



IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS

Aleix Gimeno Vives

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

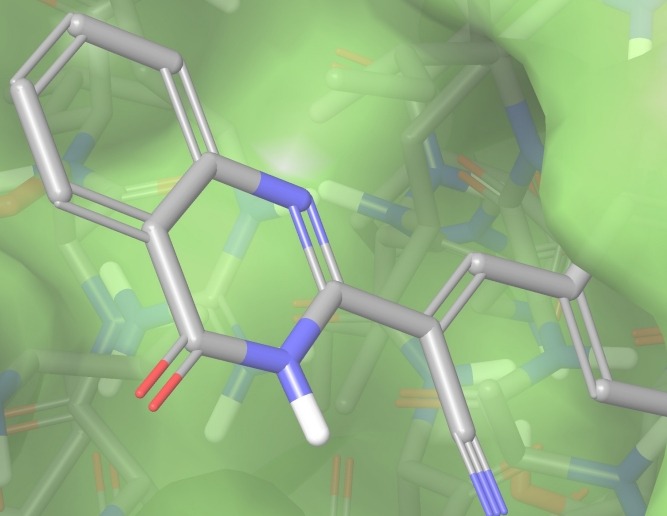
ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



UNIVERSITAT
ROVIRA I VIRGILI

Identification by virtual screening of
protein tyrosine phosphatase 1B and
matrix metalloproteinase 13 inhibitors
for the treatment of obesity and
obesity-associated disorders



Aleix Gimeno
Doctoral Thesis
2018

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Aleix Gimeno Vives

**Identification by virtual screening of
protein tyrosine phosphatase 1B and
matrix metalloproteinase 13 inhibitors
for the treatment of obesity and
obesity-associated disorders**

Doctoral Thesis

Supervised by

Dr. Gerard Pujadas Anguiano and Dr. Santiago Garcia Vallvé

Cheminformatics & Nutrition Research Group
Biochemistry & Biotechnology Department



UNIVERSITAT ROVIRA i VIRGILI

Tarragona 2018

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo



UNIVERSITAT
ROVIRA i VIRGILI

WE STATE that the present study, entitled "*Identification of protein tyrosine phosphatase 1B and matrix metalloproteinase 13 inhibitors by virtual screening for the treatment of obesity and obesity-associated disorders*", presented by Aleix Gimeno Vives for the award of the degree of Doctor, has been carried out under our supervision at the Biochemistry and Biotechnology Department of this university.

FEM CONSTAR que aquest treball, titulat "*Identification of protein tyrosine phosphatase 1B and matrix metalloproteinase 13 inhibitors by virtual screening for the treatment of obesity and obesity-associated disorders*", que presenta l'Aleix Gimeno Vives per a l'obtenció del títol de Doctor, ha estat realitzat sota la nostra direcció al Departament de Bioquímica i Biotecnologia d'aquesta universitat.

HACEMOS CONSTAR que el presente trabajo, titulado "*Identification of protein tyrosine phosphatase 1B and matrix metalloproteinase 13 inhibitors by virtual screening for the treatment of obesity and obesity-associated disorders*", que presenta Aleix Gimeno Vives para la obtención del título de Doctor, ha sido realizado bajo nuestra dirección en el Departamento de Bioquímica y Biotecnología de esta universidad.

Tarragona, November 21st, 2018

Doctoral Thesis Supervisors

Dr. Gerard Pujadas Anguiano

Dr. Santiago Garcia Vallvé

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Agraïments

Gràcies Gerard i Santi per totes les idees, la professionalitat i sobretot per la vostra proximitat. Aquesta tesi és tant meva com vostra. Gerard, gràcies per encomanar les ganes de treballar amb la teva dedicació i les ganes de fer-ho millor amb el teu ull clínic. Gràcies a la millor professora de manualitats del PDB, la MJosé. Pels teus consells, el bon humor i el teu caràcter lluitador. Gràcies a l'Adrià, per la paciència del qui té sempre la resposta correcta. Gràcies Andrea i Miquel, per guiar aquest bioinformàtic pel laboratori. Gràcies Sarah, Manuel, Joel, Sandra, Dmytro i tots els que heu fet del laboratori de quimioinformàtica més que un lloc. Gràcies papa, mama, *Lushi*, avis, família. Per ser-hi, per donar-me totes les facilitats i animar-me a seguir endavant. Miri, gràcies també per la correcció de l'anglès d'aquesta tesi. Gràcies Pep, Lara i Milo, per les estones d'estudi i d'esbarjo. I gràcies Mireia, per fer-me compartir, somniar i somriure.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Table of Contents

Summary – Resum – Resumen.....	9
Introduction.....	15
Manuscript 1.....	33
<i>The light and dark sides of virtual screening: what is there to know?</i>	
Hypotheses and Objectives.....	75
Results.....	79
Manuscript 2.....	81
<i>Combined ligand- and receptor-based virtual screening methodology to identify structurally diverse protein tyrosine phosphatase 1B inhibitors</i>	
Manuscript 3.....	131
<i>Understanding the variability of the S1' pocket to improve matrix metalloproteinase inhibitor selectivity profiles</i>	
Manuscript 4.....	219
<i>Identification of selective MMP-13 inhibitors by virtual screening</i>	
Summarizing Discussion.....	253
Conclusions.....	257
Publication List.....	261

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Summary

Obesity is one of the major public health problems in the 21st century. The great economic expansion of the last decades in developed countries has contributed to the increased consumption of unhealthy foods and the excessive usage of energy-saving technologies. These have in turn led to the development of unhealthy lifestyles and the consequent increase of obesity prevalence. Thus, obesity has emerged as a natural response to an unnatural environment. With the continuous increase in obese population in each generation, the prevalence of obesity-associated disorders such as type II diabetes and osteoarthritis is also on the rise, and the prospect of developing a medical therapy specific for each patient earns increasing interest. In this regard, the targets protein tyrosine phosphatase 1B (PTP1B) and matrix metalloproteinase 13 (MMP-13) are involved in both obesity and, respectively, type II diabetes mellitus and osteoarthritis. The present doctoral thesis focuses on developing virtual screening strategies to identify compounds that modulate the activity of these two targets which may have a positive influence on both obesity and its associated disorders.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

°°

Resum

L'obesitat és un dels principals problemes de salut pública del segle XXI. La gran expansió econòmica de les últimes dècades en els països desenvolupats ha contribuït a l'increment del consum d'aliments poc saludables i a l'ús excessiu de tecnologies d'estalvi d'energia. Aquests canvis han generat estils de vida poc saludables i el consegüent augment de la prevalença d'obesitat. Així doncs, l'obesitat sorgeix com una resposta natural a un entorn antinatural. Amb l'augment continu de la població obesa en cada generació, la prevalença de trastorns associats a l'obesitat com la diabetis tipus II i l'artrosi també augmenta, i la perspectiva de desenvolupar una teràpia mèdica específica per a cada pacient va guanyant interès. En aquest sentit, les dianes proteïna tirosina fosfatasa 1B (PTP1B) i la metaloproteasa de la matriu 13 (MMP-13) estan implicades tant en l'obesitat com, respectivament, la diabetis mellitus de tipus II i l'artrosi. La present tesi doctoral es centra en el desenvolupament d'estratègies de cribratge virtual per tal d'identificar compostos que modulin l'activitat d'aquestes dues dianes i puguin influir positivament en l'obesitat i els trastorns associats a l'obesitat.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

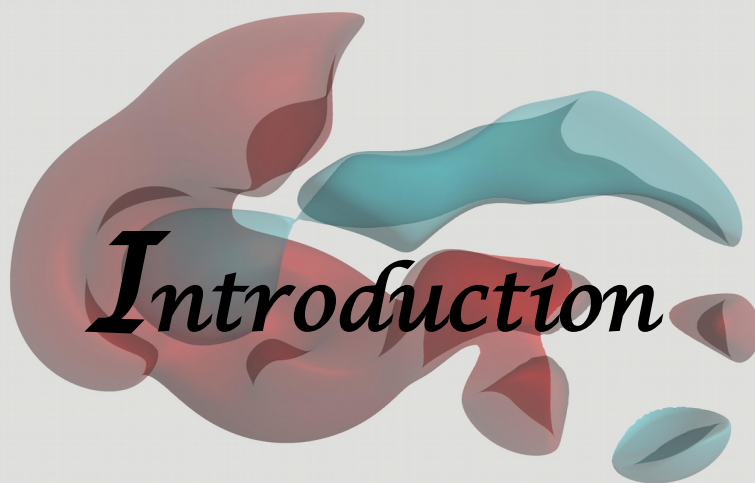
oo

Resumen

La obesidad es uno de los principales problemas de salud pública del siglo XXI. La gran expansión económica de las últimas décadas en los países desarrollados a contribuido al incremento del consumo de alimentos poco saludables y al uso excesivo de tecnologías de ahorro de energía. Estos cambios han generado estilos de vida poco saludables y el consiguiente aumento de la prevalencia de obesidad. Así pues, la obesidad surge como una respuesta natural a un entorno antinatural. Con el aumento continuo de la población obesa en cada generación, la prevalencia de trastornos asociados a la obesidad como la diabetes tipo II y la artrosis también aumenta, y la perspectiva de desarrollar una terapia médica específica para cada paciente va ganando interés. En este sentido, las dianas proteína tirosina fosfatasa 1B (PTP1B) y la metaloproteasa de la matriz 13 (MMP-13) están implicadas tanto en la obesidad como, respectivamente, la diabetes mellitus de tipo II y la artrosis. La presente tesis doctoral se centra en el desarrollo de estrategias de cribado virtual para identificar compuestos que modulen la actividad de estas dos dianas y puedan influir positivamente en la obesidad y los trastornos asociados a la obesidad.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo



UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Overweight and obesity are defined as abnormal or excessive fat accumulation that may impair health.¹ The prevalence of obesity has been continuously increasing worldwide in the last decades,² having nearly tripled since 1975.¹ In 2016, 39% of adults aged 18 years and over (1.9 billion) were overweight, and 13% (650 million) were obese.¹ Obesity is associated with multiple pathologies, such as type II diabetes mellitus, osteoarthritis, cardiovascular diseases, asthma, gallbladder disease, chronic back pain and cancer.³⁻⁵ The increasing prevalence of childhood obesity is also a matter of concern, as it is associated with greater risk and earlier onset of chronic disorders.² Consequently, as childhood obesity continues to rise,² obesity-associated morbidities are expected to develop earlier and more frequently in future generations, thus increasing the need for more effective pharmacological treatment. In fact, most of the world's population live in countries where overweight and obesity are associated with a higher mortality rate than underweight.¹ This supposes a great economic burden and establishes obesity as one of the major public health problems of the 21st century.⁴

General recommendations body weight reduction include a healthy diet and regular physical activity.⁶ However, if the patient presents obesity-associated morbidities and is not able to achieve this reduction of body weight by lifestyle alone, medical treatments including pharmacological therapies and bariatric surgery are available. As bariatric surgery has more associated health risks and costs,^{7,8} it is reserved for clinically severe obesity, making pharmacological therapy the initial treatment option for obese patients with comorbidities.⁶ Some of the drugs used to treat obesity include Phentermine, Orlistat, Lorcaserin, Naltrexone/bupropion, Lorcaserin and Liraglutide (see Table 1).^{6,9,10}

Introduction

Table 1. Current common pharmacological treatments for obese patients^{6,9,10}

Pharmacological treatment	Mechanism of action	Effects	Adverse effects	Recommended patients
Orlistat	Pancreatic and gastric lipase inhibitor	Pancreatic and gastric lipase inhibitor	Fecal urgency, oily stool, flatus with discharge, fecal incontinence	Patients with hypercholesterolemia and/or constipation who can limit their intake of dietary fat
Phentermine/topiramate	Adrenergic agonist and neurostabilizer	Increases resting energy expenditure and suppresses appetite	Paresthesias, dizziness, dysgeusia, insomnia, constipation, dry mouth	Younger patients who need assistance with appetite suppression
Lorcaserin	Selective serotonin receptor agonist	Reduces appetite and increases satiety	Headache, dizziness, fatigue, nausea, dry mouth, constipation	Patients who would benefit from appetite suppression
Naltrexone/bupropion	Opioid antagonist and dopamine and norepinephrine reuptake inhibitor	Reduces appetite and food cravings	Nausea, vomiting, constipation, headache, dizziness, insomnia, dry mouth	Patients who describe addictive behaviors related to food
Liraglutide	Glucagonlike peptide-1 gastrointestinal hormone mimic	Reduces hunger, decreases food intake, and delays gastric emptying	Nausea, vomiting, diarrhea, constipation, dyspepsia, abdominal pain	Patients who report inadequate meal satiety

Despite the availability of these pharmacological agents, their limited efficacy and adverse effects constitute serious limitations for these drugs, and there is an unmet need for safe, efficacious and tolerable anti-obesity medications (see Table 1).⁷ Moreover, these treatments are directed exclusively at the modulation of food intake and they are not aimed at treating obesity-associated disorders comorbidities directly. Therefore, it would be interesting to identify drugs that could modulate other metabolic pathways involved in obesity and that were able to address obesity-

associated morbidities, as they should increase the effectiveness of the treatment in patients presenting the corresponding comorbidity.

Given the impact of obesity in modern society, there is a need for new pharmacological agents with different modes of action aimed at treating obesity and its associated diseases. The present thesis focuses on identifying inhibitors by virtual screening (VS) that target obesity and obesity-associated comorbidities. Thus, protein tyrosine phosphatase 1B (PTP1B) and matrix metalloproteinase 13 (MMP-13) have been selected for their role in the progression of obesity and their respective implications on type II diabetes mellitus and osteoarthritis, two morbidities associated with obesity. Here is a brief description of each target of interest and their implication in obesity and the corresponding associated pathology as well as an introduction to virtual screening techniques.

Protein tyrosine phosphatase 1B

Phosphorylation plays an essential role in cell signaling, as it modulates the activity of many proteins in the cell. Many signaling pathways involve the phosphorylation of proteins, which can occur at different stages of the cell signaling cascade. The enzymes that catalyze protein phosphorylation and dephosphorylation are known as protein kinases and protein phosphatases, respectively. In phosphorylation events, protein kinases catalyze the addition of phosphoryl groups on a protein, activating or deactivating it, depending on the target protein, thus triggering a cellular response. In dephosphorylation events, protein phosphatases remove the phosphoryl group from the protein to reverse the cycle and return the protein to its previous state. It has been estimated that the human genome contains over 1000 genes that encode for kinases and phosphatases. As the coordinated interplay between these two partners is required for the proper functioning of the cell, this illustrates the importance and complexity of phosphorylation processes, which regulate many metabolic pathways, including the metabolism of glucides, lipids and amino acids.¹¹ A particular

Introduction

phosphatase. PTP1B, has become a focus of attention in the last 20 years due to its involvement in the regulation of the insulin and leptin signaling pathways.¹²

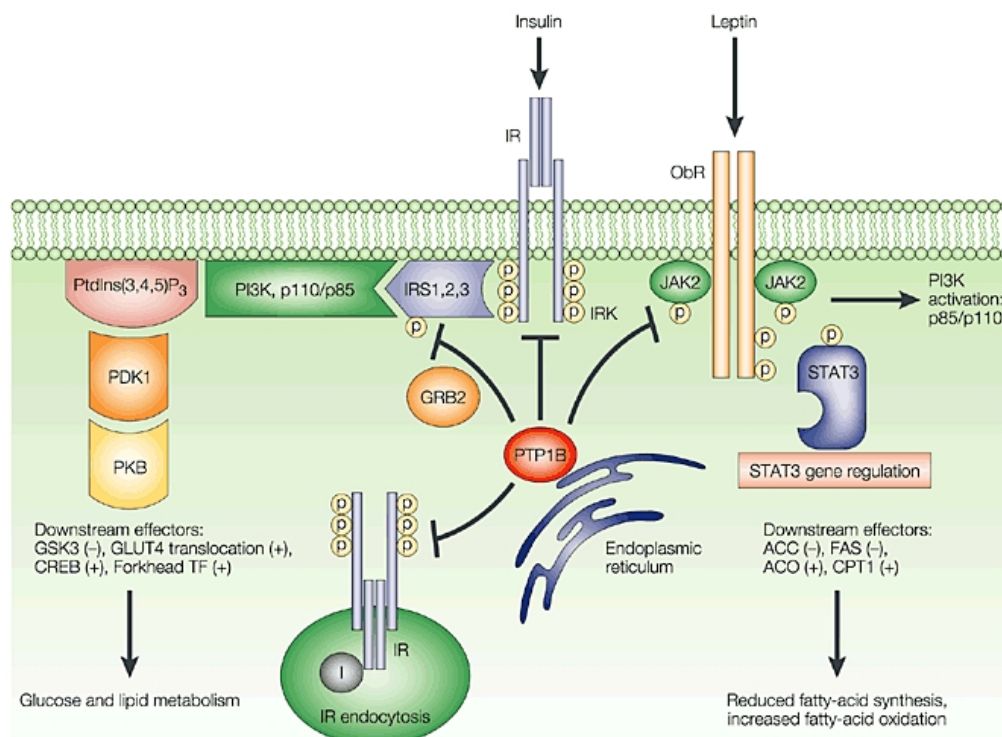


Figure 2. PTP1B involvement in the insulin and leptin signaling pathways.¹³

In the insulin signaling pathway, PTP1B dephosphorylates the insulin receptor^{14,15} and insulin receptor substrate 1 (IRS1),¹⁶ resulting in a down-regulation of the insulin signaling pathway (see Figure 2). *In vivo* studies confirm the role of PTP1B antagonizing insulin action, as PTP1B knockout (KO) mice display enhanced insulin sensitivity.¹⁷ Thus, PTP1B KO mice are more resistant to weight gain and remain insulin sensitive when they are fed with a high-fat diet. However, with the same diet, mice with the active PTP1B gene rapidly gain weight and become insulin resistant.¹⁷

In the leptin signaling pathway, PTP1B dephosphorylates the leptin receptor-associated kinase JAK2 and alters the JAK-STAT pathway, attenuating the action of

leptin (see Figure 2).^{18,19} *In vivo* studies have shown that PTP1B KO mice exhibit leptin hypersensitivity¹⁸ and an enhanced response toward leptin-mediated weight loss and suppression of feeding.¹⁹

Taken together, these findings indicate that the inhibition of PTP1B should contribute to the increase of insulin and leptin sensitivities and therefore have a positive impact in diabetic and obese patients displaying insulin and leptin resistance. Therefore, PTP1B has been established as a pharmacological target for both type II diabetes mellitus and obesity.^{20,21}

Matrix metalloproteinase 13

The extracellular matrix (ECM) consists of a network of macromolecules which not only provide physical support to the cell, but also transmit mechanical and molecular signals to communicate with the surrounding cells.²² The three major components of the ECM are: **a)** glycosaminoglycans (GAGs), usually covalently linked to protein in the form of *proteoglycans*, are large and highly charged polysaccharides that form a highly hydrated gel-like substance, resisting compressive forces and allowing the diffusion of nutrients, metabolites and hormones; **b)** fibrous proteins (primarily members of the *collagen* family), which confer both structure and elasticity to the ECM; and **c)** a large and varied assortment of *glycoproteins*, which help cells migrate, settle and differentiate in the appropriate locations (see Figure 3).²²

Introduction

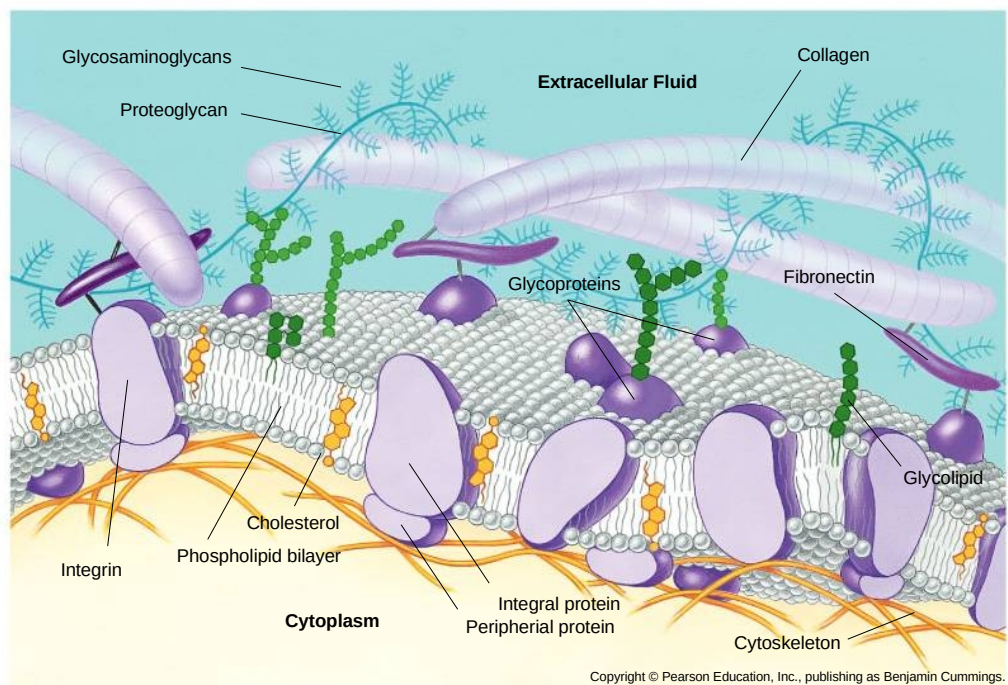


Figure 3. Components of the extracellular matrix.

