hassell, M. A. Luther, D. Weigl, and S. R. Jordan (1994) “Crystal Structures of Recombinant 19-kDa Human Fibroblast Collagenase Complexed to Itself3” *Biochemistry*, 33(27):8207-8217, doi: 10.1021/bi00193a006.

[39] B. Lovejoy, A. Cleasby, A. M. Hassell, K. Longley, M. A. Luther, D. Weigl, G. McGeehan, A. B. McElroy, D. Drewry, M. H. Lambert and S. R. Jordan (1994)

"Structure of the Catalytic Domain of Fibroblast Collagenase Complexed with an Inhibitor" *Science*, 263(5145):375-377, doi: 10.1126/science.8278810.

[40] D. Jozic, G. Bourenkov, N. Lim, R. Visse, H. Nagase, W. Bode and K. Maskos (2005) "X-ray structure of human proMMP-1: New insights into procollagenase activation and collagen binding" *J Biol Chem*, 280(10):9578–9585, doi: 10.1074/jbc.M411084200.

[41] S. Iyer, R. Visse, H. Nagase, and K. R. Acharya (2006) "Crystal Structure of an Active Form of Human MMP-1" *J Mol Biol*, 362(1):78–88, doi: 10.1016/j.jmb.2006.06.079.

[42] S. Iyer, S. Wei, K. Brew, and K. R. Acharya (2007) "Crystal structure of the catalytic domain of matrix metalloproteinase-1 in complex with the inhibitory domain of tissue inhibitor of metalloproteinase-1" *J Biol Chem*, 282:364–371, doi: 10.1074/jbc.M607625200.

[43] N. Borkakoti, F. K. Winkler, D. H. Williams, A. D'Arcy, M. J. Broadhurst, R. A. Brown, W. H. Johnson, and E. J. Murray (1994) "Structure of the catalytic domain of human fibroblast collagenase complexed with an inhibitor" *Nat Struct Biol*, 1(2): 106–110, doi: 10.1038/nsb0294-106.

[44] I. Bertini, V. Calderone, L. Cerofolini, M. Fragai, C. F. G. C. Geraldes, P. Hermann, C. Luchinat, G. Parigi, and J. M. C. Teixeira (2012) "The catalytic domain of MMP-1 studied through tagged lanthanides" *FEBS Open Bio*, 586(5):557-567, doi: 10.1016/j.febslet.2011.09.020.

[45] S. W. Manka, F. Carafoli, R. Visse, D. Bihan, N. Raynal, R. Farndale, G. Murphy, J. J. Enghild, E. Hohenester, and H. Nagase (2012) "Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1" *Proc Natl Acad Sci U S A*, 109(31):12461–12466, doi: 10.1073/pnas.1204991109.

[46] B. Lovejoy, A. R. Welch, S. Carr, C. Luong, C. Broka, R. T. Hendricks, J. A. Campbell, K. A. Walker, R. Martin, H. V. Wart, and M. F. Browner (1999) "Crystal structures of MMP-1 and-13 reveal the structural basis for selectivity of collagenase inhibitors" *Nat Struct Biol*, 6(3):217-221, doi: 10.1038/6657.

[47] F. J. Moy, P. K. Chanda, S. Cosmi, M. R. Pisano, C. Urbano, J. Willhelm, and R. Powers (1998) "High-Resolution Solution Structure of the Inhibitor-Free Catalytic

Fragment of Human Fibroblast Collagenase Determined by Multidimensional NMR" *Biochemistry,* 37:1495-1504, doi: 10.1021/bi972181w.

[48] D. Mendoza-Juárez, M. Sánchez-Gutiérrez, A. J. Izquierdo-Vega, E. O. Madrigal-Santillán, C. Velázquez-González, and J. A. Izquierdo-Vega (2025) "Matrix Metalloproteinase Inhibitors and Their Potential Clinical Application in Periodontitis" *Diseases* 13(9), doi: 10.3390/diseases13090296.

[49] T. Fujisawa, S. Katakura, S. Odake, Y. Morita, J. Yasuda, I. Yasumatsu, and T. Morikawa (2001) "Design and Synthesis of Carboxylate Inhibitors for Matrix Metalloproteinases" *Chem Pharm Bull* 49(10):1272-1279, doi: 10.1248/CPB.49.1272.

[50] T. Singh, O. A. Adekoya, and B. Jayaram (2015) "Understanding the binding of inhibitors of matrix metalloproteinases by molecular docking, quantum mechanical calculations, molecular dynamics simulations, and a MMGBSA/MMBappl study" *Mol Biosyst*, 11(4):1041–1051, doi: 10.1039/c5mb00003c.

[51] C. Rouanet-Mehoua, B. Czarny, F. Beau, E. Cassar-Lajeunesse, E. A. Stura, V. Dive, and L. Devel (2017) "Zinc-Metalloproteinase Inhibitors: Evaluation of the Complex Role Played by the Zinc-Binding Group on Potency and Selectivity" *J Med Chem*, 60:403–414, doi: 10.1021/acs.jmedchem.6b01420.

[52] S. Cheng, Q. Zhang, H. Min, W. Jiang, J. Liu, C. Liu, and Z. Wang (2024) "Development of a Predictive Model for N-Dealkylation of Amine Contaminants Based on Machine Learning Methods" *Toxics*, 12(12), doi: 10.3390/toxics12120931.

Appendix A

MMP-1 QSAR Training Set (Pages: 36-39)