As important as the ability of cells to build and bind to the ECM is their ability to degrade it. ECM degradation is required in many cellular processes as cells may need to either stretch out in order to divide, detach from other cells in order to migrate, or remove cellular material in order for the tissue to grow, be repaired, and be maintained in a continuous turnover of ECM components. Proteases are responsible for the degradation of matrix components. The largest group of proteases that degrade the ECM are the matrix metalloproteases (MMPs), a family of proteases dependent on Ca^{2+} or Zn^{2+} that degrade different components of the ECM with different specificity (see Table 2).²²

Table 2. MMPs, their alternative names, functional classification, main substrates and some of the pathologies in which they are involved.

MMP	Alternative names	Functional classification	Main substrates	Related Pathologies
MMP-1	- Interstitial collagenase - Collagenase 1 - Vertebrate collagenase	- Collagenases	- Collagen types I, II and III	- Rheumatoid arthritis ²⁴ - Cancer ^{23,30}
MMP-2	- Gelatinase A - 72 kDa gelatinase - Type IV collagenase	- Gelatinases	- Collagen type IV - Gelatin - Fibrinogen	- Asthma ^{24,31} - Cancer ³² - Cardiovascular diseases ²⁴ - Heart failure ²⁴ - Liver fibrosis ³³
MMP-3	- Stromelysin 1 - Proteoglycanase - Transin	- Stromelysins	- Proteoglycan - Fibronectin - Collagen types I, III, IV, V and IX	- Atherosclerosis ³⁴ - Coronary artery disease ³⁶ - Inflammatory bowel disease ²⁴ - Periodontitis ³⁵
MMP-7	- Matrilysin - PUMP-1 - Putative metalloproteinase-1 - Uterine metalloendopeptidase	- Matrilysins	- Elastin - Fibronectin - Casein - Laminin	- Cancer ³² - Inflammatory bowel disease ²⁷ - Lung fibrosis ²⁴
MMP-8	- Neutrophil collagenase	- Collagenases	- Collagen types I and III	- Asthma ²⁴ - Cancer ²³ - Periodontitis ²⁴ - Rheumatoid arthritis ³⁴
MMP-9	- Gelatinase B - 92 kDa gelatinase - Type IV collagenase - Macrophage gelatinase	- Gelatinases	- Collagen types I, III, IV, V and XI - Gelatin types I and V - Laminin	- Asthma ²⁴ - Cancer ³³ - Heart failure ²⁴ - Inflammatory bowel disease ³⁵ - Rheumatoid arthritis ³¹ - Liver fibrosis ³⁷
MMP-10	- Stromelysin 2 - Transin 2	- Stromelysins	- Fibronectin - Proteoglycan - Gelatin types I, III, IV, V - Collagen types I, III, IV and V	- Atherothrombosis ³⁹ - Chronic obstructive pulmonary disease ³⁸

Introduction

MMP-12	- Macrophage elastase - Metalloelastase	- Metalloelastases	- Elastin - Casein - Fibronectin - Gelatin - Laminin - Collagen type IV	- Chronic obstructive pulmonary disease ⁴⁰ - Neurological diseases ²⁴
MMP-13	- Collagenase 3	- Collagenases	- Collagen types I, II, III - Fibrillin types 1 and 2	- Cancer ²³ - Inflammatory bowel disease ²⁴ - Osteoarthritis ²⁷ - Rheumatoid arthritis ²⁶ - Obesity ²⁵
MMP-14	- Membrane-type matrix metalloproteinase-1	- Membrane-type MMPs	- Collagen type I - Fibronectin - Laminin	- Aortic aneurysm ²⁸ - Cancer ²⁹

This information has been obtained from the BRENDA,⁴¹ KEGG,⁴² ExpASY⁴³ and MEROPS⁴⁴ databases.

As MMPs are expressed in different tissues and have different substrate specificities, their uncontrolled activity, and thus the excessive degradation of different ECM components in different tissues may originate a wide range of pathologies (see Table 2). In these cases, the inhibition of MMPs could help reduce the degradation of particular ECM components responsible for these diseases and have a positive outcome on patient prognosis. Thus, at present, inhibitors of many members of the MMP family are being searched for their therapeutical interest.

However, as each MMP fulfills its biological role in the cell, only the MMP (or MMPs) involved in the pathology should be targeted by these inhibitors in order not to alter biological processes other than the one which causes the disease in question. This becomes specially relevant when developing drugs directed at systemic administration, as MMP inhibitors that target undesired MMPs on multiple tissues alter their respective biological functions and may originate side effects. In fact, in the first years of MMP inhibitor drug design, selectivity of MMP inhibitors was not considered a priority and this caused the failure of many clinical trials as patients developed musculoskeletal syndrome (MSS), whose cause, although it remains unknown, has been speculated to be the result of the nonselective inhibition of multiple MMPs.^{45–47} Moreover, some MMPs such as MMP-3, MMP-8 and MMP-9 confer a protective role to the cell and their

inhibition should be avoided, as it could lead to the development of further pathologies.⁴⁸⁻⁵⁰

In light of these facts, the current trend is to search for selective MMP inhibitors. Nevertheless, this supposes a big challenge, as all these enzymes belong to the same family and their binding sites present a high degree of homology. In order to design inhibitors that are specific for a particular MMP, it is crucial to first identify the unique features of that MMP that could allow its targeting while sparing other MMPs.

One MMP for which selective inhibitors are of interest is MMP-13. This metalloprotease is highly expressed in osteoarthritis patients, which induces an excessive breakdown of collagen that results in an imbalance between collagen synthesis and degradation in the joint, leading to the progressive degeneration of articular cartilage and causing pain, swelling, ankylosis and limited mobility of the joint.⁵¹ Therefore, selective MMP-13 inhibitors are sought for the treatment of osteoarthritis.

Obesity is regarded as a major risk factor for the incidence and progression of osteoarthritis.^{5,52} Apart from the fact that increased body mass index causes inflammation in joints that bear body weight, such as the knees and the hip,^{53,54} leptin produced in the joint white adipose tissue acts as a pro-inflammatory adipokine and has been shown to induce the expression of MMP-13, thus contributing to the progression of osteoarthritis.⁵⁵⁻⁵⁸ Interestingly, MMP-13 has recently been identified to play a role in obesity, as Shih *et al.*²⁶ have reported that the inhibition of MMP-13 prevents diet-induced obesity in mice and suppresses adipogenesis in 3T3-L1 preadipocytes, proposing the inhibition of MMP-13 as a potential strategy to prevent obesity. Considering the roles of MMP-13 in both obesity and osteoarthritis and the well established association between these disorders, the selective inhibition of MMP-13 could address both of these conditions simultaneously, making MMP-13 an ideal target for patients with obesity and obesity-induced osteoarthritis.

General strategies for the treatment of osteoarthritis combine both pharmacological and non-pharmacological therapies.⁵⁹ Non-pharmacological treatment options include exercise to achieve weight loss,^{60,61} arthroplasty,^{62,63} spa and mud-bath therapies,^{64,65}

Introduction

ultrasounds⁶⁶ and intra-articular hyaluronic acid injections.⁶⁷⁻⁷⁰ Regarding pharmacological treatment in osteoarthritis, the primary choice is the administration of non-steroidal anti-inflammatory drugs (NSAIDs) such as diclofenac, ibuprofen, naproxen, mefenamic acid and celecoxib.⁷¹ Nevertheless, little is known of the disease-modifying properties of some of these agents and, although they may provide modest pain relief, these drugs have often been associated with unwanted cardiovascular and gastrointestinal side effects and they present a challenge at evaluating the risk-benefit ratio when prescribed to patients.⁷²

Due to these limitations of NSAIDs, there is no consensus on the best pharmacological treatment for osteoarthritis, thus highlighting the need for a more effective pharmacological agent. Contrary to current treatments, which focus on symptom management and pain relief, research on osteoarthritis treatments should be directed towards halting disease progression in order to improve the overall status of the patient. This is, for instance, the case of mesenchymal stem cell therapy, which benefits from the potential of stem cells to regenerate damaged tissue by injecting these cells directly to the synovial fluid, where they generate growth factors and cytokines that initiate the repair process in the joint.⁷³⁻⁷⁷ Similarly, the exploration of the metabolic pathways involved in the development of osteoarthritis should allow the development of novel pharmacological agents with alternative mechanisms of action directed at halting and reversing joint degradation in osteoarthritis patients. In this sense, as MMP-13 is actively involved in the degradation of collagen in the joint, selective MMP-13 inhibitors may constitute a novel therapy for osteoarthritis patients that should restrain disease progression while avoiding adverse side effects.

Virtual screening

Overall, the identification of compounds able to inhibit PTP1B and MMP-13 is expected to provide novel pharmacological alternatives for obesity, type II diabetes mellitus and osteoarthritis. However, randomly testing the inhibitory activity of thousands of compounds in the laboratory for these two targets would be an expensive and

challenging task. Virtual screening can prove useful in this regard, as it allows to extract biologically active compounds from large compound databases with the use of computational tools. In the next section, the most common techniques employed in virtual screening will be introduced.

References

1. WHO: World Health Organization. Available at: <http://www.who.int/>.
2. Abarca-Gómez, L. *et al.* Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* **390**, 2627–2642 (2017).
3. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
4. Chu, D.-T. *et al.* An update on physical health and economic consequences of overweight and obesity. *Diabetes Metab. Syndr. Clin. Res. Rev.* **12**, 1095–1100 (2018).
5. Bliddal, H., Leeds, A. R. & Christensen, R. Osteoarthritis, obesity and weight loss: evidence, hypotheses and horizons - a scoping review. *Obes. Rev.* **15**, 578–86 (2014).
6. Kushner, R. F. Weight Loss Strategies for Treatment of Obesity. *Prog. Cardiovasc. Dis.* **56**, 465–472 (2014).
7. Kakkar, A. K. & Dahiya, N. Drug treatment of obesity: Current status and future prospects. *Eur. J. Intern. Med.* **26**, 89–94 (2015).
8. Hocking, S., Dear, A. & Cowley, M. A. Current and emerging pharmacotherapies for obesity in Australia. *Obes. Res. Clin. Pract.* **11**, 501–521 (2017).
9. Saunders, K. H., Umashanker, D., Igel, L. I., Kumar, R. B. & Aronne, L. J. Obesity Pharmacotherapy. *Med. Clin. North Am.* **102**, 135–148 (2018).
10. Lucas, K. H. & Kaplan-Machlis, B. Orlistat — A Novel Weight Loss Therapy. *Ann. Pharmacother.* **35**, 314–328 (2001).
11. Hancock, J. T. *Cell signalling*. (Oxford University Press, 2017).
12. Bakke, J. & Haj, F. G. Protein-tyrosine phosphatase 1B substrates and metabolic regulation. *Semin. Cell Dev. Biol.* **37**, 58–65 (2015).
13. Johnson, T. O., Ermolieff, J. & Jirousek, M. R. Protein tyrosine phosphatase 1B inhibitors for diabetes. *Nat. Rev. Drug Discov.* **1**, 696–709 (2002).
14. Bandyopadhyay, D. *et al.* Protein-tyrosine phosphatase 1B complexes with the insulin receptor in vivo and is tyrosine-phosphorylated in the presence of insulin. *J. Biol. Chem.* **272**, 1639–45 (1997).
15. Ahmad, F., Li, P. M., Meyerovitch, J. & Goldstein, B. J. Osmotic loading of neutralizing antibodies demonstrates a role for protein-tyrosine phosphatase 1B in negative regulation of the insulin action pathway. *J. Biol. Chem.* **270**, 20503–8 (1995).
16. Goldstein, B. J., Bittner-Kowalczyk, A., White, M. F. & Harbeck, M. Tyrosine dephosphorylation and deactivation of insulin receptor substrate-1 by protein-tyrosine phosphatase 1B. Possible facilitation by the formation of a ternary complex with the Grb2 adaptor protein. *J. Biol. Chem.* **275**, 4283–9 (2000).

17. Elchebly, M. *et al.* Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283**, 1544–8 (1999).
18. Zabolotny, J. M. *et al.* PTP1B regulates leptin signal transduction in vivo. *Dev. Cell* **2**, 489–95 (2002).
19. Cheng, A. *et al.* Attenuation of leptin action and regulation of obesity by protein tyrosine phosphatase 1B. *Dev. Cell* **2**, 497–503 (2002).
20. Combs, A. P. Recent advances in the discovery of competitive protein tyrosine phosphatase 1B inhibitors for the treatment of diabetes, obesity, and cancer. *J. Med. Chem.* **53**, 2333–2344 (2010).
21. Koren, S. & Fantus, I. G. Inhibition of the protein tyrosine phosphatase PTP1B: potential therapy for obesity, insulin resistance and type-2 diabetes mellitus. *Best Pract. Res. Clin. Endocrinol. Metab.* **21**, 621–40 (2007).
22. Alberts, B. *et al.* *Molecular Biology of the Cell.* (W. W. Norton & Company, 2014).
23. Ala-aho, R. & Kähäri, V.-M. Collagenases in cancer. *Biochimie* **87**, 273–86 (2005).
24. Hu, J., Van den Steen, P. E., Sang, Q.-X. a & Opdenakker, G. Matrix metalloproteinase inhibitors as therapy for inflammatory and vascular diseases. *Nat. Rev. Drug Discov.* **6**, 480–498 (2007).
25. Li, H., Wang, D., Yuan, Y. & Min, J. New insights on the MMP-13 regulatory network in the pathogenesis of early osteoarthritis. *Arthritis Res. Ther.* **19**, 248 (2017).
26. Shih, C.-L. M. & Ajuwon, K. M. Inhibition of MMP-13 prevents diet-induced obesity in mice and suppresses adipogenesis in 3T3-L1 preadipocytes. *Mol. Biol. Rep.* **42**, 1225–1232 (2015).
27. Rath, T. *et al.* Enhanced expression of MMP-7 and MMP-13 in inflammatory bowel disease: a precancerous potential? *Inflamm. Bowel Dis.* **12**, 1025–35 (2006).
28. Rabkin, S. W. The Role Matrix Metalloproteinases in the Production of Aortic Aneurysm. *Prog. Mol. Biol. Transl. Sci.* **147**, 239–265 (2017).
29. Solovyeva, N. I., Timoshenko, O. S., Gureeva, T. A. & Kugaevskaya, E. V. Matrix metalloproteinases and their endogenous regulators in squamous cervical carcinoma (review of the own data). *Biomeditsinskaya Khimiya* **61**, 694–704 (2015).
30. Arakaki, P. A., Marques, M. R. & Santos, M. C. L. G. MMP-1 polymorphism and its relationship to pathological processes. *J. Biosci.* **34**, 313–20 (2009).
31. Radosinska, J., Barancik, M. & Vrbjar, N. Heart failure and role of circulating MMP-2 and MMP-9. *Panminerva Med.* **59**, 241–253 (2017).
32. Overall, C. M. & López-Otín, C. Strategies for MMP inhibition in cancer: innovations for the post-trial era. *Nat. Rev. Cancer* **2**, 657–72 (2002).
33. Kurzepa, J. *et al.* Role of MMP-2 and MMP-9 and their natural inhibitors in liver fibrosis, chronic pancreatitis and non-specific inflammatory bowel diseases. *Hepatobiliary Pancreat. Dis. Int.* **13**, 570–579 (2014).

Introduction

34. Romero, A. M., Mastromatteo-Alberga, P., Escalona, L. & Correnti, M. [MMP-3 and MMP-8 levels in patients with chronic periodontitis before and after nonsurgical periodontal therapy]. *Invest. Clin.* **54**, 138–48 (2013).
35. Siloși, I. *et al.* Matrix metalloproteinases (MMP-3 and MMP-9) implication in the pathogenesis of inflammatory bowel disease (IBD). *Rom. J. Morphol. Embryol.* **55**, 1317–24 (2014).
36. Beton, O., Arslan, S., Acar, B., Ozbilum, N. & Berkan, O. Association between MMP-3 and MMP-9 polymorphisms and coronary artery disease. *Biomed. reports* **5**, 709–714 (2016).
37. Huang, H. Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances. *Sensors (Basel)*. **18**, 3249 (2018).
38. Gharib, S. A., Manicone, A. M. & Parks, W. C. Matrix metalloproteinases in emphysema. *Matrix Biol.* **73**, 34–51 (2018).
39. Rodriguez, J. A. *et al.* Metalloproteinases and atherothrombosis: MMP-10 mediates vascular remodeling promoted by inflammatory stimuli. *Front. Biosci.* **13**, 2916–21 (2008).
40. Chelluboina, B. *et al.* MMP-12, a Promising Therapeutic Target for Neurological Diseases. *Mol. Neurobiol.* **55**, 1405–1409 (2018).
41. Placzek, S. *et al.* BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res.* **45**, D380–D388 (2017).
42. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* doi:10.1093/NAR/GKY962
43. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603 (2012).
44. Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).
45. Cathcart, J. M. & Cao, J. MMP Inhibitors: Past, present and future. *Front. Biosci. (Landmark Ed.)* **20**, 1164–78 (2015).
46. Pulkoski-Gross, A. E. Historical Perspective of Matrix Metalloproteases. *Front. Biosci.* **7**, 125–149 (2015).
47. Coussens, L. M., Fingleton, B. & Matrisian, L. M. Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science* **295**, 2387–92 (2002).
48. Alcántara, M. B. & Dass, C. R. Pigment epithelium-derived factor as a natural matrix metalloproteinase inhibitor: a comparison with classical matrix metalloproteinase inhibitors used for cancer treatment. *J. Pharm. Pharmacol.* **66**, 895–902 (2014).
49. Adhikari, N. *et al.* 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. *Bioorg. Med. Chem.* **24**, 4291–4309 (2016).
50. Fabre, B. *et al.* Progress towards water-soluble triazole-based selective MMP-2 inhibitors. *Org. Biomol. Chem.* **11**, 6623 (2013).

51. Takaishi, H., Kimura, T., Dalal, S., Okada, Y. & D'Armiento, J. Joint Diseases and Matrix Metalloproteinases: A Role for MMP-13. *Curr. Pharm. Biotechnol.* **9**, 47–54 (2008).
52. Pottie, P. *et al.* Obesity and osteoarthritis: more complex than predicted! *Ann. Rheum. Dis.* **65**, 1403–5 (2006).
53. Sartori-Cintra, A. R., Aikawa, P. & Cintra, D. E. C. Obesity versus osteoarthritis: beyond the mechanical overload. *Einstein (Sao Paulo)*. **12**, 374–9 (2014).
54. Jiang, L. *et al.* The relationship between body mass index and hip osteoarthritis: A systematic review and meta-analysis. *Jt. Bone Spine* **78**, 150–155 (2011).
55. Hui, W. *et al.* Leptin produced by joint white adipose tissue induces cartilage degradation via upregulation and activation of matrix metalloproteinases. *Ann. Rheum. Dis.* **71**, 455–62 (2012).
56. Pallu, S. *et al.* Obesity affects the chondrocyte responsiveness to leptin in patients with osteoarthritis. *Arthritis Res. Ther.* **12**, R112 (2010).
57. Conde, J. *et al.* Adipokines and osteoarthritis: novel molecules involved in the pathogenesis and progression of disease. *Arthritis* **2011**, 203901 (2011).
58. Berenbaum, F., Eymard, F. & Houard, X. Osteoarthritis, inflammation and obesity. *Curr. Opin. Rheumatol.* **25**, 114–118 (2013).
59. Taylor, N. Nonsurgical Management of Osteoarthritis Knee Pain in the Older Adult: An Update. *Rheum. Dis. Clin. North Am.* **44**, 513–524 (2018).
60. Bartholdy, C. *et al.* Poor replicability of recommended exercise interventions for knee osteoarthritis: a descriptive analysis of evidence informing current guidelines and recommendations. *Osteoarthr. Cartil.* (2018). doi:10.1016/J.JOCA.2018.06.018
61. Edge, R. & Farrah, K. *Exercise for the Management of Knee Osteoarthritis: A Review of Clinical Effectiveness*. *Exercise for the Management of Knee Osteoarthritis: A Review of Clinical Effectiveness* (Canadian Agency for Drugs and Technologies in Health, 2017).
62. Sun, Z., Li, N., Zhang, T., Xin, J. & Ma, X. [Progress of total ankle arthroplasty for end-stage ankle osteoarthritis]. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi* **31**, 1 (2018).
63. Karpinski, K., Müller-Rath, R., Niemeyer, P., Angele, P. & Petersen, W. Subgroups of patients with osteoarthritis and medial meniscus tear or crystal arthropathy benefit from arthroscopic treatment. *Knee Surgery, Sport. Traumatol. Arthrosc.* 1–15 (2018). doi:10.1007/s00167-018-5086-0
64. Fraioli, A. *et al.* Efficacy of Spa Therapy, Mud-Pack Therapy, Balneotherapy, and Mud-Bath Therapy in the Management of Knee Osteoarthritis. A Systematic Review. *Biomed Res. Int.* **2018**, 1–9 (2018).
65. Antonelli, M., Donelli, D. & Fioravanti, A. Effects of balneotherapy and spa therapy on quality of life of patients with knee osteoarthritis: a systematic review and meta-analysis. *Rheumatol. Int.* **38**, 1807–1824 (2018).
66. Zhou, X.-Y. *et al.* Effects of Low-Intensity Pulsed Ultrasound on Knee Osteoarthritis: A Meta-Analysis of Randomized Clinical Trials. *Biomed Res. Int.* **2018**, 1–7 (2018).