NAME	QSAR ID	PubChem CID	Canonical SMILE	pIC_{50}	REFERENCE (DOI)
CHEMBL1801412	A1	22485302	CCOC1=CC=C(C=C1)OC2=CC=C(C=C2)S(=O)(=O)C3(CCN(CC3)C4CC4)C(=O)NO	6	10.1021/jm100669j
CHEMBL366111	A2	9982414	C1COCC(C1S(=O)(=O)C2=CC=C(C=C2)OCC3=C(C=C(C=C3)Cl)Cl)(C(=O)NO)O	6.04	10.1016/j.bmcl.2004.06.081
CHEMBL79433	A3	9957653	C1CN(CCC1(C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC3=CC=C(C=C3)Cl)CC4=CC=CC=C4	6.1	10.1021/jm0205550
CHEMBL399351	A4	44447353	C1COCC1OC(=O)N2CCC(CC2)(CS(=O)(=O)N3CCN(CC3)C4=CC=CC=C4)C(=O)NO	6.11	10.1016/j.bmcl.2007.11.086
CHEMBL198778	A5	44404553	C1=CC=C(C=C1)OC2=CC=C(C=C2)S(=O)(=O)NCC(=O)NO	6.14	10.1016/j.ejmech.2012.12.058
beta-sulfone 7d	A6	11548265	C1CN(CCC1(CS(=O)(=O)C2=CC=C(C=C2)OC3=CC=CC=C3)C(=O)NO)CCC4=CC=CC=C4	6.16	10.1016/j.bmcl.2006.10.004
CHEMBL3617404	A7	44634267	CC(C)ON(C(CCNC(=O)C)C(=O)NO)S(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)OC	6.18	10.1021/acs.jmedchem.5b00367
CHEMBL550148	A8	25049753	C1=CC=C(C=C1)C2=CC=C(C=C2)S(=O)(=O)NCC(=O)NO	6.21	10.1021/jm900261f
CHEMBL148171	A9	44364250	COC1=C(C=C2CCN(C(C2=C1)C(=O)NO)S(=O)(=O)C3=CC=CC=C3)O	6.22	10.1016/j.bmcl.2003.10.026
CHEMBL227598	A10	44422723	CC(C)C(NS(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)OC(C)C)P(=O)(O)O	6.22	10.1016/j.bmc.2006.10.047
CHEMBL234529	A11	25181080	COC1=CC=C(C=C1)S(=O)(=O)N(CC(=O)NCCC2=CC=C(C=C2)S(=O)(=O)N)CC(=O)NO	6.27	10.1021/jm800964f
CHEMBL234735	A12	25181081	C1=CC=C(C=C1)C2=CC=C(C=C2)S(=O)(=O)N(CC(=O)NCCC3=CC=C(C=C3)S(=O)(=O)N)CC(=O)NO	6.27	10.1021/jm800964f
ARP 101	A13	11292680	CC(C)C(C(=O)NO)N(OC(C)C)S(=O)(=O)C1=CC=C(C=C1)C2=CC=CC=C2	6.31	10.1021/jm900335a
CHEMBL227824	A14	44422720	CCOC1=CC=C(C=C1)C2=CC=C(C=C2)S(=O)(=O)NC(C(C)C)P(=O)(O)O	6.35	10.1016/j.bmc.2006.10.047
alpha-Sulfone 21	A15	9822926	C1COCCC1(C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC3=CC=C(C=C3)Cl	6.36	10.1021/jm100669j
CHEMBL179552	A16	11752434	CC1(CC(CN(C1C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OCC3=C(C=C(C=C3)Cl)Cl)O)C	6.37	10.1016/j.bmcl.2005.03.105
CP-544439	A17	9866250	C1COCCC1(C(=O)NO)NS(=O)(=O)C2=CC=C(C=C2)OC3=CC=C(C=C3)F	6.38	10.1016/j.bmcl.2004.04.083
CHEMBL1088878	A18	45378254	CC#CCOC1=CC=C(C=C1)S(=O)(=O)N(CC(C)C)C(CCN2C(=O)C3=CC=CC=C3C2=O)C(=O)NO	6.4	10.1021/jm901868z
CHEMBL1089843	A19	45378253	CC#CCOC1=CC=C(C=C1)S(=O)(=O)NC(CCN2C(=O)C3=CC=CC=C3C2=O)C(=O)NO	6.46	10.1021/jm901868z
CHEMBL227501	A20	44422743	CC1=CC(=CC=C1)C2=CC=C(C=C2)S(=O)(=O)NC(C(C)C)P(=O)(O)O	6.5	10.1016/j.bmc.2006.10.047
CHEMBL228188	A21	44422737	CC(C)C(NS(=O)(=O)C1=CC=C(C=C1)C2=CC(=CC=C2)Cl)P(=O)(O)O	6.5	10.1016/j.bmc.2006.10.047
CHEMBL7108	A22	10362663	CC(C)(C)N1CN(C(CC1=O)C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC	6.5	10.1016/s0960-894x(01)00137-8
CHEMBL187092	A23	44395506	CC1(CCCN(C1C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OCC3=C(C=C(C=C3)F)Cl)O	6.51	10.1021/jm300449x
CHEMBL1801044	A24	22485417	C1COCCC1(C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC3=CC=CC=C3	6.57	10.1021/jm100669j
CHEMBL251820	A25	11605631	C1COCC1OC(=O)N2CCC(CC2)(CS(=O)(=O)N3CCC(=CC3)C4=CC=CC=C4)C(=O)NO	6.59	10.1016/j.bmcl.2007.11.086