Introduction

67. Li, Q., Qi, X. & Zhang, Z. Intra-articular oxygen-ozone versus hyaluronic acid in knee osteoarthritis: A meta-analysis of randomized controlled trials. *Int. J. Surg.* **58**, 3–10 (2018).
68. Maheu, E. *et al.* Why we should definitely include intra-articular hyaluronic acid as a therapeutic option in the management of knee osteoarthritis: Results of an extensive critical literature review. *Semin. Arthritis Rheum.* (2018). doi:10.1016/J.SEMARTHRT.2018.06.002
69. Salmon, J.-H. *et al.* Cost Effectiveness of Intra-Articular Hyaluronic Acid and Disease-Modifying Drugs in Knee Osteoarthritis. *Pharmacoeconomics* **36**, 1321–1331 (2018).
70. Smith, C. *et al.* Combined intra-articular injection of corticosteroid and hyaluronic acid reduces pain compared to hyaluronic acid alone in the treatment of knee osteoarthritis. *Knee Surgery, Sport. Traumatol. Arthrosc.* 1–10 (2018). doi:10.1007/s00167-018-5071-7
71. Nakata, K. *et al.* Disease-modifying effects of COX-2 selective inhibitors and non-selective NSAIDs in osteoarthritis: a systematic review. *Osteoarthr. Cartil.* **26**, 1263–1273 (2018).
72. Haroon, N., Kim, T.-H. & Inman, R. D. NSAIDs and radiographic progression in ankylosing spondylitis Bagging big game with small arms? *Ann. Rheum. Dis.* **71**, 1593–5 (2012).
73. Shah, K., Zhao, A. G. & Sumer, H. New Approaches to Treat Osteoarthritis with Mesenchymal Stem Cells. *Stem Cells Int.* **2018**, 1–9 (2018).
74. Iijima, H., Isho, T., Kuroki, H., Takahashi, M. & Aoyama, T. Effectiveness of mesenchymal stem cells for treating patients with knee osteoarthritis: a meta-analysis toward the establishment of effective regenerative rehabilitation. *NPJ Regen. Med.* **3**, 15 (2018).
75. Lopa, S., Colombini, A., Moretti, M. & de Girolamo, L. Injective mesenchymal stem cell-based treatments for knee osteoarthritis: from mechanisms of action to current clinical evidences. *Knee Surgery, Sport. Traumatol. Arthrosc.* 1–18 (2018). doi:10.1007/s00167-018-5118-9
76. Jayaram, P., Ikpeama, U., Rothenberg, J. B. & Malanga, G. A. Bone Marrow-Derived and Adipose-Derived Mesenchymal Stem Cell Therapy in Primary Knee Osteoarthritis: A Narrative Review. *PM&R* (2018). doi:10.1016/J.PMRJ.2018.06.019
77. Damia, E. *et al.* Adipose-Derived Mesenchymal Stem Cells: Are They a Good Therapeutic Strategy for Osteoarthritis? *Int. J. Mol. Sci.* **19**, 1926 (2018).

Manuscript 1

The light and dark sides of virtual screening: what is there to know?

Aleix Gimeno^[a], María José Ojeda-Montes^[a], Sarah Tomás-Hernández^[a], Adrià Cereto-
Massagué^[a], Raúl Beltrán-Debón^[a], Miquel Mulero^[a], Gerard Pujadas^{[a],[b],*}, Santiago
Garcia-Vallvé^{[a],[b]}

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Abstract

Virtual screening consists in the usage of computational tools in order to extract potentially bioactive compounds from large small-molecule databases. Virtual screening is becoming increasingly popular in the field of drug discovery as *in silico* techniques are continuously being developed, improved and made available. As most of these techniques are easy to use, both private and public organizations resort to virtual screening methodologies in order to save resources in the laboratory. However, it is often the case that the techniques implemented in virtual screening workflows are restricted to those most known by the research team. Moreover, although the software is often easy to use, each methodology has a series of pitfalls that should be avoided so that false results or artifacts are not produced. Here, we will review the most common methodologies used in virtual screening workflows in order to both introduce the inexperienced researcher to new methodologies and advise the experienced researcher on how to prevent common mistakes and the improper usage of virtual screening methodologies.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

1. Virtual screening

Virtual screening (VS) consists in the usage of computational tools in order to extract potentially bioactive compounds from large small-molecule databases. The use of computers allows to process thousands of compounds in a matter of hours while decreasing the cost of experimentally testing the biological activity of all compounds. Like high-throughput screenings, VS protocols are normally used as an early step in the drug discovery process in order to enrich the initial library with active compounds.¹ Thus, they should not be expected to obtain highly potent compounds to be successful (which should be achieved in subsequent hit-to-lead and lead optimization stages), but rather generate a diverse library with a high proportion of active molecules which may be used as starting points for drug design.¹

VS is usually approached hierarchically in the form of a workflow, incorporating methods of different nature in a sequential manner, which act as filters that discard undesirable compounds (see Figure 1). This allows to take advantage of strengths and avoid limitations of individual methods.^{1,2} Compounds that survive all the filters of the VS are usually referred to as *hit compounds* and they need to be tested experimentally in the laboratory to confirm their biological activity. Virtual screening methods can be classified in two major groups: **a)** ligand-based methods, which rely on the similarity of the compounds of interest with active compounds, and **b)** receptor-based methods, which focus on the complementarity of the compounds of interest with the target protein. The most common methodologies belonging to these two categories will be summarized in this review. Table 1 shows a list of popular software used for each methodology.

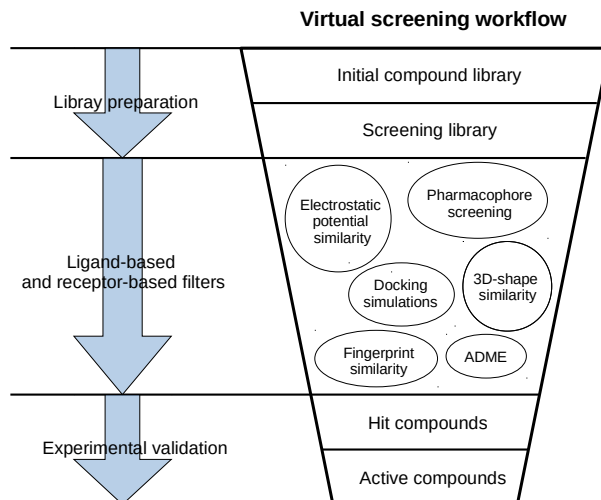


Figure 1. General scheme of a virtual screening workflow

2. First steps

Before developing a VS, a thorough analysis of the data available should be undergone in order to familiarize with the target of interest and determine which methodologies can or can not be incorporated in the VS workflow:

- **Bibliographic research.** First, a bibliographic research on the receptor is recommended, considering aspects such as its biological function, natural ligands and catalytic mechanism, as well as its involvement in pathological processes. This information can be found in databases such as UniProt³ or Brenda.⁴ It is also important to review previous attempts to develop compounds that modulate the activity of the receptor of interest and their mechanisms of action, as well as the current challenges that these compounds face and their limitations. In this regard, the analysis of structure–activity relationship (SAR) studies can provide useful insights on how to design inhibitors for a given target. SAR studies are experiments in which a compound is modified by adding a series of substituents of different nature in one or several parts of the molecule and evaluate the activity of the resulting compounds towards the target

The light and dark sides of virtual screening: what is there to know?

of interest. This provides information on the compound substituents that are preferred by the target in each part of the molecule and therefore reveals which modifications could be applied to a compound to further increase its activity towards the target and which ones should not be applied, as they would result in activity losses. Although in some cases molecular visualization software could aid us in rationalizing the causes of the activity changes observed in SAR studies by analyzing them within the protein environment with the naked eye, there are methods that help us to better understand the nature of the interactions between the ligand and the protein. This is for example the case of Flare,⁵ in which the electrostatic potentials and hydrophobicity of both the ligand and the protein can be represented and compared to determine whether the introduction of a particular functional group is favorable for activity or not. While this type of information may be important to establish an appropriate VS strategy, it could also be crucial in the final steps of the VS to: **a)** select hits that present features which have previously been reported to be important for activity; and **b)** avoid selecting compounds that would perform unfavorable interactions with the target that can have possibly been overlooked by earlier steps of the VS workflow.

- **Activity and structural data collection.** On the one hand, activity data of previously reported inhibitors as well as their structure should be retrieved from databases such as ChEMBL,⁶ Reaxys,⁷ BindingDB⁸ or PubChem,⁹ as a high amount and structure variability of compounds will result in an improved performance of the developed ligand-based models and a more representative computational validation of the used VS methods.¹ On the other hand, it is important to determine whether the 3D structure of the receptor has been elucidated and, if that is the case, the available quantity and quality of crystallographic structures. A careful inspection of these structures will allow us to assess the flexibility of the receptor and evaluate whether receptor-based approaches can be implemented to the VS. Crystallographic models of protein and protein-ligand complexes can be obtained from the PDB^{10,11} database, but it should be kept in mind that the atoms in these models correspond to the representation of the electron density maps created by the crystallographer and they are in some cases susceptible to interpretations. Therefore, the validation of

Manuscript 1

crystallized structures with specialized visualization software such as VHELIBS¹² is recommended before the use of crystal structures in VS.

- **Library preparation.** The collection of compounds to which the VS workflow will be applied is referred to as the *VS library*. Apart from retrieving the structures and activities of compounds with known activity for VS development and validation purposes, the structures of compounds to which the VS should be applied also need to be obtained, generating the so-called virtual screening library. This library of compounds can either proceed from an in-house collection of compounds of interest, it can be obtained either from different databases, such as ZINC¹³ or Reaxys,⁷ or directly from a compound supplier. In many cases, the structures of these compounds are collected in 2D format, but many VS methods require the 3D conformation of compounds (i.e. the arrangement of atoms of a molecule in space). Thus, the 3D conformations that molecules adopt usually need to be predicted in a process known as conformational sampling, in which conformers are first generated by determining bond-lengths, bond angles and torsion angles and then ranked to prioritize the low energy conformations that are accessible with a reasonable probability at room temperature.¹⁴ Software commonly used to perform these tasks are shown in Table 1. In a recent benchmarking of conformer generator ensembles,¹⁵ commercial ensemble generators like OMEGA¹⁶ or ConfGen¹⁷ showed a high performance, closely followed by the freely available implementation of the Distance Geometry¹⁸ algorithm by RDKit,¹⁹ which also showed a high robustness. Generating a sufficiently wide set of conformations for each compound is crucial to cover the compound's conformational space and achieve optimal results in many VS methodologies that depend on the 3D conformation of compounds (e.g. 3D fingerprints, 3D-shape comparison, electrostatic potential comparison, protein-ligand docking, pharmacophore screening). Otherwise, if the bioactive conformation of interest is not included among the conformers generated for a given compound, this supposes a limitation for subsequent methods.¹ On the other hand, the generation of high energy conformations with a low probability of being accessed by the molecule at room temperature should be avoided as they may be misleading and cause false positive results.¹ In addition to the spatial distribution of

The light and dark sides of virtual screening: what is there to know?

atoms, other aspects need to be considered when preparing molecules for VS purposes. Many VS methodologies are dependant on the charge of molecules (e.g. electrostatic potential comparison, protein-ligand docking, pharmacophore screening), so it is important to ensure that the charges of compounds are properly defined as they may not be present or they may not be assigned correctly. The different possible protonation states at the pH of interest also need to be generated for each molecule, as well as its tautomeric states.¹ In addition, other aspects such as stereochemistry and the presence of salt and solvent fragments also need to be considered during molecule preparation. Software such as Standardizer²⁰ or LigPrep²¹ and tools such as MolVS²² can be used for this purpose.

Table 1. Popular and useful software in virtual screening.

Method	Software	Developer
Graphical user interface	Flare ⁵	Cresset
	Maestro ³⁶	Schrödinger
	VIDA ³⁷	OpenEye
Validation set preparation	DecoyFinder ³⁸	Universitat Rovira I Virgili
Crystal structure validation	VHELIBS ¹²	Universitat Rovira I Virgili
Molecule standardization	Standardizer ²⁰	ChemAxon
	LigPrep ²¹	Schrödinger
	MolVS ²²	RDKit
Conformer generation	OMEGA ¹⁶	OpenEye
	ConfGen ¹⁷	Schrödinger
	Distance Geometry (DG) ¹⁸	RDKit
	ETKDG ²³	RDKit
ADME property prediction	QikProp ²⁴	Schrödinger
	SwissADME ²⁵	Swiss Institute of Bioinformatics
	FAFDrugs4 ²⁶	ChemAxon
Shape similarity	ROCS ²⁷	OpenEye
	Shape screening ²⁸	Schrödinger
Electrostatic potential similarity	EON ²⁹	OpenEye
Pharmacophore	Phase ³⁰	Schrödinger
	Ligandscout ³¹	Inte:Ligand GmbH

Manuscript 1

	Glide ³²	Schrödinger
Docking	GOLD ³³	The Cambridge Crystallographic Data Centre
	DOCK ³⁴	University of California San Francisco
	Autodock ³⁵	The Scripps Research Institute

3. Ligand-based virtual screening

Ligand-based VS methods measure the similarity of the compounds in the library with reference compounds which are active towards a target of interest or present desired properties. The basis for these methods lies in the *similar property principle* introduced by Johnson and Maggiora, which states that similar compounds have similar properties.³⁹ Thus, compounds with high similarity to reference compounds are likely to behave in a similar fashion or act through the same mechanism as those and therefore have similar effects. Similarity is a subjective concept, and different methods use different similarity measures to determine the similarity of two compounds. In this section, the most common ligand-based methods used in virtual screening will be summarized. Ligand-based VS methods are relatively cheap computationally compared to receptor-based methods, as no macromolecules are involved in the calculations. For this reason, they are used early in the VS workflow process when the amount of compounds in the starting library is the highest.

3.1. Fingerprint-Based Methods

Molecular fingerprints constitute a ligand-based method for similarity searching which uses patterns in the structure of compounds in order to compare them. Fingerprints are sequences of bits, in which each one includes certain information regarding a molecule (see Figure 2).⁴⁰ As these bits are quantifiable, this allows to draw comparisons between two molecules (A and B) and determine their similarity. According to nature of bits, fingerprints can be classified as:

The light and dark sides of virtual screening: what is there to know?

- **Substructure keys-based fingerprints.** In these type of fingerprints, the bit string corresponds to a series of predefined structural keys and each bit relates to the presence or the absence of a given feature in the molecule. Therefore, these fingerprints are effective when the structural keys used by the fingerprint are present in the molecules to be compared, but they are not that meaningful otherwise. Examples of these type of fingerprints include MACCS^{41,42}, PubChem fingerprints⁹ or BCI fingerprints.⁴³
- **Topological or path-based fingerprints.** In these types of fingerprints, bits are defined from fragments of the molecules themselves. For every atom in a molecule, fragments are obtained by progressively increasing the length up to a determined number of bonds usually following a linear path. Then, these fragments are hashed to generate the fingerprint. As fingerprints are generated from the molecules themselves, every molecule produces a meaningful fingerprint, and its length can be adjusted. However, in topological fingerprints with a reduced number of bits, bit collisions may occur as a result of assigning more than one different feature to a given bit. Examples of these type of fingerprints include the Daylight fingerprint⁴⁴ and OpenEye's Tree fingerprints.⁴⁵
- **Circular fingerprints.** In these type of fingerprints, bits are also defined from molecule fragments, but these fragments are obtained from the environment of each atom up to a determined radius instead of a path. Examples of these type of fingerprints include Molprint2D,^{46,47} extended-connectivity fingerprints (ECFP) and functional-class fingerprints (FCFP).⁴⁸
- **Pharmacophore fingerprints.** These fingerprints incorporate the features of the molecule in the fingerprint. The distances between features are considered in order to encode 3D information into the fingerprint.⁴⁹

Apart from the fingerprint types listed above, other fingerprint types exist, such as fingerprints based on molecule SMILES⁵⁰ or protein-ligand interaction fingerprints, which encode information regarding the type of interactions between the protein and

Manuscript 1

the ligand.^{51,52} Moreover, fingerprints can also derive from a combination of the approaches mentioned above, constituting what are known as hybrid fingerprints.⁵³

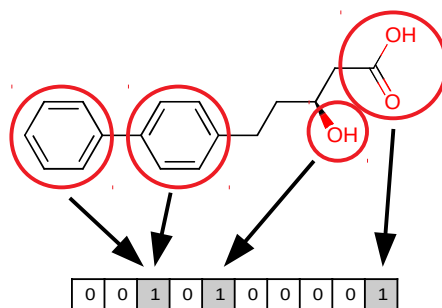


Figure 2. Illustration of how fingerprint bits are derived from the structure of a molecule.

Once the fingerprints for each molecule have been calculated, different methods can be used to enrich a library in active compounds through the use of fingerprints:

a) Similarity metrics

Several similarity metrics can be used to compare the fingerprints of two compounds (e.g. Euclidean distance, Manhattan distance or Sørensen–Dice coefficient), but the most popular one is the Tanimoto coefficient.⁵⁴ The Tanimoto coefficient is a value in a range from 0 to 1 that represents the similarity between two compounds based on the fingerprint bits they match (the higher the value, the more similar the compounds). It is expressed by:

$$\text{Tanimoto coefficient} = \frac{c}{(a+b+c)}$$

where given the fingerprints of compounds *A* and *B*, *a* equals the amount of bits set to 1 in *A*, *b* equals the amount of bits set to 1 in *B* and *c* equals the amount of bits set to 1 in both *A* and *B*.

With the help of these similarity metrics, the fingerprints of the compounds in a library can be compared to the fingerprints of active compounds for the target of interest, thus assessing the similarity between them. Next, in order to select the compounds in the library with a higher probability of being active, compounds can be sorted according to the Tanimoto coefficient or other similarity metrics and a cutoff can be applied, thus keeping the compounds with a higher similarity to known actives, while discarding

The light and dark sides of virtual screening: what is there to know?

compounds with a lower similarity to known actives. While this approach is easy to use and well founded, it has some limitations:

- A high similarity coefficient value does not always imply that two compounds will have the same activity. Some minor structural changes could greatly modulate the activity of a compound depending on how they affected the interactions with the protein. These great changes in activity due to small changes in compound structure are commonly known as activity cliffs (see section 4) and they constitute the major reason why activity values can not always be inferred from similarity measures. Thus, the comparison of similarity coefficients should be seen as a simplification attempt that bypasses medicinal chemistry in order to establish an automated approximation of the activity of compounds.^{40,55}
- There is not a universal cutoff value to determine that compounds with a determined range of similarity to reference compounds will have similar activity values. As fingerprints are designed in different ways, if we compare the similarities between two compounds with a given similarity coefficient using two different types of fingerprints, the similarity values obtained will most likely differ. A similar situation will occur when comparing two compounds using the same fingerprint, but a different similarity metric.^{40,55} Therefore, as the similarity value between two compounds is affected by both the type of fingerprint and the similarity coefficient used, the optimal cutoff will also be dependent on these two factors and will have to be evaluated on a case by case basis using the adequate statistical measures (see section 5). Moreover, different types of fingerprints perform differently in different situations,¹ so the fingerprint results obtained should be validated computationally in order to choose the appropriate fingerprint (see section 5).
- Similarity coefficients assign equal importance to all fingerprint bits. This supposes a limitation in the following two situations: **a)** on the one hand, compounds that do not possess the critical features for activity that are present in the active compounds used as reference but still accomplish good similarity values by matching most of the fingerprint bits will be wrongly predicted to be active; and **b)** on the other hand, compounds that only match the critical features for activity with the active compounds

Manuscript 1

used as reference and are structurally different from them will be wrongly predicted to be inactive.^{1,40,55}

Nevertheless, as fingerprint-based similarity approaches display a virtual screening performance similar to other more complex methods while being computationally cheaper, they still are the preferred choice in many VS approaches.⁴⁰

b) Supervised machine learning

Another common method for incorporating the information encoded in fingerprints is the use of supervised machine learning methods. These consist on different algorithms that, given a sample of the fingerprints of compounds with different activities, are to obtain a model that relates the fingerprint to the observed activity value of the compound in order to apply the model to a new set of compounds and predict their activity based on their fingerprints. Some examples of supervised machine learning methods include random forest, support vector machines, naive Bayes, *k*-nearest neighbors or artificial neural networks.

Supervised learning can be divided into classification and regression tasks, depending on the desired output. In classification tasks the objective is to identify to which class a particular input belongs to (discrete output). In this case, an arbitrary threshold can be established to divide compounds into *active* and *inactive* and the new compounds will be classified in one of these two categories based on their fingerprints. In regression tasks, the objective is to assign a continuous output value from the input. In this case, the model will be asked to predict the corresponding activity value of a compound with a given fingerprint.

In order to build and validate the model, the input data (in this case, compounds with known activity and their fingerprints) needs to be divided into training and test sets, both sets containing active and inactive compounds. While the training set is used to build the model, the test set is used to validate it and evaluate its performance. The different metrics to evaluate model performance are discussed in section 5. Contrary to typical similarity approaches, which are based on the overall similarity of compounds placing equal importance to all parts of the molecule,¹ machine learning methods

The light and dark sides of virtual screening: what is there to know?

circumvent this problem as they are built from multiple active molecules and fingerprint bits are related independently to the bioactivity of compounds. Thus, they are able to recognize the fingerprint bits that are critical for bioactivity. This constitutes their major strength compared to other ligand-based methods.

3.2. 3D-shape similarity

In a ligand-receptor interaction, the shape of the ligand is crucial as the ligand needs to fit in the binding pocket of the receptor to establish key interactions for the binding to occur. The basis of 3D-shape similarity lies in the fact that two molecules with similar shape are likely to fit in the same binding pocket and thereby exhibit similar biological activity.⁵⁶ In this approach, the 3D shape of the compounds in the VS library is compared to the 3D shape of known active compounds, which are used as a reference. Despite being based on similarity, different to other ligand-based methods, this method does not take into account the particular structure or properties of the reference ligands and only relies on the shape of the molecules. This constitutes its major advantage as it makes it a suitable method to identify new scaffolds that may overlap well with known ligands which may be active towards the target of interest (see Figure 3). As new chemotypes are pursued in medicinal chemistry to expand the horizon, structure novelty is highly valued in virtual screening and 3D shape-based similarity analysis is gaining attention in virtual screening campaigns nowadays.⁵⁶ Software commonly used to perform 3D shape-based similarity comparisons are shown in Table 1.

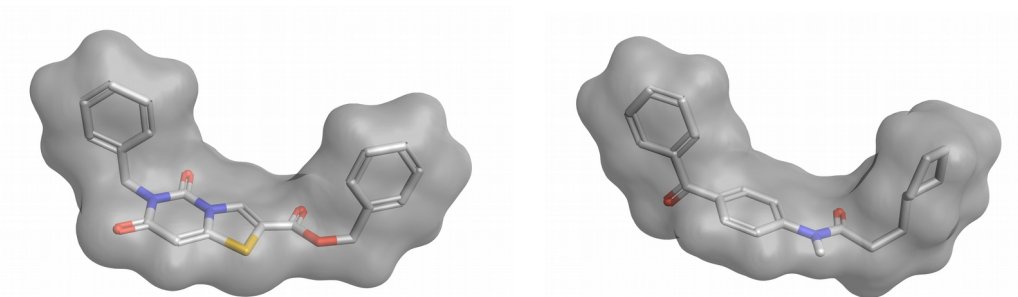


Figure 3. Example of two compounds with a different molecular structure but a high 3D shape similarity. This figure was obtained with Flare.⁵

Shape comparison methods can be classified in two major categories:

- **Alignment-free or non-superposition methods.** As these methods are independent of the position and orientations of molecules, they are much faster and could be used to screen large compound databases
- **Alignment or superposition-based methods.** These methods require a superposition between the reference compounds and the compounds in the database. Although they are highly effective, they are computationally expensive and a sub-optimal alignment may lead to errors in comparing the molecules. These methods allow the visualization of the alignment together with their similarity values, which can aid in the design of new molecules and guide in their further optimization. As an alignment is performed, these methods can also include comparisons of surface properties such as hydrophobicity and polarity.

Several methods exist to attain the common end of evaluating shape similarity. Here is a description of the most commonly used:

- **Atomic Distance-Based Shape Similarity Methods.** As the shape of a molecule can be described by the relative positions of its atoms, these methods rely on the computation and comparison of interatomic distance descriptors to determine shape similarity. These methods do not require the alignment of the molecules involved and therefore are faster than alignment-based methods.

- **Volume-Based Shape Similarity Methods:** Two molecules will possess similar shape if they have a similar volume. Therefore, shape similarity can be described in terms of volume occupancy. The most widely adopted models to describe shape similarity in terms of volume are the hard sphere model^{57,58} and the Gaussian sphere model.^{59,60} The hard sphere model treats each atom in the molecule as a sphere, and the volume of each molecule is calculated based on the unions and intersections of their volumes. The Gaussian sphere model represents a molecule as a set of overlapping Gaussian spheres. The inclusion-exclusion principle is applied to obtain the volume of the molecule by calculating the volume of all Gaussians and their intersections.
- **Surface Based Shape Similarity Methods:** Shape similarity can also be analyzed by comparing the molecular surfaces of two molecules. Some surface definitions commonly used for this purpose include the solvent-accessible surface⁶¹ and the van der Waals surface.^{62,63}

3.3. Electrostatic potential similarity

Another method to measure the similarity between two compounds is to compare their electrostatic potentials. Electrostatic interactions often play a critical role in the binding of the ligand as the target presents a particular electrostatic environment that must be matched by the ligand in order for the binding to occur. Thus, using the electrostatic potential of the ligand as a reference, we can obtain compounds that present a similar electrostatic distribution and could potentially match the electrostatic environment of the target, therefore being candidates of having an action on that target. Software commonly used for electrostatic potential comparison is shown in Table 1.

Although structurally similar compounds are likely to have a similar electrostatic potential, some small changes in structure can have a great impact on the electrostatic distribution of the compound (see Figure 4A). Therefore, pairing this methodology with 2D fingerprints or 3D-shape analysis should result in a reduction of false positives. More interestingly, compounds with a completely different 2D structure can present similar electrostatic potentials (see Figure 4B). This allows to search for novel inhibitors

with the same electrostatic properties as known ligands but with different structure that should be able to bind to the target of interest.

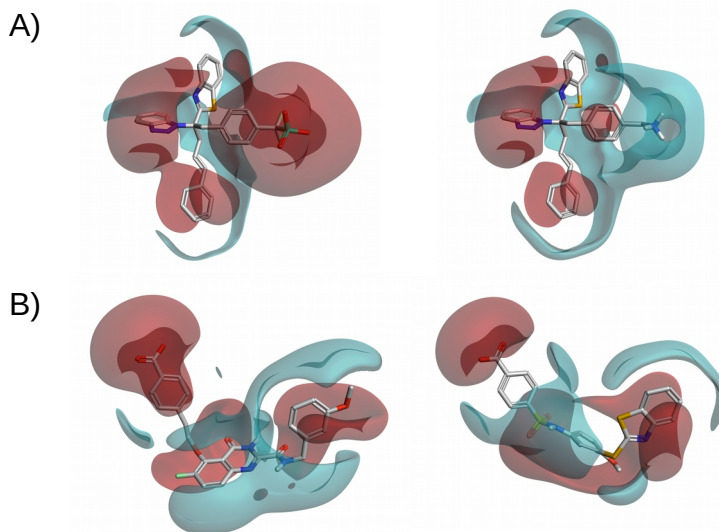


Figure 4. Electrostatic potential comparisons of two pairs of molecules. Panel A shows two structurally similar compounds with a different electrostatic potential. Panel B shows two compounds with a different molecular structure but a high electrostatic potential similarity. This figure was obtained with Flare.⁵

3.4. Ligand-based pharmacophores

According to the IUPAC, a pharmacophore is “*the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response.*”⁶⁴ Pharmacophores which are obtained from one or a set of ligands are called ligand-based pharmacophores. These incorporate the common features throughout a set of ligands that present bioactivity towards a common target. It is then assumed that these features are responsible for the activity of the ligand and the pharmacophore is used to search for other compounds presenting the same distribution of features in the VS library (see Figure 5). As the compounds that match the pharmacophore contain the same features as known ligands, they are expected to perform the same interactions with the biological target and their binding is expected to result in the same biological response. Software commonly used in pharmacophore-based virtual screenings is

The light and dark sides of virtual screening: what is there to know?

shown in Table 1. The generation of a ligand-based pharmacophore consists of several steps:

- 1. Select a set of active ligands to generate the pharmacophore.** First a set of active ligands is obtained, from which the pharmacophore features will be generated, usually based on the common features of the ligands. Ligands that are not active for the target of interest may also be used to discard pharmacophore hypotheses.
- 2. Generate 3D conformations of the ligands.** A determined number low energy conformations is generated for each bioactive compound so that the conformation in which the ligand binds to the receptor is likely to be included.
- 3. Identify ligand features.** The substructures and functional groups of the ligand are transformed to pharmacophoric features. These features often include: **a)** hydrogen bond donor feature, **b)** hydrogen bond acceptor feature, **c)** negative charge feature, **d)** positive charge feature, **e)** hydrophobic feature and **f)** aromatic ring feature. The distances between the features of the ligand are computed and the combination of features and distances is used as an abstract representation of the ligand.
- 4. Superimpose ligands.** Ligand representations are superimposed so that a maximum number of features occupy the same regions of space.
- 5. Generate pharmacophore.** Features that occupy the same region and are present in the majority of ligands are included in the pharmacophore.
- 6. Validation.** The pharmacophore needs to be validated to ensure that it is able to discriminate active compounds from inactive compounds. This is usually done by screening a set of actives and a set of decoys, which are compounds similar to actives that do not present bioactivity for the target of interest (see section 5).

Manuscript 1

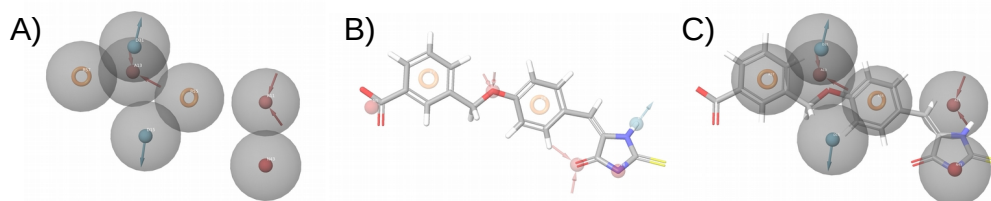


Figure 5. Pharmacophore model and fitting compound. Panel A shows the pharmacophore model. Panel B shows the compound and its pharmacophoric features. Panel C shows a superposition of the compound and the pharmacophore, showing that the compound matches 4 sites in the pharmacophore. Hydrogen bond acceptor, hydrogen bond donor, aromatic and negative ionizable features are shown in red, blue and orange and red, respectively. The arrows in hydrogen bond acceptor and hydrogen bond donor features indicate the direction of the hydrogen bond. This figure was obtained with Maestro.³⁶

As the definition of pharmacophoric features is arbitrary (the user determines the number of pharmacophore features and their tolerances), it is important to correctly evaluate the strictness of the pharmacophore model. While a very strict model should lead to better activity results but poor structural diversity, a very fuzzy model is more likely to retrieve a larger number of false positives but achieve a higher structural diversity. Therefore, an adequate trade-off should be found between strict and loose criteria (see section 5).¹ This can be achieved by prioritizing features that show a better performance or that are associated with higher compound activity. For instance, it is possible to develop and evaluate the performance of a particular pharmacophore with and without the presence of certain features or adjusting their tolerances in order to determine the importance of each feature for the performance of the model. In order to prioritize pharmacophoric features relevant for compound activity, information obtained from SAR studies regarding ligand-receptor interactions critical for activity may be implemented in the pharmacophore model.

Although a pharmacophore model can be developed from a single or few active molecules, it is recommended to use a rather large set of actives to develop the model based on their common features,¹ because using one or few ligands that do not present certain features relevant for activity as the only reference compound/s may lead to missing out certain features important for bioactivity and a lower

enrichment in active compounds after applying the pharmacophore to the VS library.

4. Receptor-based virtual screening

Previously we have seen how ligand-based VS methods take advantage of the *similar property principle* in order to identify active compounds towards a particular target. Nevertheless, although similar compounds often present similar activities, this is not always the case, as some compound modifications may be prejudicial for the ligand-target interaction and therefore result in a loss of activity for the target of interest. Thus, this can lead to erroneous predictions if the *similar property principle* is applied to determine the activity of the new compound.¹ For instance, if a negatively charged carboxylic acid group is introduced in a region of the molecule that is close to an acidic residue of the target protein, such as Glu or Asp, even though the new compound and the original compound will be structurally similar as they present the same substructure, this will most likely result in a loss of activity of the compound for that target due to the electrostatic repulsion of the negative charges in the new compound and the protein. These situations in which a small modification of the compound results in a drastic change of activity are known as activity cliffs. To avoid these type of incompatibilities between a compound and the receptor, it is important to take the receptor into account. Nevertheless, as the receptor is a macromolecule, more information needs to be processed and, thus, receptor-based methods are more computationally expensive than ligand-based methods.

4.1. Protein-ligand docking

Protein-ligand docking is the most popular structure-based technique in virtual screening.⁶⁵ This methodology uses the crystallized structure of a protein to predict how the compounds in the VS library would bind to the binding site (see Figure 6). The different compound orientations in the binding site generated by docking are referred to

Manuscript 1

as *docked poses*. Software commonly used to perform protein-ligand docking are shown in Table 1. Protein-ligand docking consists of the following steps:

1. Protein preparation. As experimental structures, X-ray crystallographic protein structures present problems such as missing hydrogen atoms, missing residues, incomplete side chains, undefined protonation states or the presence of crystallization products that are not found *in vivo*. These aspects need to be corrected before the crystallographic structure can be used to perform protein-ligand docking.

2. Grid generation. A grid is defined around the area of the protein where the ligand is expected to bind. Docked poses will be restricted to the space occupied by the grid.

3. Conformational sampling. A search algorithm is responsible for identifying the possible conformations (docked poses) in which each compound may fit in the grid. Constraints can be defined during docking to require the resulting docked poses to bind to a certain region of the binding site or to perform a certain interaction with the receptor.

4. Scoring. Finally, the affinity of each docked pose for the target is approximated with a scoring function that predicts the strength of the interaction, generating a score for each docked pose. Then, the docked poses are ranked according to the score provided by the docking function to obtain the docked pose that is most likely to represent the real binding mode of the compound.

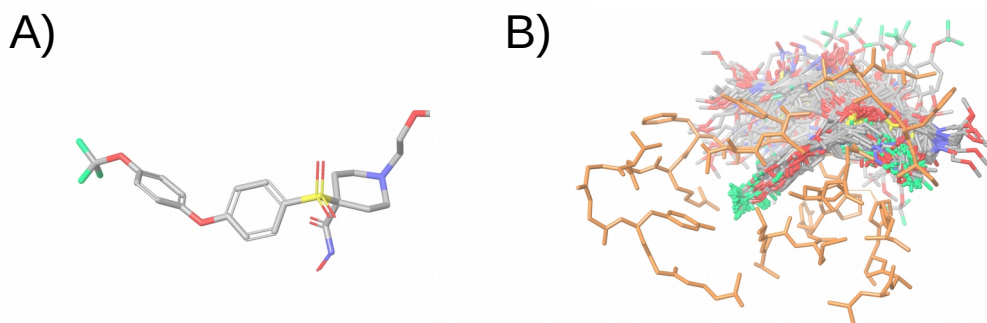


Figure 6. Illustration of the docking simulation of a compound. Panel A shows the molecular structure of the compound. Panel B shows the docked poses obtained after docking the compound in the binding site of the

The light and dark sides of virtual screening: what is there to know?

protein. The compound is colored in the CPK color scheme and the protein is colored in orange. Non-polar hydrogen atoms have been omitted in the representation. This figure was obtained with Maestro.³⁶

As docking is one of the most popular and available VS methods, docking results are often misinterpreted by inexperienced researchers for the following reasons:

- Although the search algorithm provides potential orientations of the compound in the binding site, this does not imply that the real binding mode of the compound is among the docking poses, as the search algorithm can fail to predict it. Thus, docking should be seen as a means of generating hypotheses on how the compound may bind in the binding site of the target (i.e. docked poses), but not as definite proof that the compound binds in a determined fashion.^{1,66} In order to determine the binding mode of a compound, the experimental 3D structure of the complex formed between the protein and the compound needs to be obtained with experimental methods, such as X-ray crystallography or nuclear magnetic resonance.
- Docking software are often wrongly used to predict the activity of compounds based on the score provided by docking functions. Although this may be possible in some cases for structurally similar compounds, scoring functions are not accurate enough to predict the binding affinity of compounds that have different structures and different predicted binding modes. The low success rate of scoring functions at predicting the binding affinity of compounds should be taken into account when performing a docking simulation.^{1,66} Instead of aiming at predicting compound activity, protein-ligand docking should be used to enrich the initial library in active compounds by discarding compounds that are not able to fit in the binding site of the protein and by keeping the compounds that are more likely to show a good binding affinity as predicted by the scoring function. The latter can be achieved, for instance, by establishing a docking score threshold.
- The flexibility of the protein can be accounted for in docking procedures in an approach known as flexible docking. However, this is often not the case in virtual screening, as it would imply an added computational cost, and the receptor atoms are usually not allowed to change their spatial location. This approach is known as rigid

Manuscript 1

docking and it is important to respect its limitations.¹ Therefore, in proteins with a flexible binding site able to accommodate very diverse ligands, this may not be the right approach and allowing the movement of some protein residues may be considered. A possible workaround to account for the flexibility of the protein while using a rigid docking approach may be to use all the available receptor conformations for docking.

4.2. Structure-based pharmacophores

Pharmacophores can also be obtained taking into account the receptor. Although the process is similar to a ligand-based pharmacophore, the main difference is the way of obtaining the distribution of features. Instead of obtaining the features through the alignment of ligand conformations, the features that will constitute the pharmacophore can be obtained by one of these processes:

- a)** Using conformations of ligands that are co-crystallized with the receptor. In this case, the pharmacophore features are also obtained from active compounds, but as they are co-crystallized with the receptor, their bioactive conformations are already known and there is no need to generate conformations.
- b)** Using the residues in the binding site to determine pharmacophore features. In this case, the pharmacophore features are not obtained from the ligands, but from the receptor. This method allows the comparison of pharmacophores obtained from different crystallographic structures of the same target.
- c)** Docking fragments into the binding site of the receptor. In this case, the docking of a fragment library is performed, the docked fragments are clustered, and pharmacophore features are generated from these clusters. The nature of the interaction of the cluster of fragments determines the type of feature and the docking scores of the fragments in the cluster determine the relevance of the feature. This method allows probe the binding site and identify interactions not considered in the design of previous inhibitors which could potentially be important

for activity. An interesting utility to evaluate the potential activity contribution of each pharmacophoric feature in a fragment-based pharmacophore is the E-pharmacophores⁶⁷ utility from Schrödinger, which assigns an energetic contribution to each pharmacophore feature based on the docking scores of the fragments.

An added advantage of receptor-based pharmacophores is that the receptor cavity can be considered in order to directly discard compound conformations that would not fit in the binding site. This can be done by introducing excluded volumes in the regions occupied by the protein atoms that can not be occupied by the compound as they would result in steric impediments with the protein and obstruct the binding.