CHEMBL456774	A26	25178357	CC(C)C(C(=O)NO)N(CC(=O)O)S(=O)(=O)C1=CC=C(C=C1)C2=CC=CC=C2	6.63	10.1021/jm800964f
CHEMBL1090647	A27	45380190	CC(=O)NCCC(C(=O)NO)NS(=O)(=O)C1=CC=C(C=C1)C2=CC=CC=C2	6.71	10.1016/j.bmc.2018.11.041
CHEMBL1078281	A28	46175170	CC1C(C1(C(=O)O)NS(=O)(=O)C2=CC=C(S2)N3C=C(C=N3)Cl)C4=CC=CC=C4	6.75	10.1021/jm101609j
CHEMBL3617407	A29	24936157	CC(C)ON(C(CCN1C(=O)C2=CC=CC=C2C1=O)C(=O)NO)S(=O)(=O)C3=CC=C(C=C3)C4=CC=C(C=C4)OC	6.77	10.1016/j.bmc.2018.11.041
CHEMBL374377	A30	44422724	CC(C)C(NS(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)F)P(=O)(O)O	6.8	10.1016/j.bmc.2006.10.047
CHEMBL147882	A31	44364122	CC1=CC=C(C=C1)S(=O)(=O)N2CCC3=C(C2C(=O)NO)C=CC(=C3)O	6.82	10.1016/j.bmcl.2003.10.026
CHEMBL202932	A32	6852147	CC(C)C(NS(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)OC)P(=O)(O)O	6.82	10.1016/j.bmc.2006.10.047
Hydroxamate 14	A33	10042482	CC(C(=O)NO)NS(=O)(=O)C1=C(C(=C(C(=C1F)F)F)F)F	6.82	10.1016/j.bmc.2007.01.011
CHEMBL234530	A34	25178145	C1=CC=C(C=C1)OC2=CC=C(C=C2)S(=O)(=O)N(CC(=O)NCCC3=CC=C(C=C3)S(=O)(=O)N)CC(=O)NO	6.84	10.1021/jm800964f
Hydroxamate 20	A35	11792968	CC(C)C(C(=O)NO)NS(=O)(=O)C1=C(C(=C(C(=C1F)F)F)F)F	6.86	10.1016/j.bmc.2007.01.011
GI-254023X	A36	9952396	CC(C(CCCC1=CC=CC=C1)C(=O)NC(C(=O)NC)C(C)(C)C)N(C=O)O	6.9	10.1016/j.ejmech.2016.01.053
CHEMBL148214	A37	44364249	COC1=CC=C(C=C1)S(=O)(=O)N2CCC3=CC(=C(C=C3C2C(=O)NO)OC)O	6.91	10.1016/j.bmcl.2003.10.026
CHEMBL317508	A38	9890226	C1=CC=C2C(=C1)C(=O)N(N=N2)CCC(COC3=CC=C(C=C3)C4=CC=C(C=C4)C#N)C(=O)NO	6.94	10.1016/s0960-894x(00)00646-6
CHEMBL433262	A39	10550399	CC(C)C(C(=O)NO)NS(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)Br	6.96	10.1021/jm9903141
CHEMBL296911	A40	44291166	CC(C)CN(C(CCSCC1=CC=CC=C1)C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC	6.98	10.1016/s0960-894x(99)00259-0
CHEMBL267939	A41	44263913	COC1=CC=C(C=C1)S(=O)(=O)N2CCN(CCC2C(=O)NO)C(=O)OCC3=CC=CC=C3	7.01	10.1016/s0960-894x(01)00137-8
CHEMBL358539	A42	10367355	COC1=CC=C(C=C1)S(=O)(=O)N2CCC3=C(C2C(=O)NO)C=CC(=C3)OCC4=CC=CC=C4	7.06	10.1016/j.bmcl.2003.10.026
CHEMBL376557	A43	9822461	CC(C)C(NS(=O)(=O)C1=CC=C(C=C1)C2=CC=C(C=C2)Cl)P(=O)(O)O	7.11	10.1016/j.bmc.2006.10.047
CHEMBL515980	A44	25180867	CC(C)C(C(=O)NO)N(CC(=O)O)S(=O)(=O)C1=CC=C(C=C1)OC2=CC=CC=C2	7.23	10.1021/jm800964f
Hydroxamate 18	A45	15296597	CC(C(=O)NO)N(CC1=CC=CC=C1)S(=O)(=O)C2=CC=C(C=C2)OC	7.24	10.1016/j.bmc.2007.01.011
CHEMBL432882	A46	6539670	CC(C)C(C(=O)NO)NS(=O)(=O)C1=CC=C(C=C1)OC2=CC=CC=C2	7.3	10.1021/jm9903141
CTS-1027	A47	3342298	C1COCCC1(CS(=O)(=O)C2=CC=C(C=C2)OC3=CC=C(C=C3)Cl)C(=O)NO	7.47	10.1021/acs.jmedchem.6b01007
Apratastat	A48	11452716	CC1(C(N(CCS1)S(=O)(=O)C2=CC=C(C=C2)OCC#CCO)C(=O)NO)C	7.48	10.1016/j.bmcl.2005.12.020
UNII-9P3H27MP46	A49	448002	CC(C)CN(CC(=O)NO)S(=O)(=O)C1=CC=C(C=C1)OC	7.51	10.1016/j.ejmech.2019.111875
CHEMBL200274	A50	44406655	CC(C)CC(C(C(=O)NO)O)C(=O)NC1CCCCCCCCCCNC1=O	7.52	10.1021/jm050363f
Hydroxamate 10	A51	10837740	C1=CC=C(C=C1)CN(CC(=O)NO)S(=O)(=O)C(C(C(C(F)(F)F)(F)F)(F)F)(F)F	7.52	10.1016/j.bmc.2007.01.011

CHEMBL427616	A52	10006134	COC1=CC=C(C=C1)S(=O)(=O)N2CCC(N(CC2)S(=O)(=O)C3=CC=C(C=C3)OC)C(=O)NO	7.54	10.1016/s0960-894x(01)00137-8
CHEMBL11306	A53	194777	CC(C)CC(CC(=O)NO)C(=O)NC(CC1=CC=CC=C1)C(=O)NC	7.6	10.1016/j.ejmech.2007.07.002
Mmp inhibitor II	A54	4218	CC1(CN(C(N(C1)S(=O)(=O)C2=CC=C(C=C2)OC)C(=O)NO)S(=O)(=O)C3=CC=C(C=C3)OC)C	7.62	10.1021/acs.jmedchem.7b00315
CHEMBL20154	A55	10045150	CC(C)(C)C(C(=O)NC)NC(=O)C(CCCC1=CC=CC=C1)CC(=O)NO	7.68	10.1021/jm970404a
CHEMBL176602	A56	9543420	CC(C)CC(C(=O)NO)N(CC1=CC=CC=C1)P(=O)(C)C2=CC=CC=C2	7.69	10.1021/jm980142s
SC-44463	A57	128564	CC(C)CC(CC(=O)NO)C(=O)NC(CC1=CC=C(C=C1)OC)C(=O)NC	7.7	10.1021/jm0308038
CHEMBL3889936	A58	136054516	CC(C)C(CS(=O)(=O)C1=CC=C(C=C1)C2=CC=CC(=C2)CNC(=O)C3=NC4=CC=CC=C4C(=O)N3)N(C=O)O	7.75	10.1016/j.bmc.2016.09.009
CHEMBL95237	A59	44329018	CC(C)C=C(CC(=O)NO)C(=O)NC(CC1=CNC2=CC=CC=C21)C(=O)NC	7.75	10.1016/j.bmc.2007.05.001
CHEMBL415482	A60	44263926	COC1=CC=C(C=C1)S(=O)(=O)N2CCN(C(=O)CC2C(=O)NO)CC3=CC=CC=C3	7.87	10.1016/s0960-894x(01)00137-8
CHEMBL3932562	A61	137287436	CC(C)C(C(=O)NO)NS(=O)(=O)C1=CC=C(C=C1)C2=CC=CC(=C2)CNC(=O)C3=NC4=CC=CC=C4C(=O)N3	7.96	10.1016/j.bmc.2016.09.009
CHEMBL311266	A62	10164317	COCCOC(=O)N1CCC(CC1)C(C(=O)O)NS(=O)(=O)C2=CC=C(C=C2)C3=CC=C(C=C3)OC	8.09	10.1021/jm015531s
Prinomastat	A63	466151	CC1(C(N(CCS1)S(=O)(=O)C2=CC=C(C=C2)OC3=CC=NC=C3)C(=O)NO)C	8.09	10.1016/j.bmcl.2005.01.024
CHEMBL471537	A64	25147773	CC(C)C(C(=O)NO)N(CC(=O)NCCC1=CC=C(C=C1)S(=O)(=O)N)S(=O)(=O)C2=CC=C(C=C2)OC3=CC=CC=C3	8.11	10.1021/jm800964f
Hydroxamate 17	A65	23645529	CC(C(=O)NO)N(CC1=CC=CC=C1)S(=O)(=O)C2=C(C(=C(C(=C2F)F)F)F)F	8.15	10.1016/j.bmc.2007.01.011
Hydroxamate 38	A66	23645533	CC(C(=O)NO)N(CC1=CC=CC=C1[N+](=O)[O-])S(=O)(=O)C2=C(C(=C(C(=C2F)F)F)F)F	8.16	10.1016/j.bmc.2007.01.011
CHEMBL98044	A67	10453826	CN(CC(C(CC1CCCC1)C(=O)N2CCCCC2)C(=O)NO)S(=O)(=O)C	8.22	10.1016/j.bmc.2008.08.058
Hydroxamate 32	A68	10671441	C1=CC=C(C(=C1)CN(CC(=O)NO)S(=O)(=O)C2=C(C(=C(C(=C2F)F)F)F)F)[N+](=O)[O-]	8.22	10.1016/j.bmc.2007.01.011
CHEMBL115189	A69	44342614	CC(C)CC(C(C(=O)NO)O)C(=O)NC(C(=O)NC)C(C)(C)C	8.3	10.1021/jm990412m
CHEMBL441662	A70	10549114	CC1(C(N(CCCCCS1)S(=O)(=O)C2=CC=C(C=C2)OC)C(=O)NO)C	8.52	10.1021/jm990330y
CHEMBL45631	A71	9887544	CC(C)CC1C(CCCOC2=CC=C(CC(NC1=O)C(=O)NC)C=C2)C(=O)NO	8.52	10.1016/s0960-894x(98)00396-5
Hydroxamate 35	A72	10456675	C1=CC(=CC=C1CN(CC(=O)NO)S(=O)(=O)C2=C(C(=C(C(=C2F)F)F)F)F)[N+](=O)[O-]	8.52	10.1016/j.bmc.2007.01.011
CHEMBL140152	A73	44360784	CC1(CCCCN(C1C(=O)NO)S(=O)(=O)C2=CC=C(C=C2)OC)C	8.62	10.1021/jm990330y
Cipemastat	A74	9824350	CC1(C(=O)N(C(=O)N1C)CC(C(CC2CCCC2)C(=O)N3CCCCC3)C(=O)NO)C	8.64	10.1021/jm801394m
CHEMBL432079	A75	9908389	CC(C)CC(C(CC=C)C(=O)NO)C(=O)NC(CC1=CC=CC=C1)C(=O)NC	8.66	10.1016/s0960-894x(98)00255-8
CHEMBL342210	A76	11795621	CC1(C(N(CCCS1(=O)=O)S(=O)(=O)C2=CC=C(C=C2)OC)C(=O)NO)C	8.72	10.1021/acs.jmedchem.7b00315