5. Computational validation

Because virtual screening workflows consist of a series of computational methodologies, ultimately virtual screening hits are predictions which need to be validated both *in silico* and *in vitro* or *in vivo* to prove their correctness. *In silico* validation is often performed by applying the virtual screening parallelly to a set of active compounds and a set of inactive or decoy compounds:

- **Actives.** Active compounds (or simply actives) are compounds which have been reported to present a high activity towards the target of interest. Compounds used as a reference to generate the virtual screening fall into this category. The exact activity threshold over which a compound is considered to be active is arbitrary, but compounds are often considered as actives if they have IC_{50} , K_i or EC_{50} activity values around the micromolar and nanomolar range. The higher the threshold selected, the more restrictive the virtual screening. It should be taken into account that even if a compound has been reported to have a certain activity for the target of interest, the mechanism of action may differ from the mechanism of action sought by the VS (e.g. the compound may exert its action by binding to an allosteric site of the protein instead of the catalytic site). The VS should not be able to identify these compounds as actives

Manuscript 1

as their binding mode is different and this would result in a decrease of performance of the VS. Therefore, active compounds with unreported protein binding modes represent intrinsic limitations of the VS validation.¹

- **Inactives.** Inactive compounds (or simply inactives) are compounds which have been reported to present a low activity towards the target of interest. Hit compounds that are similar to inactives are considered to have a higher chance of being inactive and therefore should be avoided. Analogously to actives, an arbitrary activity threshold under which compounds are believed to be inactive should be predefined. The higher the threshold, the more demanding the virtual screening, as it will be asked to discern between active and inactive compounds with higher accuracy due to the lower difference of activity between the two groups. PubChem⁹ and ChEMBL⁶ are databases of chemical compounds that include inactive compounds.
- **Decoys.** Decoy compounds (or simply decoys) are compounds that resemble active compounds but for which the activity towards the target of interest has not been reported and, since they are likely to be inactive, they are presumed to be so.³⁸ Decoys are typically obtained by searching compounds that have similar physical descriptors (e.g. molecular weight, number of rotational bonds, total hydrogen bond donors, total hydrogen bond acceptors and octanol–water partition coefficient) to active compounds, but that are chemically different from them (which can be determined by fingerprint similarity).³⁸ In validation protocols, decoys are usually used instead of inactives due to the low amount of null results reported in the literature and the consequent lack of data on inactive compounds. Similarly to what occurs with active compounds that act through different modes of action than the one assessed by the VS, as decoys are putatively inactive compounds but their activity for the target of interest has not been determined, a small portion of them may actually be active, and therefore this also constitutes an intrinsic limitation to assess VS performance.¹ Decoys can be obtained either directly from databases such as DUD-E⁶⁸ or through the use of tools such as DecoyFinder,³⁸ which allows to obtain sets of decoys that match the provided sets of active compounds.

The light and dark sides of virtual screening: what is there to know?

Because each VS step essentially behaves as a binary classifier which labels the output compounds as active (positives) or inactive (negatives), once the active and inactive/decoy groups have been defined and the VS has been applied, each compound will fall into one of the following four classes:

- **True positives.** Active compounds which are predicted to be active.
- **True negatives.** Inactive compounds which are predicted to be inactive.
- **False positives.** Inactive compounds which are predicted to be active.
- **False negatives.** Active compounds which are predicted to be inactive.

Table 2. Confusion matrix of a binary classifier.

		True condition	
		Positive	Negative
Predicted condition	Positive	True positives (TP)	False positives (FP)
	Negative	False negatives (FN)	True negatives (TN)

These classes conform the so-called confusion matrix of the binary classifier (see Table 2) and they can be used to calculate a series of statistical measures which can in turn be used to assess the performance of each VS step:

- **Sensitivity.** Also referred to as recall, hit rate or true positive rate (TPR), it measures the proportion of actual positives that are correctly identified as such:

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR$$

Manuscript 1

- **Specificity.** Also referred to as selectivity or true negative rate (TNR), it measures the proportion of actual negatives that are correctly identified as such:

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} = 1 - FPR$$

- **Precision.** Also referred to as positive predictive value (PPV), it measures the proportion of positive results that correspond to actual positives:

$$PPV = \frac{TP}{TP+FP}$$

- **Negative predictive value (NPV).** It measures the proportion of negative results that correspond to actual negatives:

$$NPV = \frac{TN}{TN+FN}$$

- **False negative rate (FNR).** Also referred to as miss rate, it measures the proportion of actual positives incorrectly classified as such (it complements sensitivity):

$$FNR = \frac{FN}{P} = \frac{FN}{FN+TP} = 1 - TPR$$

- **Fall-out.** Also referred to as false positive rate (FPR), it measures the proportion of actual negatives that are incorrectly classified as such (it complements specificity):

$$FPR = \frac{FP}{N} = \frac{FP}{FP+TN} = 1 - TNR$$

- **Accuracy.** It measures the proportion of correctly predicted results among the total number of cases examined:

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$$

The light and dark sides of virtual screening: what is there to know?

- **F1 score.** It is a measure obtained from the harmonic mean of the sensitivity and the precision statistical measures. It considers both measures in order to determine the performance of the classifier:

$$F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

- **Matthews correlation coefficient (MCC).** It is a correlation coefficient between the observed and predicted binary classifications that returns a value between -1 and +1. A coefficient of +1 represents a perfect prediction, a coefficient of 0 indicates that the prediction is no better than a random prediction and a coefficient of -1 indicates total disagreement between prediction and observation. The MCC is generally regarded as a balanced measure that can be used even if the classes are of very different sizes and it is calculated using the following formula:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Apart from these statistical measures, other methods are commonly used to assess the performance of binary classifiers:

- **Enrichment factor (EF).** The EF is a measure of how much the sample is enriched with actives after applying a determined filter or a series of filters. It is calculated as the ratio between the proportion of actives after and before the VS step.

$$EF = \frac{\frac{a_2}{a_2 + d_2}}{\frac{a_1}{a_1 + d_1}}$$

in which:

a_1 = actives before the VS step

a_2 = actives after the VS step

d_1 = decoys before the VS step

d_2 = decoys after the VS step

Manuscript 1

With the help of these measures, we are able to evaluate not only the overall performance of the model, but also other characteristics such as its ability to retrieve actives or discard inactives. Depending on the priorities of the particular step of the virtual screening in question, a determined measure can be prioritized over another in order to select the adequate model in each situation. For instance, at the beginning of the VS, priority may be given to discarding inactive compounds as the number of inactive compounds in the initial library is higher. In this case, specificity would be prioritized over sensitivity. On the other hand, in latter stages of the VS workflow, priority may be given to the retrieval of active compounds, as the proportion of active compounds should be higher and compounds that are predicted to be active are also expected to be active based on previous filters. In this case, sensitivity would be prioritized over specificity.

Based on these statistical measures, model parameters can be tweaked until a more satisfactory result is achieved. For instance, if after applying a determined workflow filter, the resulting number of compounds was considered to be too high, the parameters of that filter could be set to more restrictive parameters in order to decrease the number of inactives that surpass the filter (therefore reducing the number of false positives) at the expense of losing the ability to correctly predict a proportion of the actives (therefore increasing the number of false negatives). In this situation we would be prioritizing the precision of the model over its accuracy. In a different situation, for instance, if the proportion of actives that surpassed the filter was considered to be low (meaning that the model is too restrictive), the parameters of the model could be tweaked to try to increase the proportion of actives that surpass the filter while avoiding the retrieval of decoys in an attempt to find the optimum compromise between sensitivity and specificity.

Overall, these measures allow to evaluate different aspects of the model in an objective manner and adjust its the parameters of according to the preferences of the user in order to obtain the most suitable model for each situation.

6. Hit selection

In order to demonstrate the usefulness of the VS, the activity of VS hits for the target of interest needs to be determined *in vitro* (see Figure 1). This is usually done to a sample of the hit compounds. Subjective cherry picking should be avoided in the selection of compounds in order to obtain a representative sample that allows an adequate evaluation of VS performance.⁶⁹ For this, it is crucial to select compounds that are different from one another, and this can be achieved by grouping the hit compounds according to their structure using clustering. Clustering is an unsupervised learning method in which the algorithm is provided with input data and its goal is to find patterns in the data and divide the input into groups. Fingerprint data can, for instance, be used as input to cluster compounds according to their structures. This allows to: **a)** select hit compounds that belong to different clusters than known actives and are therefore more novel and have a greater interest; **b)** avoid selecting compounds that cluster together with inactives and therefore have a greater chance of being inactive; and **c)** select hit compounds that belong to different clusters to ensure structural diversity.

Different clustering methods exist, such as hierarchical clustering, *k*-means clustering or HDBSCAN,⁷⁰ and the choice of the method is generally influenced by the characteristics of the dataset and the limitations of each method. For instance, in *k*-means clustering the number of clusters is predefined and they are circumferential, whereas in HDBSCAN⁷⁰ the minimum cluster size can be modified to alter the number of clusters and it is also indicated for outlier detection. In hierarchical clustering, an arbitrary similarity threshold can be established to define the desired number of clusters.

7. Experimental validation

As previously mentioned, VS hits are compounds which are expected to have an action on the target of interest, but this has to be demonstrated. The activity of the compound for the target of interest can be tested in the laboratory and this is usually

Manuscript 1

accomplished by using a commercial enzymatic activity kit or by developing an in-house method for the detection of enzymatic activity and comparing the activities of the target with and without the presence of the compound that was obtained as a hit.

Nevertheless, the result of a single experiment does not always guarantee that the compound has an action of the target itself, as some compounds may give a positive experimental result by interfering with the assay. Unlike in the case of a true drug, which inhibits or activates a protein by fitting into its binding site, these compounds give positive experimental results without performing a specific, drug-like interaction with the protein.⁷¹ Compounds that produce such results are referred to as pan-assay interference compounds (or simply PAINS),⁷¹ and they can be classified by their mode of action:

- Fluorescent or highly colored compounds. These compounds may give false readouts in fluorimetric and colorimetric assays, giving a positive signal even when no protein is present.
- Compounds that trap the toxic or reactive metals used to synthesize molecules in a screening library or used as reagents in assays. These metals give rise to signals that have nothing to do with the compound's interaction with the protein.
- Compounds that sequester reactive metal ions necessary for the reaction.
- Compounds that coat the protein, altering it chemically and affecting its function in an unspecific manner without fitting into its binding site.

Based on previous experience, a series of recommendations have been given on how to discern between hits and PAINS:

- **Identify potential PAINS based on structure.** Most PAINS fall into 16 different categories according to their chemotypes. Therefore, hits that are candidates of being PAINS can be identified by checking if they present one of these chemotypes.⁷¹ While this is more effectively done by eye, several *in silico* tools that implement chemical similarity and substructure searches have also been developed for this purpose.^{26,72–74} Nevertheless, this does not ensure that all PAINS are discarded and experimental

The light and dark sides of virtual screening: what is there to know?

testing will ultimately be needed to identify whether hit compounds giving positive experimental results are acting as PAINS or not.

- **Perform more than one assay.** To have more certainty that a hit compound which gives a positive experimental result is not a PAIN compound, it is advisable to conduct at least one more assay that detects activity with a different readout in order to check whether the compound is interfering with the assay or not. It is also advisable to check the activity of hits against unrelated targets and if the inhibition of the target of interest is competitive to determine whether the binding of the hit compound is specific or not.

Another reason for which a compound may give a false positive experimental result is aggregation. Some compounds form aggregates that adsorb and denature the protein, inhibiting it in an unspecific manner.⁷² Molecules that act as aggregators can be predicted computationally⁷⁵ and detected experimentally using different tests:

- Aggregates can be observed directly by dynamic light-scattering as they form particles from 50 to 400 nm in diameter.^{72,76}
- Inhibition by colloidal aggregates can be significantly attenuated by small amounts of a non-ionic detergent such as Triton-X or Tween-20.⁷⁷
- Inhibition by colloidal aggregates can also be attenuated by increasing enzyme concentration, whereas this should not affect the inhibition by well-behaved inhibitors if the receptor concentration to K_i ratio is high.⁷⁷
- Inhibition by colloidal aggregates is non-competitive. If the binding of the compound in question is competitive, the compound is unlikely to be an aggregator^{72,77}

The combination of several of these tests is recommended to determine with a greater confidence whether the compound acts as an aggregator or a true drug-like compound.⁷⁷

Overall, although potential PAINS and aggregators can be discarded *in silico*, further experimental tests should be performed upon confirming the activity of a hit compound in order to determine whether the observed activity is a result of the desired interaction

Manuscript 1

of the compound with the protein or, on the contrary, it corresponds to a false positive result as the compound behaves as a PAIN compound or an aggregator. From a VS design and validation perspective, the fact that a substantial number of molecules reported to be active against a target protein are actually PAINS and aggregators that provided false positive experimental results is a limitation¹ that should be taken into account when: **a)** selecting compounds as a reference to perform similarity searches; and **b)** evaluating the false negatives obtained in the VS validation, as PAINS and aggregators may fall into this category.

8. ADME

Even if a compound is able to specifically bind to the target of interest and its activity is confirmed *in vitro*, this does not imply that it will have the desired effect *in vivo*. First, the compound will need to be properly absorbed by the organism and distributed to the tissue of interest while avoiding being metabolized and excreted. The properties of a compound that have an influence on its *in vivo* activity through the modulation of one of these stages are commonly referred to as ADME (*Absorption, Distribution, Metabolism and Excretion*) properties and they can be used to determine the druglikeness of the compound, that is, how a compound resembles an actual drug and therefore can be processed as such by the organism.

ADME properties include physical properties of the compound like its solubility and hydrophobicity. Generally, a drug needs to be soluble enough to be carried by the blood stream, but also lipophilic enough to penetrate the lipidic bilayer that composes the cell membrane. As these properties are inherent to the particular structure of the compound, they can be predicted *in silico* with the help of mathematical algorithms (see Table 1). Other ADME properties such as the skin, gut-blood and blood brain barrier permeabilities of the compound can also be approximated computationally based on the reported data for known drugs.

The light and dark sides of virtual screening: what is there to know?

One of the most critical aspects regarding the effectiveness of an oral drug, apart from its potency, is its absorption and the proportion of the drug that reaches the blood stream, also referred to as bioavailability. If a compound presents high activity towards a target but it has low bioavailability, it will not exert the desired effect. The most popular method to predict the bioavailability of a compound is the so-called Lipinski's rule of 5, which was developed 20 years ago by Lipinski *et al.*⁷⁸ The rule is based on the ADME properties of known drugs, stating that most of the orally active drugs in humans fulfill 3 of the following 4 criteria:

1. A maximum of 5 hydrogen bond donors.
2. A maximum of 10 hydrogen bond acceptors.
3. A molecular weight of less than 500 daltons.
4. An octanol-water partition coefficient not greater than 5.

Therefore, compounds in the screening library that fulfill Lipinski's rule of 5 are more likely to be orally active and can be filtered either at early stages or at the end of the VS. It should be kept in mind that Lipinski's rule of 5 only applies to oral bioavailability, and that drugs designed aiming at other administration routes routinely fall outside the scope of this rule.¹

9. Conclusions

In this review, we have introduced the most common methodologies used in VS and given recommendations on how to design, build and validate a VS workflow. With these guidelines we hope to encourage the usage of VS, help researchers familiarize with its capabilities and, most importantly, raise awareness of common mistakes in order to promote the proper usage of VS techniques.

Manuscript 1

Acknowledgements

AG contract is supported by grant 2015FI_B00655 of the Catalonia Government.

References

- (1) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52* (4), 867–881.
- (2) Kumar, A.; Zhang, K. Y. J. Hierarchical Virtual Screening Approaches in Small Molecule Drug Discovery. *Methods* **2015**, *71*, 26–37.
- (3) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alpi, E.; Antunes, R.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; et al. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45* (D1), D158–D169.
- (4) Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. BRENDA in 2017: New Perspectives and New Tools in BRENDA. *Nucleic Acids Res.* **2017**, *45* (D1), D380–D388.
- (5) Flare, v2.0 , Cresset®, Litlington, Cambridgeshire, UK, ; <http://www.cresset-group.com/flare/>; Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.*
- (6) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Motow, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; et al. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45* (D1), D945–D954.
- (7) Reaxys <https://www.reaxys.com/>.
- (8) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Res.* **2016**, *44* (D1), D1045–D1053.
- (9) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202–D1213.
- (10) RCSB PDB <http://www.rcsb.org>.
- (11) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (12) Cereto-Massagué, A.; Ojeda, M. J.; Joosten, R. P.; Valls, C.; Mulero, M.; Salvado, M. J.; Arola-Arnal, A.; Arola, L.; Garcia-Vallvé, S.; Pujadas, G. The Good, the Bad and the Dubious: VHELIBS, a Validation Helper for Ligands and Binding Sites. *J. Cheminform.* **2013**, *5* (1), 36.
- (13) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324–2337.
- (14) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57* (8), 1747–1756.
- (15) Friedrich, N.-O.; de Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J. Chem. Inf. Model.* **2017**, *57* (11), 2719–2728.

Manuscript 1

- (16) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572–584.
- (17) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50* (4), 534–546.
- (18) Blaney, J. M.; Dixon, J. S. Distance Geometry in Molecular Modeling; Wiley-Blackwell, 2007; pp 299–335.
- (19) RDKit: Open-source cheminformatics <http://www.rdkit.org>.
- (20) Standardizer 16.10.10.0, 2016, ChemAxon <http://www.chemaxon.com>.
- (21) Schrödinger Release 2018-3: LigPrep, Schrödinger, LLC, New York, NY, 2018.
- (22) <https://molvs.readthedocs.io/en/latest/>.
- (23) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55* (12), 2562–2574.
- (24) Schrödinger Release 2018-3: QikProp, Schrödinger, LLC, New York, NY, 2018.
- (25) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7* (1), 42717.
- (26) <http://fafdrugs4.mti.univ-paris-diderot.fr/>.
- (27) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50* (1), 74–82.
- (28) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *J. Chem. Inf. Model.* **2011**, *51* (10), 2455–2466.
- (29) EON 2.2.0.5: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>. <http://www.eyesopen.com>.
- (30) Dixon, S. L.; Smondryev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided. Mol. Des.* *20* (10–11), 647–671.
- (31) Gerhard Wolber*, † and; Langer‡, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. **2004**.
- (32) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
- (33) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.

The light and dark sides of virtual screening: what is there to know?

- (34) Allen, W. J.; Balias, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132–1156.
- (35) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
- (36) Schrödinger Release 2018-1: Maestro, Schrödinger, LLC, New York, NY, 2018.
- (37) VIDA 4.4.0: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.
- (38) Cereto-Massagué, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. DecoyFinder: An Easy-to-Use Python GUI Application for Building Target-Specific Decoy Sets. *Bioinformatics* **2012**, *28* (12), 1661–1662.
- (39) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; 1990.
- (40) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular Fingerprint Similarity Search in Virtual Screening. *Methods* **2014**, *71* (August), 6–11.
- (41) Accelrys, MACCS structural keys, (n.d.).
- (42) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280.
- (43) Gianti, E.; Sartori, L. Identification and Selection of “Privileged Fragments” Suitable for Primary Screening. *J. Chem. Inf. Model.* **2008**, *48* (11), 2129–2139.
- (44) Daylight chemical information systems, Daylight. <http://www.daylight.com>.
- (45) OpenEye scientific software, OEChem, 2013.
- (46) Andreas Bender, *; Hamse Y. Mussa, A.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. **2003**.
- (47) Andreas Bender; Hamse Y. Mussa, A.; Glen*, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. **2004**.
- (48) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.
- (49) McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 2. Application to Primary Library Design. *J. Chem. Inf. Comput. Sci.* **40** (1), 117–125.
- (50) Schwartz, J.; Awale, M.; Reymond, J.-L. SMI1p (SMILES Fingerprint) Chemical Space for Virtual Screening and Visualization of Large Databases of Organic Molecules. *J. Chem. Inf. Model.* **2013**, *53* (8), 1979–1989.
- (51) Chemical Computing Group Inc., Molecular operating environment (MOE), 2013.
- (52) Zhan Deng; Claudio Chuaqui, A.; Singh*, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein–Ligand Binding Interactions. **2003**.

Manuscript 1

- (53) Ling Xue, †; Jeffrey W. Godden, †; Florence L. Stahura, † and; Jürgen Bajorath*, †. Design and Evaluation of a Molecular Fingerprint Involving the Transformation of Property Descriptor Values into a Binary Classification Scheme. **2003**.
- (54) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminform.* **2015**, *7* (1), 20.
- (55) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57* (8), 3186–3204.
- (56) Kumar, A.; Zhang, K. Y. J. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Front. Chem.* **2018**, *6* (July), 1–21.
- (57) Connolly, M. L. Computation of Molecular Volume. *J. Am. Chem. Soc.* **1985**, *107* (5), 1118–1124.
- (58) Masek, B. B.; Merchant, A.; Matthew, J. B. Molecular Shape Comparison of Angiotensin II Receptor Antagonists. *J. Med. Chem.* **1993**, *36* (9), 1230–1238.
- (59) Grant, J. A.; Pickup, B. T. A Gaussian Description of Molecular Shape. *J. Phys. Chem.* **1995**, *99* (11), 3503–3510.
- (60) GRANT, J. A.; GALLARDO, M. A.; PICKUP, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1996**, *17* (14), 1653–1666.
- (61) Mezey, P. G. *Molecular Surfaces*; John Wiley & Sons, Ltd, 2007; pp 265–294.
- (62) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55* (3), 379-404.
- (63) Connolly, M. L. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221* (4612), 709–713.
- (64) Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70* (5), 1129–1143.
- (65) Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J. Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications. *J. Med. Chem.* **2010**, *53* (24), 8461–8467.
- (66) Kolb, P.; Irwin, J. Docking Screens: Right for the Right Reasons? *Curr. Top. Med. Chem.* **2009**, *9* (9), 755–770.
- (67) Salam, N. K.; Nuti, R.; Sherman, W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *J. Chem. Inf. Model.* **2009**, *49* (10), 2356–2368.
- (68) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594.
- (69) U.S. National Library of Medicine. National Institutes of Health. MedlinePlus. Metformin.
- (70) Campello, R. J. G. B.; Moulavi, D.; Sander, J. *Density-Based Clustering Based on Hierarchical Density Estimates*; Springer, Berlin, Heidelberg, 2013; pp 160–172.

The light and dark sides of virtual screening: what is there to know?

- (71) Baell, J.; Walters, M. A. Chemistry: Chemical Con Artists Foil Drug Discovery. *Nature* **2014**, *513* (7519), 481–483.
- (72) Panaceas, I. M. The Ecstasy and Agony of Assay Interference Compounds. **2017**.
- (73) <http://www.cbligand.org/PAINS/>.
- (74) <http://zinc15.docking.org/patterns/home>.
- (75) <http://advisor.docking.org>.
- (76) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.* **2002**, *45* (8), 1712–1722.
- (77) Feng, B. Y.; Shoichet, B. K. A Detergent-Based Assay for the Detection of Promiscuous Inhibitors. *Nat. Protoc.* **2006**, *1* (2), 550–553.
- (78) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23* (1–3), 3–25.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo



Hypotheses
And
Objectives

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

°°

According to the involvement of PTP1B and MMP-13 in obesity and its associated disorders, three hypotheses have been proposed:

1. Structurally diverse PTP1B inhibitors can be obtained by virtual screening.
2. Targeting the S1' pocket of MMPs constitutes a means of achieving selectivity against other members of the MMP family.
3. Selective MMP-13 inhibitors can be obtained through virtual screening by targeting its S1'' pocket.

In order to test these hypotheses, the following objectives have been established:

1. Identify structurally diverse PTP1B inhibitors by virtual screening (**Manuscript 2**).
2. Assess the variability of the S1' pocket among the members of the MMP family and determine the characteristics of the S1' pocket that govern the selectivity of previously reported MMP inhibitors (**Manuscript 3**).
3. Identify selective MMP-13 inhibitors through virtual screening by targeting its S1'' pocket (**Manuscript 4**).

D'acord amb la involucració de PTP1B i MMP-13 en l'obesitat i els seus trastorns associats, s'han proposat tres hipòtesis:

1. Es poden obtenir inhibidors PTP1B estructuralment diversos mitjançant tècniques de cribratge virtual.
2. La butxaca S1' de les MMPs es pot utilitzar per obtenir inhibidors selectius contra altres membres de la família de les MMPs.
3. Es poden obtenir inhibidors selectius de MMP-13 per mitjà de cribratge virtual utilitzant la butxaca S1''.

Per comprovar aquestes hipòtesis, s'han establert els següents objectius:

1. Identificar inhibidors estructuralment diversos de PTP1B mitjançant tècniques de cribratge virtual (**Manuscrit 2**).
2. Avaluar la variabilitat de la butxaca de S1' entre els membres de la família de les MMPs i determinar les característiques de la butxaca S1' que regeixen la selectivitat dels inhibidors de MMP prèviament descrits (**Manuscrit 3**).
3. Identificar inhibidors selectius de MMP-13 mitjançant tècniques de cribratge virtual utilitzant la butxaca S1'' (**Manuscrit 4**).



UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Manuscript 2

Combined ligand- and receptor-based virtual screening methodology to identify structurally diverse protein tyrosine phosphatase 1B inhibitors

Aleix Gimeno,^[a] Andrea Ardid-Ruiz,^[b] María José Ojeda-Montes,^[a] Sarah Tomás-Hernández,^[a] Adrià Cereto-Massagué,^[a] Raúl Beltrán-Debón,^[a] Miquel Mulero,^[a] Cristina Valls,^[a] Gerard Aragonès,^[b] Manuel Suárez,^[b] Gerard Pujadas,^{*,[a, c]} and Santiago Garcia-Vallvé^[a, c]

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]Nutrigenomics Research Group, Department of Biochemistry and Biotechnology, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[c]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

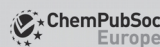
*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo



DOI: 10.1002/cmdc.201800267

CHEM MED CHEM
Full Papers

Combined Ligand- and Receptor-Based Virtual Screening Methodology to Identify Structurally Diverse Protein Tyrosine Phosphatase 1B Inhibitors

Aleix Gimeno,^[a] Andrea Ardid-Ruiz,^[b] María José Ojeda-Montes,^[a] Sarah Tomás-Hernández,^[a] Adrià Cereto-Massagué,^[a] Raúl Beltrán-Debón,^[a] Miquel Mulero,^[a] Cristina Valls,^[a] Gerard Aragonès,^[b] Manuel Suárez,^[b] Gerard Pujadas,^{*,[a, c]} and Santiago Garcia-Valle^[a, c]

Protein tyrosine phosphatase 1B (PTP1B) is a potential drug target for diabetes and obesity. However, the design of PTP1B inhibitors that combine potency and bioavailability is a great challenge, and new leads are needed to circumvent this problem. Virtual screening (VS) workflows can be used to find new PTP1B inhibitors with little chemical similarity to existing inhibitors. Unfortunately, previous VS workflows for the identification of PTP1B inhibitors have several limitations, such as a

small number of experimentally tested compounds and the low bioactivity of those compounds. We developed a VS workflow capable of identifying 15 structurally diverse PTP1B inhibitors from 20 compounds, the bioactivity of which was tested in vitro. Moreover, we identified two PTP1B inhibitors with the highest bioactivity reported by any VS campaign (i.e., IC₅₀ values of 1.4 and 2.1 μM), which could be used as new lead compounds.

Introduction

Diabetes mellitus and obesity have become major public health problems in today's society, affecting more than 9%^[1] and 13%^[2] of the adult population worldwide, respectively. The combination of a sedentary lifestyle and an unbalanced diet promotes the development of obesity because it increases lipid storage and encourages the expansion of adipose tissue and the progressive loss of leptin sensitivity.^[3] This triggers the secretion of cytokines, which ultimately results in a systemic inflammation driven by the immune response.^[4,5] This inflammation can result in increased insulin resistance,^[6,7] decreased insulin secretion by the pancreatic islets,^[8] and the increased permeability of the vascular endothelium,^[3] favoring the development of type 2 diabetes mellitus (T2DM)^[9,10] and cardiovascular disease (CVD).^[11] Moreover, the higher waist circumference of obese patients is thought to be related to elevated blood pres-

sure and risk of CVD.^[12] Elevated waist circumference as a measure of central obesity, elevated blood pressure, atherogenic dyslipidemia (i.e., elevated TAG and lowered HDL cholesterol) and increased fasting glucose levels (as an outcome of increased insulin resistance and decreased insulin secretion) constitute a cluster of cardiovascular risk factors that define metabolic syndrome, which in turn confers a fivefold increase of risk for T2DM.^[13] All the cardiovascular risk factors involved in metabolic syndrome are part of a wide range of diseases that should not be treated separately, but recognized as components of a greater single disease.^[14]

Protein tyrosine phosphatase 1B (PTP1B) is a phosphatase whose increased activity and expression are associated with resistance to the hormones insulin^[15] and leptin.^[16,17] PTP1B acts at several stages of the insulin and leptin signaling pathways.^[18] In the insulin signaling pathway, PTP1B dephosphorylates the insulin receptor, thus diminishing insulin action. It also dephosphorylates IRS1/2, inhibiting insulin signal transduction.^[19,20] As a result, both the Akt/PKB pathway and the Ras-MAPK pathway are affected. In the leptin signaling pathway, PTP1B dephosphorylates JAK2, altering the JAK-STAT pathway and therefore inhibiting the anorexigenic action of leptin in the hypothalamus, resulting in an increase in food intake and a decrease in energy expenditure.^[21,22] Consequently, PTP1B inhibition has emerged as a promising strategy for the treatment of diabetes and obesity.^[23,24]

Over the last two decades, different classes of PTP1B inhibitors have been identified through rational drug design.^[25] However, achieving a compromise between the activity and pharmacokinetics of PTP1B inhibitors has proved a major challenge in medicinal chemistry due to the characteristics of the

[a] A. Gimeno, Dr. M. J. Ojeda-Montes, Dr. S. Tomás-Hernández, Dr. A. Cereto-Massagué, Dr. R. Beltrán-Debón, Dr. M. Mulero, Dr. C. Valls, Dr. G. Pujadas, Dr. S. Garcia-Valle
Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona Catalonia (Spain)
E-mail: gerard.pujadas@urv.cat

[b] A. Ardid-Ruiz, Dr. G. Aragonès, Dr. M. Suárez
Nutrigenomics Research Group, Department of Biochemistry and Biotechnology, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona Catalonia (Spain)

[c] Dr. G. Pujadas, Dr. S. Garcia-Valle
EURECAT, TECNIO, CEICS, Avinguda Universitat 1, 43204 Reus Catalonia (Spain)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/cmdc.201800267>.

PTP1B binding site. The combination of the backbone amide protons of the P-loop (defined by residues 214 to 221 with HCSAGIGR sequence) and the basic nature of the Arg221 side chain create a highly positive-charged environment that has a preference for ligands with highly acidic groups (see Figure 1).

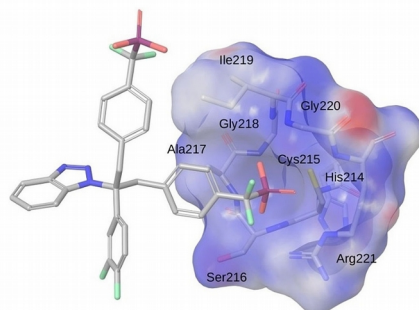


Figure 1. Representation of the crystal structure with PDB^[34,35] code 1Q6M^[36] showing the P-loop residues of PTP1B and the co-crystallized ligand. The ligand and the P-loop residues are shown in sticks and the molecular surface of the P-loop residues has been represented and colored according to their Poisson-Boltzmann electrostatic potentials from red (negative) to blue (positive), where potentials range from -50.0 to 50.0 . This image was obtained with Maestro^[37] v10.7.

VS workflows capable of further exploring the chemical space by identifying new chemical entities with diverse structures may prove useful in the pursuit of molecules that show potential for use as new lead compounds for the design of PTP1B inhibitors with a good balance between potency and bioavailability. To date, several VS workflows have been developed for the purpose of discovering novel PTP1B inhibitors.^[26–33] Despite their success in identifying PTP1B inhibitors, they often present drawbacks such as few^[31–33] or no compounds^[26,27,29] tested in vitro, low percentages of active compounds^[28,30] and relatively low activity values of the identified inhibitors^[28,30,33] (see Table 1).

Table 1. Summary of the manuscripts that have been published containing VSs for PTP1B inhibitors.

VS	Compounds tested in vitro	Compounds with PTP1B activity	Compounds with activity in the 1–10 μM range
Floriano et al. ^[26]	0	–	–
Bharatham et al. ^[27]	0	–	–
Park et al. ^[28]	225	62	0
Rao et al. ^[29]	0	–	–
Reddy et al. ^[30]	43	4	0
Ma et al. ^[31]	8	8	2
Balaramnavar et al. ^[32]	10	10	2
Chandra et al. ^[33]	5	5	0
Present study	20	15	2

Therefore, the aim of this study was to design a new VS methodology and validate it by means of the in vitro testing of a structurally diverse set of compounds, which addresses the need to find new PTP1B inhibitors with increased potency that can be used as lead compounds in the search for new PTP1B inhibitors.

Results and Discussion

We have developed a VS workflow capable of identifying PTP1B inhibitors from a given library of compounds. The workflow consists of a series of successive filters, namely a molecular weight (MW) filter, a random forest (RF) model based on fingerprints, protein–ligand docking and an electrostatic similarity analysis, in which the output molecules of each filter were the input of the next one (i.e., those molecules that do not accomplish one filter are eliminated for subsequent filters; see Figure 2). The performances of the RF model and the electrostatic similarity analysis were validated in silico with two different sets of actives and decoys. Finally, a library of 212672 compounds obtained from the Specs^[38] company was screened for the purpose of validating the VS protocol in vitro by finding new PTP1B inhibitors with previously undescribed scaffolds for such bioactivity.

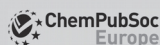
Workflow steps	Number of compounds
Starting library	212,672
MW filter	65,223
Random Forests	30,838
Docking	11,429
Electrostatic similarity analysis	717
Activity assay	20

Figure 2. Diagram of the virtual screening indicating the different filters used and the number of compounds that overcome each of them.

1. MW filter

In the first VS step, the Specs library was filtered according to MW to discard molecules that were considered either too small or too large to fit on the PTP1B binding site. To determine the specific molecular weight range, all inhibitors of human PTP1B with an average pX bioactivity greater than or equal to 4 were obtained from Reaxys^[39] (a total of 6091 molecules) and the distribution of their MW was plotted (see Figure S1). Using the MW of these known actives as a reference, molecules with a MW lower than 200 Da and higher than

Combined ligand- and receptor-based VS to identify PTP1B inhibitors



CHEM MED CHEM
Full Papers

900 Da were removed from the library, leaving 65223 remaining compounds and discarding about 70% of the initial compounds (see Figure 2). Although no validation was performed for this step, the rapid exclusion of compounds from the scope of the VS justifies the use of this filter at the beginning of the workflow.

2. RF model

Next, an RF model based on circular fingerprints (FPs) was developed to rapidly rule out the compounds least likely to be active. Circular FPs (also known as Morgan FPs because they are based on the Morgan algorithm) record the environment of each atom in the molecule up to a determined radius.^[40] As hashed topological fingerprints, they have the advantage of being built around each molecule; therefore, any molecule can produce a meaningful FP. Despite the drawback that FP bits do not relate to the presence or absence of a particular substructure, because the RF model from these FPs was developed for the specific purpose of enriching the library in actives, circular FPs have been used to develop the model as they are one of the highest ranked FPs in terms of FP performance.^[40]

As a supervised machine learning algorithm based on fingerprints, in this RF model, FP bits are related to bioactivity and the output probabilities are a function of the presence of important structural characteristics for the bioactivity of known actives. Therefore, the model should be capable of identifying molecules with similar structural characteristics to known actives, but with different overall structures, recognizing compounds that present crucial features for activity as compounds with a high probability of being active and vice versa. Because the RF model is not the last step of the VS workflow (see Figure 2), compounds presenting a probability of being active of less than 20% were excluded. For this threshold, in a cross-validation using actives and decoys, the RF model performed with average precision and sensitivity values of 94.1% and 98.7% (see Table 2 for performance details). Despite this relatively low threshold, about 50% of the compounds from the Specs library that passed the MW filter were discarded and

Table 2. Statistical parameters of the RF model cross-validation.

Parameter	5-Fold cross-validation ^[a]
Sensitivity	0.9875
Specificity	0.9379
Precision	0.9408
Fall-out	0.0621
False negative rate	0.0125
False discovery rate	0.0592
Accuracy	0.9627
F1 score	0.9636
Matthews correlation coefficient	0.9265

[a] Values are the means of the 5-fold cross-validation for each parameter.

30838 compounds remained (see Figure 2). Therefore, the development of this predictive model allowed us to quickly dismiss a large number of molecules with low probabilities of inhibiting PTP1B due to the low computational cost of FP calculation and machine learning, thus proving a good filtering option for application at early stages of the VS workflow.

3. Protein–ligand docking

The compounds resulting from the RF filter were docked onto the binding site of PTP1B using the protein from the crystal structure with PDB^[34,35] code 1Q6M.^[36] It has previously been reported that the use of pharmacophoric constraints during docking (i.e., a polar interaction with Arg221 and at least one interaction with the backbone amides of the residues of the P-loop) contributes to an increase in VS enrichment when searching for PTP1B inhibitors.^[41] Thus, with this idea in mind, the ligand interactions with the P-loop residues were analyzed for the crystal structures containing the most potent PTP1B inhibitors (see Table 3). In this analysis, a series of hydrogen bond interactions were identified to be common among most of these inhibitors and, therefore, the protein atoms involved in these interactions were used to set hydrogen bond constraints during the protein–ligand docking. A total of six protein residues were used to define these hydrogen bond constraints (see the **Grid generation** section below for more de-

Table 3. Summary of the interactions between crystallized PTP1B inhibitors with an IC_{50} or K_i lower than or equal to 100 nM and their target.^[a]

	1Q6J	1Q6M	1Q6N	1Q6P	1Q6S	1Q6T	2FJN	2QBP	2QBQ	2VEU	2VEW	2VEY	2ZMM	2ZN7
His214	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Cys215	SG	✓	SG	SG	SG	SG	SG	✓	✓	✓, SG	✓	✓	✓	✓
Ser216	N	N	N	N	N	N	N	✓	✓	N	N	N	✓	✓
Ala217	✓, N	✓, N	✓, N	✓, N	✓, N	✓, N	✓, N	✓	✓	✓, N	✓, N	✓, N	✓	✓
Gly218	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Ile219	✓, N	✓, N	✓, N	✓, N	✓, N	✓, N	✓, N	–	–	✓, N	N	✓, N	✓	✓
Gly220	N	N	N	N	N	N	N	–	✓	N	N	N	✓	✓
Arg221	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH1	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2	N, NE, NH2

[a] Data were obtained from the LigPlot^[42] and PoseView^[43] diagrams in the PDBsum^[44] and PDB^[34,35] databases. Non-bonded contacts are indicated with check marks, and hydrogen bonds are indicated with the PDB^[34,35] label of the protein atom that is involved in the interaction. The crystallized ligands of these protein–ligand complexes were used as a reference to perform electrostatic comparisons. The ligands of all 14 protein–ligand complexes were used to perform 14 separate validations (see Figure S2), but only the ligands of the 11 protein–ligand complexes in bold were selected as queries for the electrostatic similarity analysis.



tails). Because the next step of the VS is a 3D-based similarity analysis with actives, hydrogen bond constraints are also helpful in orienting each molecule at the binding site and dismissing the molecules that are unable to form hydrogen bond interactions with the P-loop. These molecules are not of interest to us because their binding mode would be different to that of the molecules used as a reference in the subsequent electrostatic similarity analysis.

Although this docking step allows us to discard compounds that would not fit in the binding site as well as those that would not be capable of establishing hydrogen bonds with the P-loop residues, it is not considered a filter per se, included to increase enrichment in actives, but instead it is considered a means of generating hypotheses on how each compound may bind to the binding site. Therefore, no validation was performed for this step of the VS. After applying this filter to the remaining compounds from the Specs library, 114823 docked poses corresponding to 11429 compounds survived this step (see Figure 2).

4. Electrostatic similarity analysis

Once the docked poses were obtained, an electrostatic similarity analysis was performed to compare the electrostatic potential of the 114823 docked poses for the remaining 11429 molecules to those of the experimental poses for potent PTP1B inhibitors (see Figure 2). Thus, 11 crystallized complexes between PTP1B and PTP1B inhibitors with IC_{50} or K_i activity values between 1 and 100 nM were superposed to the crystal structure with PDB^[34,35] code 1Q6M^[36] (the same PDB^[34,35] structure that was used during the protein–ligand docking step of the VS) and used as a reference for the electrostatic comparison (see Table 3). Comparing the superposed experimental poses directly with the 114823 docked poses of the 11429 compounds allows us to overcome the limitations of docking scoring functions by selecting the most reliable docking pose for each compound based on its similarity to the known binding mode of an active compound.

To validate the use of these co-crystallized PTP1B inhibitors as reference compounds for the electrostatic potential comparison in silico, a validation set of 1424 actives and 1500 decoys was prepared and docked to 1Q6M^[36] in the same conditions as the library compounds and compared with the set of the 11 experimental poses (see Table 3) in order to obtain the electrostatic Tanimoto (i.e., EON_ET_pb) coefficients of the docked poses for actives and decoys from the validation set for each comparison. Once the EON_ET_pb values were computed for each molecule, only the docked pose that presented the highest EON_ET_pb value regardless of the reference compound used for the comparison was kept. After discarding all other docking poses, histograms of the electrostatic Tanimoto coefficients for the actives and the decoys in the validation set were plotted, separating the actives into 3 groups depending on whether their bioactivity had a pX value lower than 4, between 4 and 7 or higher than 7 (see Figure 3). As Figure 3 shows, applying an electrostatic Tanimoto (i.e., EON_ET_pb) cutoff of 0.4 to the validation set allowed us to discard 1436 decoys (96%

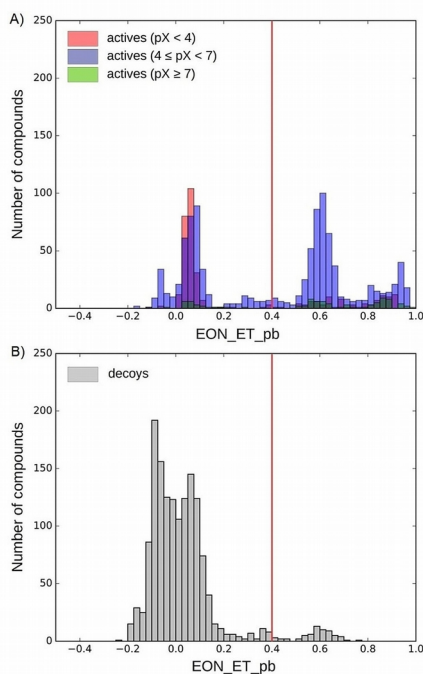


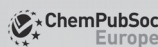
Figure 3. Histogram representation of the highest electrostatic Tanimoto (i.e., EON_ET_pb) values obtained in the comparison of the validation set to all queries. A) Histogram for the actives in the validation set. Actives with a pX lower than 4 are in red, actives with a pX from 4 to 7 are in blue, and actives with a pX higher than 7 are in green. B) Histogram for the decoys in the validation set. The EON_ET_pb cutoff of 0.4 is represented as a red line.

of the initial decoys) and retain 739 actives (52% of the initial actives). Therefore, this cutoff was applied to our compound library after performing the same electrostatic potential comparison in order to enrich it with active molecules, leaving 717 remaining compounds (see Figure 2).