Galardin	A77	132519	CC(C)CC(CC(=O)NO)C(=O)NC(CC1=CNC2=CC=CC=C21)C(=O)NC	8.82	10.1016/j.bmc.2007.05.001
Marimastat	A78	119031	CC(C)CC(C(C(=O)NO)O)C(=O)NC(C(=O)NC)C(C)(C)C	8.82	10.1021/jm0307638
Batimastat	A79	5362422	CC(C)CC(C(CSC1=CC=CS1)C(=O)NO)C(=O)NC(CC2=CC=CC=C2)C(=O)NC	9	10.1021/jm0308038
CHEMBL337594	A80	44360287	CC1(C(N(CC(CS1)OCOC)S(=O)(=O)C2=CC=C(C=C2)OC)C(=O)NO)C	9	10.1021/jm990330y
CHEMBL140970	A81	9976697	CC1(C(N(CCCS1)S(=O)(=O)C2=CC=C(C=C2)OC)C(=O)NO)C	9.1	10.1021/acs.jmedchem.7b00315
CHEMBL279078	A82	10047585	CC(C)CC(C(C)C(=O)NO)C(=O)NC(CC1=CC=CC2=C1CCCC2)C(=O)NC	9.52	10.1021/jm970404a
CHEMBL147489	A83	44364119	C1CN(C(C2=C1C=C(C=C2)O)C(=O)NO)S(=O)(=O)C3=CC=C(C=C3)[N+](=O)[O-]	10.39	10.1016/j.ejmech.2012.10.016