5. Clustering and hit selection

After having computed their EON_ET_pb similarities with experimental poses and applied the cutoff, the resulting 717 hit molecules were clustered using the HDBSCAN^[46] algorithm together with the 1424 active compounds used in the validation of the electrostatic similarity filter. Thus, hit compounds that were grouped in the same cluster as known active compounds were excluded. Clusters that contained only hit molecules were identified and the molecule with the highest EON_ET_pb value for one of the 11 experimental poses was selected from

Combined ligand- and receptor-based VS to identify PTP1B inhibitors



CHEM MED CHEM
 Full Papers

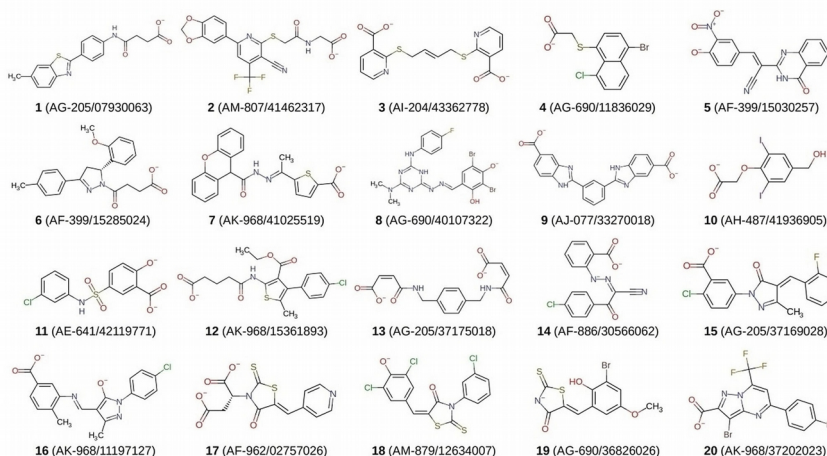


Figure 4. 2D structures of the 20 hit compounds. Each compound is identified with its Specs ID number. MarvinSketch⁴⁹¹ was used to draw the structures. The protonation state of each compound corresponds to the protonation state of the docked pose selected for that compound.

each cluster. The selected molecules were sorted according to electrostatic similarity and were visually inspected to examine the diversity of their structures. Finally, the top 20 molecules were selected for activity tests (see Figure 4). Figure 5 shows their structural diversity in a fingerprint-based dendrogram. This allowed us to identify the molecules that had similar electrostatic potential to active compounds, but with a different structure. These molecules may present similar chemical properties to those of active compounds and may thus be able to establish similar interactions with the protein environment, de-

spite their low structural similarity to known PTP1B inhibitors (see Figure S3).

6. Activity assays and comparison with other VS looking for PTP1B inhibitors

The 20 hit compounds were purchased from the Specs³⁸¹ company. Figure S4 shows the LC-UV/Vis and Q-TOF spectra and each compound. This data was obtained at the Center for Omic Sciences (COS, Reus, Catalonia) and shows that all compounds have a purity above 95.0% (see Table S1) except for compounds **13** (87.96%) and **17** (93.56%). The PTP1B inhibitory activity of these 20 hit compounds was analyzed in vitro at concentrations of 100 μM and 10 μM (see Table 4). Fifteen of the 20 hit compounds (**2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 18, 19** and **20**) showed PTP1B inhibition at a concentration of 100 μM . Although compounds **2, 3, 4, 6, 9, 10, 12, 13, 14, 15, 18, 19** and **20** displayed decreased or no inhibition at the concentration of 10 μM , interestingly, compounds **5** (with a purity of 97.77%) and **7** (with a purity of 99.64%) displayed complete inhibition of the enzyme at both concentrations and were selected for IC_{50} determination (see Table 4). Compound **8** was not selected because it inhibited the enzyme at the concentration of 10 μM , but not at the concentration of 100 μM . The IC_{50} values of compound **5** and compound **7** were 1.4 and 2.1 μM , respectively (see Table 4).

To our knowledge, these compounds represent the highest inhibitory potency for PTP1B obtained by any VS method. Moreover, although previous VS protocols^{31,32} have also obtained PTP1B inhibitors with activity values in the low μM range, they were similar in structure. For example, Ma et al.³¹

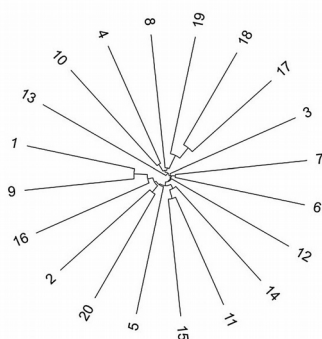


Figure 5. Dendrogram based on fingerprints showing the structural diversity of the 20 hit compounds. The FP used to obtain the distance matrix was the OpenEyePath FR. ITOL⁴⁷² was used to draw the dendrogram.

Table 4. Results of the activity assays performed on the 20 hit compounds showing the percentages of inhibition at the substrate concentrations of 100 and 10 μM , and the IC_{50} values of compounds 5 and 7. The experimental data used to calculate the IC_{50} values in this table can be found in Figure S7.

Compound	PTP1B inhibition [%]		PTP1B IC_{50} [μM]
	100 μM	10 μM	
1	0 ^[a]	0 ^[a]	ND
2	12	0	ND
3	27	0 ^[a]	ND
4	44	0 ^[a]	ND
5	100 ^[b]	100 ^[b]	1.4
6	5	0 ^[a]	ND
7	100 ^[b]	100 ^[b]	2.1
8	0 ^[a]	77	ND
9	29	4	ND
10	72	11	ND
11	0 ^[a]	0 ^[a]	ND
12	37	0 ^[a]	ND
13	27	3	ND
14	89	0 ^[a]	ND
15	26	0 ^[a]	ND
16	0 ^[a]	0 ^[a]	ND
17	0 ^[a]	0 ^[a]	ND
18	16	0 ^[a]	ND
19	57	42	ND
20	40	0 ^[a]	ND
Suramin ^[c]	90	93	ND

[a] Negative inhibition values were considered as 0% inhibition. [b] Inhibition values higher than 100% were considered as 100% inhibition. [c] Suramin was used as a positive control. ND: not determined.

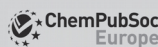
reported two inhibitors (labeled 6 and 4 by the authors) with IC_{50} values of 4.2 and 4.5 μM , respectively, but the compounds were structurally similar to one another and to previously identified PTP1B inhibitors (see Figure S5). Moreover, Balaramnavar et al.^[32] reported two other inhibitors (labeled 115 and 116 by the authors) with IC_{50} values of 7.5 and 8.3 μM , respectively, that share the same core structure and a different aryl sulfonyl substituent (see Figure S6). In contrast, the PTP1B inhibitors obtained in this work are more potent than those found by Ma et al.^[31] or Balaramnavar et al.,^[32] and they are structurally diverse and different from previously described PTP1B inhibitors (see Figure 5). Remarkably, 15 of the 20 compounds tested show bioactivity as PTP1B inhibitors. Therefore, this method offers unprecedented performance when using VS for exploring the chemical space in pursuit of new scaffolds with PTP1B inhibitory activity.

7. Performance of our VS workflow on compounds identified by other VSs

We have analyzed how our VS workflow acts on the hit compounds identified by previous PTP1B VSs (see Table 1). At this point, it is worth remarking that although the bioactivity of a total of 291 compounds was assayed in these studies, no data is available on the 2D structure for: a) the false positive compounds; and b) 53 out of the 62 actives identified by Park et al.^[28] Thus, we can only describe the performance for 36 PTP1B inhibitors from Table 1:

- Park et al.^[28]: out of the nine compounds with reported PTP1B inhibitory activity available from this study, five of them were identified as hits by our VS methodology (i.e., compounds 1, 3, 4, 8 and 9). Compounds 5 and 6 from this study were discarded in the protein–ligand docking step of our VS workflow because no docked poses were able to fulfill the hydrogen bond constraints specified in the protein–ligand docking step. This could be explained by the fact that hydrogen bond acceptors in both compounds are surrounded by bulky groups which may prevent them from interacting with the hydrogen bond donors of the P-loop. Compounds 2 and 7, on the other hand, fulfilled the docking constraints, but their docked poses showed low electrostatic similarity with the co-crystallized ligands used as queries and, therefore, they were discarded.
- Reddy et al.^[30]: out of the four compounds with reported PTP1B inhibitory activity in this study, three of them were identified as hits by our VS methodology (i.e., compounds AU-247, AU-2525 and AU-2439). Compound AU-008 was the only one that the VS was not able to predict as active and it was discarded by the RF model, possibly due to its low structural similarity with known PTP1B inhibitors.
- Ma et al.^[31]: out of the eight compounds with reported PTP1B inhibitory activity in this study, only one of them was identified as a hit by our VS methodology (i.e., compound 5). Compounds 1 and 2 were discarded by the RF model and compounds 3, 4, 6, 7 and 8 fulfilled the docking constraints but displayed an electrostatic similarity value lower than the defined cutoff value. As these five compounds present high structural similarity, it is no surprise that all of them are discarded for the same reason. Compound 5 (the only hit from Ma et al. that succeed in our VS workflow) is a special case in the sense that, despite its structural similarity with compounds 3, 4, 6, 7 and 8, it is the only compound in that study that has a COOH group which is not only able to fulfill the protein–ligand docking constraints but also provides a negative environment close to the P-loop residues that is similar to the one provided by the co-crystallized compounds used as queries during the electrostatic comparison.
- Balaramnavar et al.^[32]: out of the ten compounds with reported PTP1B inhibitory activity in this study, only one of them was identified as a hit by our VS methodology (i.e., compound 117). Compounds 116 and 117 reported by this work were the only ones able to fulfill our protein–ligand docking constraints due to the presence of a NO_2 group and a COOH group, respectively at their variable substructure. Interestingly, the protein–ligand docking results obtained by Balaramnavar et al.^[32] suggest that all ten compounds used the hydroxy group present in the substructure shared by all of them to act as a hydrogen bond acceptor of the protein donors located at the P-loop. In contrast, no pose for any of these 10 compounds in our protein–ligand docking protocol was able to use this hydroxy group to make hydrogen bonds with the P-loop residues. Moreover, using our protocol with the same PDB structure that Balaramnavar et al.^[32] used (i.e., PDB code 1AAx^[48]) also gave no

Combined ligand- and receptor-based VS to identify PTP1B inhibitors



CHEM MED CHEM
Full Papers

docked poses with the hydroxy group accepting protons from the P-loop and only docked poses for compounds **116** and **117** were again obtained. Therefore, such strong differences between protein–ligand docking results could be the result of differences in how the programs used to perform this step consider the steric hindrance between protein binding site and ligand (which would allow certain poses for one program that are forbidden for another one). Finally, only compound **117** surpassed the electrostatic similarity filter. This is surely the consequence of using query ligands at this step with negatively charged groups to interact with the P-loop (as compound **117** does with its COOH substituent in the docked pose we obtained).

- Chandra et al.^[33]: out of the five compounds with reported PTP1B inhibitory activity in this study, three of them were identified as hits by our VS methodology (i.e., compounds **SB017945C**, **JFD02789** and **JFD03705**). Compounds **BTB12807** and **JFD02644** were discarded in the last step of the workflow as they did not show enough electrostatic similarity with the reference compounds.

Overall, our VS workflow is able to identify 13 out of the 36 compounds identified by other VS methodologies as PTP1B inhibitors. From the 23 discarded molecules, 3 and 10 of them are eliminated by the RF and the electrostatic similarity filters, respectively. Remarkably, the protein–ligand docking step (that was not specifically designed as a filter and whose aim was to orient each molecule at the binding site and dismiss those that were unable to form hydrogen bond interactions with the P-loop) also discarded 10 compounds. However, as mentioned above, this could be related to differences in how the programs consider the steric hindrance between the protein binding site and the corresponding ligand, as 8 of these compounds shared the same substructure that is expected to hydrogen bond to the P-loop.^[32]

Conclusions

Due to the lack of PTP1B inhibitors presenting both a good activity and pharmacokinetic profile, new candidate compounds presenting PTP1B inhibitory activity through novel structures need to be identified for use as lead compounds. Despite previous efforts, over the last two decades very few PTP1B inhibitors resulting from VS have been reported. Our VS methodology demonstrates the potential of VS workflows because it was able to: a) find 15 compounds capable of inhibiting PTP1B at 100 μM (75% of the compounds experimentally tested); b) find 2 PTP1B inhibitors with an IC_{50} value in the 1–10 μM range; c) find PTP1B inhibitors with structural diversity; and d) find PTP1B inhibitors whose structure is not similar to that of known actives. Although further pharmacokinetic studies are required, this VS workflow has discovered several novel compounds capable of inhibiting PTP1B and has shown its potential for identifying more potent and structurally diverse molecules which could be used as lead compounds for the treatment of diabetes and obesity.

Nevertheless, although a good proportion of the tested compounds have been found to be active in this study, their activity values are not as high as those described in several structure–activity relationship studies.^[36,49] Keeping in mind that: a) the percentages of PTP1B inhibition of the 20 hits show poor or no correlation with either the docking score or the electrostatic Tanimoto (see Figure S8); and b) the required protein–ligand interactions were the establishment of at least one hydrogen bond with one of the residues of the P-loop and the implicit presence of a negative charge in that region, these results suggest that it would be necessary to consider further interactions within the binding site to achieve higher inhibitory potency during a VS workflow. For instance, Pandey et al.^[50] developed and validated different 3D-QSAR models from a set of peptidomimetic competitive inhibitors and concluded that: a) hydrogen bonding with the carboxylate group of Asp48 is an important interaction for PTP1B inhibitor activity; and b) the inclusion of steric bulks in the region around Phe182 are detrimental factors for PTP1B inhibitory activity. In another study, Gupta et al.^[51] remarked the importance of hydrogen bonding with the side chain oxygen of Gln262 by comparing the binding pose of compound **5b** obtained in that study to the binding poses of the corresponding acetamide derivatives and their respective bioactivities. Moreover, the same study shows the key role played by Asp181, Ser216 and Arg221 in the anchorage of compound **5b** to the active site. From our docking studies, we observed that most of our compounds did not perform further hydrogen bonding interactions apart from those required with the residues of the P-loop (see Figure S9). Interestingly, the compounds with more activity identified in our study are the only ones capable of performing an additional π – π interaction with Tyr46 in the active site (i.e., compounds **5** and **7**; see Figure S9) which, together with a second π – π interaction with Phe182, has been reported to be present in other complexes between PTP1B and potent inhibitors.^[51,52] Moreover, the requirement of an interaction with secondary binding site residues like Arg24 and Arg254 would contribute to the selectivity of the VS hits over the highly homologous T-cell protein tyrosine phosphatase (TCPTP).^[48,51,53] Thus, this indicates that there is still room for improvement in our VS protocol as seeking interactions with other residues in the PTP1B binding site may result in the identification of PTP1B inhibitors with higher potency and selectivity.

Experimental Section

1. RF model

To prepare the molecules for the application of the RF model, ChemAxon's Standardizer^[54] was used to generate canonical representations of each molecule and Morgan fingerprints of radius 2 were calculated with RDKit.^[55] The RF model was built and validated with a set of 4693 actives and 4676 decoys using the Python package Scikit-learn^[56] v0.17. The actives were obtained from ChEMBL^[57] and Reaxys^[59] and correspond to inhibitors of human PTP1B from those databases with bioactivity in the 1–15 range for pX, and whose activity was determined by measuring IC_{50} or K_{MW} -based decoys were obtained from the ZINC^[58] database using



Decoyfinder.^[59] Some parameters of the model were adjusted by individually evaluating their effect upon the performance of the model (see Figure S10). We chose to use 100 trees and to split the training and the test set into 80% and 20%, respectively. The model was validated via 5-fold cross-validation. The RF model performs well, with average precision and sensitivity values of 94.1% and 98.7% (see Table 2 for performance details). The output classification probabilities were calibrated using Platt scaling.^[60] Thanks to probability calibration, the RF model provides the predicted probability of a compound being active. It is worth noting that the probability threshold can be increased or decreased depending on the library size to obtain the desired amount of input compounds for the following step of the virtual screening.

2. Selection of the crystal structure used for docking

Crystal structures for PTP1B/inhibitor complexes were obtained by VHELIBS^[61] from PDB_REDO^[62] under the UniProt^[63] accession number for human PTP1B (i.e., P18031) and the fitting of the coordinates of the inhibitors and the binding site relative to their corresponding electron density map was analyzed. Protein–ligand interaction schemes from the PDBsum^[44] website were used to confirm the non-peptide and reversible character of the PTP1B inhibitor present in each complex. Furthermore, complexes with at least one mutation in their amino acid sequence were discarded. Thus, from the initial 72 structures corresponding to PTP1B–inhibitor complexes in the PDB_REDO,^[62] 11 were rejected by VHELIBS^[61] because either the inhibitor or the receptor binding site coordinates did not correctly fill their corresponding electron density map. Of the remaining 61 structures, 13 were discarded because the ligand was a peptide or covalently bound to PTP1B. From the remaining 48 structures, the one with PDB^[34,35] code 1Q6M^[36] (corresponding to the complex with an inhibitor that shows one of the lowest IC₅₀; that is, 13 nM) was selected as the target for protein–ligand docking.

3. Ligand setup for docking

Before docking, ligand molecules were prepared with Maestro's v10.7 LigPrep^[37,64] with default parameter values except for the following options: a) respect chiralities from input geometry when generating stereoisomers; b) use Epik^[65] v3.7 for ionization and tautomerization; c) use 7.0 as effective pH; and d) use 2.0 as pH tolerance for generated structures.

4. Protein preparation

The A chain of the crystal structure with PDB^[34,35] code 1Q6M^[36] was prepared by using Maestro's v10.7 Protein Preparation Wizard^[37,66] through the following procedure: a) remove original hydrogens; b) cap termini; c) generate ionization and tautomeric states of the ligand with Epik^[65] (here the original state of the ligand was selected because it showed the lowest state penalty and the highest hydrogen bond count; that is, state penalty: 0.17 Kcal mol⁻¹, Q: -4 and hydrogen bond count: 10); d) assign hydrogen bonds at pH 7 with PROPKA; e) use force field OPLS_2005 to minimize the structure at 0.30 Å; and f) remove all water molecules from the structure.

5. Grid generation

The grid for protein–ligand docking was generated with Maestro^[37] v10.7 by using default parameter values except for the following settings: a) the grid center coordinates were (0.25, 72.57, 70.0); b) aromatic hydrogens were included as hydrogen bond donors; c) halogens were included as acceptors; d) the inner box size was (10, 10, 10); e) the outer box size was (30, 30, 30); and f) hydrogen bond constraints were defined on the backbone nitrogen atoms of the residues Ser216, Ala217, Ile219, Gly220 and Arg221, as well as the side-chain nitrogen atoms of the residue Arg221 and the thiol group of the residue Cys215.

6. Docking

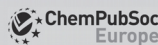
Protein–ligand docking was performed with Glide^[67] v7.2 by using default parameter values except for the following settings: a) SP precision; b) enhance planarity of conjugated π groups; c) include aromatic H as donors; d) include halogens as acceptors; e) write out at most 10 poses per ligand; f) include 50 poses per ligand in post-docking minimization; and g) require accomplishment of one hydrogen bond constraint.

7. Electrostatic similarity analysis

The software EON^[68] compares the poses for two different compounds by calculating Tanimoto coefficients associated either to their electrostatic potentials (i.e., Poisson–Boltzmann electrostatics and the coulombic part of the Poisson–Boltzmann electrostatics), to their shape, or to the combination of Poisson–Boltzmann electrostatics and their shape. The Poisson–Boltzmann electrostatics metric was used here, obtaining the electrostatic Tanimoto value (i.e., EON_ET_pb), which was in the $-1/5$ to 1 range (where a value of 1 corresponds to identical electrostatic potential overlap whereas negative values correspond to the overlap of opposite charges between the two poses). The validation set used in all electrostatic similarity comparisons was composed of 1424 actives and 1500 decoys. The actives were obtained from Reaxys^[59] and corresponded to inhibitors of human PTP1B with a bioactivity in the 1–15 range for pX and whose activity was determined by measuring IC₅₀ or K_i. MW-based decoys were obtained from the ZINC^[69] database using Decoyfinder.^[59] Prior to the analysis, the docking of the validation set was performed following the procedure described above.

To determine which crystallized ligands to use as references for the electrostatic similarity analysis, the PTP1B crystal structures containing ligands with IC₅₀ or K_i activity values between 1 and 100 nM (a total of 14 crystal structures; see Table 3) were obtained and superposed to the crystal structure with PDB^[34,35] code 1Q6M.^[36] The validation set was then used to perform 14 separate validations (i.e., one for each of the 14 crystallized ligands in the superposed crystal structures; see Figure S2) using the crystallized ligands as queries. In each validation, only the docked pose that presented the highest electrostatic Tanimoto with the corresponding query was kept for each library compound. After discarding the rest of the docking poses, histograms of the electrostatic Tanimoto coefficients for the actives and the decoys in the validation set were plotted, separating the actives into 3 groups depending on whether their pX value was lower than 4, between 4 and 7 or higher than 7 (see Figure S2). For 3 of the 14 initial crystallized actives (ligands in crystal structures with the PDB^[34,35] codes 2VEU,^[69] 2VEW,^[69] 2VEY,^[69] see Figure S2) the electrostatic Tanimoto value distribution of the actives was very similar to that of the decoys, so

Combined ligand- and receptor-based VS to identify PTP1B inhibitors



CHEM MED CHEM
Full Papers

no electrostatic Tanimoto value could be used as a cutoff to differentiate between the two groups, and for this reason these queries were discarded. The remaining 11 crystallized actives (ligands in crystal structures with the PDB^[34,35] codes 1Q6J^[36] 1Q6M^[36] 1Q6N^[36] 1Q6P^[36] 1Q6S^[36] 1Q6T^[36] 2FJN^[49] 2QBP^[70] 2QBQ^[70] 2ZMM^[71] 2ZN7^[71] see boldface PDB codes in Table 3 and Figure S2) were selected as queries for the in silico validation of the electrostatic similarity analysis.

8. Clustering

The HDBSCAN^[46] Clustering Library from Python was used for clustering. First, a distance matrix was developed using scikit-learn based on the "rogerstanimoto" metric and then HDBSCAN^[46] clustering was performed with the following settings: a) minimum cluster size: 2; b) metric: "precomputed".

9. Bioactivity assays

The colorimetric PTP1B Drug Discovery Kit (BML-AK822) was purchased from Enzo Life Sciences. The compounds were fully dissolved in DMSO at a concentration of 25 mM. For the inhibitory activity assay, each compound was dissolved in 1 × assay buffer to obtain the final concentrations of 100 μM and 10 μM in a volume of 45 μL per well with the respective DMSO concentrations of 0.4% and 0.04%. Controls with DMSO were prepared at these concentrations to ensure that DMSO had no effect on PTP1B activity (data not shown). After warming for 10 min to the assay temperature of 37 °C, 5 μL of human recombinant PTP1B stock solution (0.6 ng μM⁻¹) were added. The reaction was initiated with 50 μL of the substrate IR5 (insulin receptor β residues 1142–1153 with pY at position 1146) dissolved in 1 × assay buffer to a concentration of 60 μM. After an incubation period of 2 h at 37 °C the reaction was terminated by adding of 25 μL of Biomol Red Reagent. PTP1B dephosphorylates the IR5 substrate resulting in the release of orthophosphate, which was quantified after incubating for 30 min at 620 nm using the microtiter plate reader EON Microplate (BioTek, Vermont, USA). The maximum absorbance values were detected 10 min after exposure to Biomol Red reagent. In this initial bioactivity assay, only two measures per compound were performed (one at 10 μM and another one at 100 μM concentration; see Table 4). For the determination of IC₅₀ values of the two most active compounds, they were diluted in 1 × assay buffer to obtain the following final concentrations: 1000, 300, 100, 30, 10, 3, 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, and 0.0003 μM. Suramin was used as the positive control. The percentages of inhibition at each concentration were determined using triplicates and IC₅₀ values were calculated using a four parameter logistic regression (see Figure S7).

Acknowledgements

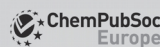
This study was supported by research grants 2014PFR-URV-B2-67 and 2015PFR-URV-B2-67 from the Universitat Rovira i Virgili, and also by grant AGL2013-40707-R from the Spanish Ministry of Economy and Competitiveness. A.G.'s contract is supported by grant 2015FL_B00655 from the Government of Catalonia. M.M. is a Serra Hunter research fellow. We thank OpenEye Scientific Software and ChemAxon Ltd. for kindly providing us with a software bursary to use their programs. This manuscript has been edited by the English language service of our university.

Conflict of interest

The authors declare no conflict of interest.

Keywords: inhibitors · obesity · protein tyrosine phosphatase · type 2 diabetes mellitus · virtual screening

- [1] International Diabetes Federation, <https://www.idf.org/> (accessed August 3, 2018).
- [2] World Health Organization, <http://www.who.int/> (accessed August 3, 2018).
- [3] L. Yao, O. Herlea-Pana, J. Heuser-Baker, Y. Chen, J. Barlic-Dicen, *J. Immunol. Res.* **2014**, *2014*, 181450.
- [4] E. K. Anderson, D. A. Gutierrez, A. H. Hasty, *Curr. Opin. Lipidol.* **2010**, *21*, 172–177.
- [5] B. K. Surmi, A. H. Hasty, *Vascul. Pharmacol.* **2010**, *52*, 27–36.
- [6] N. Dali-Youcef, M. Mecili, R. Ricci, E. Andrès, *Ann. Med.* **2013**, *45*, 242–253.
- [7] U. J. Jung, M.-S. Choi, *Int. J. Mol. Sci.* **2014**, *15*, 6184–6223.
- [8] Y. S. Lee, H. Morinaga, J. J. Kim, W. Lagakos, S. Taylor, M. Keshwani, G. Perkins, H. Dong, A. G. Kayali, I. R. Sweet, J. Olefsky, *Cell* **2011**, *153*, 413–425.
- [9] N. Esser, S. Legrand-Poels, J. Piette, A. J. Scheen, N. Paquot, *Diabetes Res. Clin. Pract.* **2014**, *105*, 141–150.
- [10] V. R. Richardson, K. A. Smith, A. M. Carter, *Immunobiology* **2013**, *218*, 1497–1504.
- [11] L. F. Van Gaal, I. L. Mertens, C. E. De Block, *Nature* **2006**, *444*, 875–880.
- [12] M. Y. Donath, É. Dalmas, N. S. Sauter, M. Böni-Schnetzler, *Cell Metab.* **2013**, *17*, 860–872.
- [13] K. G. M. M. Alberti, R. H. Eckel, S. M. Grundy, P. Z. Zimmet, J. I. Cleeman, K. A. Donato, J.-C. Fruchart, W. P. T. James, C. M. Loria, S. C. Smith, Jr., International Diabetes Federation Task Force on Epidemiology and Prevention, National Heart, Lung and Blood Institute, American Heart Association, World Heart Federation, International Atherosclerosis Society, International Association for the Study of Obesity, *Circulation* **2009**, *120*, 1640–1645.
- [14] S. Guo, *J. Endocrinol.* **2014**, *220*, T1–T23.
- [15] R. Di Paola, L. Frittitta, G. Miscio, M. Bozzali, R. Baratta, M. Centra, D. Spampinato, M. G. Santagati, T. Ercolino, C. Cisternino, T. Soccio, S. Mastrianno, V. Tassi, P. Almgren, A. Pizzuti, R. Vigneri, V. Trischitta, *Am. J. Hum. Genet.* **2002**, *70*, 806–812.
- [16] N. T. Lam, S. D. Covey, J. T. Lewis, S. Oosman, T. Webber, E. C. Hsu, A. T. Cheung, T. J. Kieffer, *J. Mol. Endocrinol.* **2006**, *36*, 163–174.
- [17] C. D. Morrison, C. L. White, Z. Wang, S.-Y. Lee, D. S. Lawrence, W. T. Cefalu, Z.-Y. Zhang, T. W. Gettys, *Endocrinology* **2007**, *148*, 433–440.
- [18] S. Koren, I. G. Fantus, *Best Pract. Res. Clin. Endocrinol. Metab.* **2007**, *21*, 621–640.
- [19] M. F. Cicirelli, N. K. Tonks, C. D. Diltz, J. E. Weiel, E. H. Fischer, E. G. Krebs, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5514–5518.
- [20] B. J. Goldstein, A. Bittner-Kowalczyk, M. F. White, M. Harbeck, *J. Biol. Chem.* **2000**, *275*, 4283–4289.
- [21] A. Cheng, N. Uetani, P. D. Simoncic, V. P. Chaubey, A. Lee-Loy, C. J. McGlade, B. P. Kennedy, M. L. Tremblay, *Dev. Cell* **2002**, *2*, 497–503.
- [22] J. M. Zabolotny, K. K. Benice-Hanulec, A. Stricker-Krongrad, F. Hajj, Y. Wang, Y. Minokoshi, Y.-B. Kim, J. K. Elmquist, L. A. Tartaglia, B. B. Kahn, B. G. Neel, *Dev. Cell* **2002**, *2*, 489–495.
- [23] P.-A. Thiebaut, M. Besnier, E. Gomez, V. Richard, *J. Mol. Cell. Cardiol.* **2016**, *101*, 50–57.
- [24] A. K. Saxena, G. Pandey, S. Gupta, A. B. Singh, A. K. Srivastava, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2320–2323.
- [25] A. P. Combs, *J. Med. Chem.* **2010**, *53*, 2333–2344.
- [26] W. B. Floriano, N. Vaidehi, G. Zamanakos, W. A. Goddard, *J. Med. Chem.* **2004**, *47*, 56–71.
- [27] K. Bharatham, N. Bharatham, K. W. Lee, *Arch. Pharmacol. Res.* **2007**, *30*, 533–542.
- [28] H. Park, B. R. Bhattarai, S. W. Ham, H. Cho, *Eur. J. Med. Chem.* **2009**, *44*, 3280–3284.



- [29] P. S. Rao, C. Muvva, K. Geethanjali, S. B. Bastipati, R. Kalashikam, *Bioinformatics* **2012**, *8*, 834–837.
- [30] M. V. Reddy, C. Ghadiyaram, S. K. Panigrahi, S. Hosahalli, L. N. Mangamoori, *Mini Rev. Med. Chem.* **2013**, *13*, 1602–1606.
- [31] Y. Ma, Y.-Y. Jin, Y.-L. Wang, R.-L. Wang, X.-H. Lu, D.-X. Kong, W.-R. Xu, *Chem. Biol. Drug Des.* **2014**, *83*, 697–709.
- [32] V. M. Balaramnavar, R. Srivastava, N. Rahuja, S. Gupta, A. K. Rawat, S. Varshney, H. Chandasana, Y. S. Chhonker, P. K. Doharey, S. Kumar, S. Gautam, S. P. Srivastava, R. S. Bhatta, J. K. Saxena, A. N. Gaikwad, A. K. Srivastava, A. K. Saxena, *Eur. J. Med. Chem.* **2014**, *87*, 578–594.
- [33] S. Chandra, J. Pandey, A. K. Tamrakar, M. I. Siddiqi, *J. Mol. Graphics Modell.* **2017**, *71*, 242–256.
- [34] RCSB Protein Data Bank (PDB), <http://www.rcsb.org> (accessed August 3, 2018).
- [35] H. M. Berman, *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [36] G. Scapin, S. B. Patel, J. W. Becker, Q. Wang, C. Desponts, D. Waddleton, K. Skorey, W. Cromlish, C. Bayly, M. Therien, J. Y. Gauthier, C. S. Li, C. K. Lau, C. Ramachandran, B. P. Kennedy, E. Asante-Appiah, *Biochemistry* **2003**, *42*, 11451–11459.
- [37] Schrödinger Release 2016–2013: Maestro, Schrödinger, LLC, New York, NY, **2016**.
- [38] Specs, <http://www.specs.net> (accessed August 3, 2018).
- [39] Reaxys, <https://www.reaxys.com/> (accessed August 3, 2018).
- [40] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, G. Pujadas, *Methods* **2015**, *71*, 58–63.
- [41] M. A. Ghattas, N. Atatreh, E. V. Bichenkova, R. A. Bryce, *J. Mol. Graphics Modell.* **2014**, *52*, 114–123.
- [42] A. C. Wallace, R. A. Laskowski, J. M. Thornton, *Protein Eng.* **1995**, *8*, 127–134.
- [43] K. Stierand, P. C. Maass, M. Rarey, *Bioinformatics* **2006**, *22*, 1710–1716.
- [44] PDBsum, www.ebi.ac.uk/pdbsum/ (accessed August 3, 2018).
- [45] Marvin 16.10.10.0, **2016**, ChemAxon, <http://www.chemaxon.com> (accessed August 3, 2018).
- [46] L. McInnes, J. Healy, S. Astels, *J. Open Source Softw.* **2017**, *2(11)*, 205; DOI: <http://doi.org/10.21105/joss.00205>.
- [47] I. Letunic, P. Bork, *Nucleic Acids Res.* **2016**, *44*, W242–5.
- [48] Y. A. Puius, Y. Zhao, M. Sullivan, D. S. Lawrence, S. C. Almo, Z. Y. Zhang, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 13420–13425.
- [49] E. Asante-Appiah, S. Patel, C. Desponts, J. M. Taylor, C. Lau, C. Dufresne, M. Therien, R. Friesen, J. W. Becker, Y. Leblanc, B. P. Kennedy, G. Scapin, *J. Biol. Chem.* **2006**, *281*, 8010–8015.
- [50] G. Pandey, A. K. Saxena, *J. Chem. Inf. Model.* **2006**, *46*, 2579–2590.
- [51] S. Gupta, K. Varshney, R. Srivastava, N. Rahuja, A. K. Rawat, A. K. Srivastava, A. K. Saxena, *MedChemComm* **2013**, *4*, 1382–1387.
- [52] S. R. Klopfenstein, A. G. Evdokimov, A.-O. O. Colson, N. T. Fairweather, J. J. Neuman, M. B. Maier, J. L. Gray, G. S. Gerwe, G. E. Stake, B. W. Howard, J. A. Farmer, M. E. Pokross, T. R. Downs, B. Kasibhatla, K. G. Peters, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1574–1578.
- [53] D. Patel, M. Jain, S. R. Shah, R. Bahekar, P. Jadav, A. Joharapurkar, N. Dhanesha, M. Shaikh, K. V. V. M. Sairam, P. Kapadnis, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 1111–1117.
- [54] Standardizer 16.10.10.0, **2016**, ChemAxon, <http://www.chemaxon.com> (accessed August 3, 2018).
- [55] RDKit: Open-source cheminformatics, <http://www.rdkit.org> (accessed August 3, 2018).
- [56] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- [57] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J. P. Overington, *Nucleic Acids Res.* **2014**, *42*, D1083–90.
- [58] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.
- [59] A. Cereto-Massagué, L. Guasch, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, *Bioinformatics* **2012**, *28*, 1661–1662.
- [60] J. Platt, *Adv. Large Margin Classif.* **1999**, *10*, 61–74.
- [61] A. Cereto-Massagué, M. J. Ojeda, R. P. Joosten, C. Valls, M. Mulero, M. J. Salvado, A. Arola-Arnal, L. Arola, S. Garcia-Vallvé, G. Pujadas, *J. Cheminf.* **2013**, *5*, 36.
- [62] PDB_REDO, A databank of updated and optimized X-ray structure models: http://www.cmbi.ru.nl/pdb_redo/ (accessed August 3, 2018).
- [63] A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-Ajee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georgioudi, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuhe, M. Doche, D. Dornevili, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuer-mann, E. Gasteiger, S. Gehant, V. Gerlitsen, A. Gos, N. Guaz-Gumowski, U. Hinz, C. Hulo, F. Junco, G. Keller, V. Lara, P. Lemerrier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Noupkiel, S. Paesano, I. Pedruzzi, S. Pilboud, M. Pozzato, M. Pruess, C. Rivoire, B. Roehert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognoli, L. Verbrugue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, J. Zhang, *Nucleic Acids Res.* **2017**, *45*, D158–D169.
- [64] Schrödinger Release 2016–2013: LigPrep, Schrödinger, LLC, New York, NY, **2016**.
- [65] Schrödinger Release 2016–2013: Epik, Schrödinger, LLC, New York, NY, **2016**.
- [66] Schrödinger Release 2016–2013: Schrödinger Suite 2016–2013 Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2016.
- [67] Schrödinger Release 2016–2013: Glide, Schrödinger, LLC, New York, NY, **2016**.
- [68] EON 2.2.0.5: OpenEye Scientific Software, Santa Fe, NM (USA), <http://www.eyesopen.com> (accessed August 3, 2018).
- [69] B. Douty, B. Wayland, P. J. Ala, M. J. Bower, J. Pruitt, L. Bostrom, M. Wei, R. Klabe, L. Gonville, R. Wynn, T. C. Burn, P. C. C. Liu, A. P. Combs, E. W. Yue, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 66–71.
- [70] D. P. Wilson, Z.-K. K. Wan, W.-X. X. Xu, S. J. Kirincich, B. C. Follows, D. Joseph-McCarthy, K. Foreman, A. Moretto, J. Wu, M. Zhu, E. Binnun, Y.-L. L. Zhang, M. Tam, D. V. Erbe, J. Tobin, X. Xu, L. Leung, A. Shilling, S. Y. Tam, T. S. Mansour, J. Lee, *J. Med. Chem.* **2007**, *50*, 4681–4698.
- [71] Z.-K. Wan, J. Lee, R. Hotchandani, A. Moretto, E. Binnun, D. P. Wilson, S. J. Kirincich, B. C. Follows, M. Ipek, W. Xu, D. Joseph-McCarthy, Y.-L. Zhang, M. Tam, D. V. Erbe, J. F. Tobin, W. Li, S. Y. Tam, T. S. Mansour, J. Wu, *ChemMedChem* **2008**, *3*, 1525–1529.

Manuscript received: April 23, 2018
Revised manuscript received: July 5, 2018
Accepted manuscript online: July 19, 2018
Version of record online: August 23, 2018

Supplementary material for

Combined ligand- and receptor-based virtual screening methodology to identify structurally diverse protein tyrosine phosphatase 1B inhibitors

Aleix Gimeno,^[a] Andrea Ardid-Ruiz,^[b] María José Ojeda-Montes,^[a] Sarah Tomás-Hernández,^[a] Adrià Cereto-Massagué,^[a] Raúl Beltrán-Debón,^[a] Miquel Mulero,^[a] Cristina Valls,^[a] Gerard Aragonès,^[b] Manuel Suárez,^[b] Gerard Pujadas,^{*,[a, c]} and Santiago Garcia-Vallvé^[a, c]

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]Nutrigenomics Research Group, Department of Biochemistry and Biotechnology, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[c]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

Table S1. Purity and exact mass of the 20 compounds whose bioactivities have been tested.

Compound	Specs ID	Purity (%)	Formula	Pseudo molecular ion adduct	Theoretical m/z	Measured m/z
1	AG-205/07930063	98.98	C18 H16 N2 O3 S	[M+H] ⁺	341.0954	341.0999
2	AM-807/41462317	98.92	C18 H12 F3 N3 O5 S	[M+H] ⁺	440.0523	440.052
3	AI-204/43362778	97.92	C16 H14 N2 O4 S2	[M+H] ⁺	363.0468	363.0466
4	AG-690/11836029	98.43	C12 H8 Br Cl O2 S	[M] ⁺	329.9111	329.9102
5	AF-399/15030257	97.77	C17 H10 N4 O4	[M+H] ⁺	335.0775	335.0766
6	AF-399/15285024	98.72	C21 H22 N2 O4	[M+H] ⁺	367.1652	367.1663
7	AK-968/41025519	99.64	C21 H16 N2 O4 S	[M+H] ⁺	393.0904	393.095
8	AG-690/40107322	95.29	C18 H16 Br2 F N7 O2	[M+H] ⁺	539.9789	539.9781
9	AJ-077/33270018	97.56	C22 H14 N4 O4	[M+H] ⁺	399.1088	399.1092
10	AH-487/41936905	96.38	C9 H8 I2 O4	[M+NH4] ⁺	451.8850	451.884
11	AE-641/42119771	99.65	C13 H10 Cl N O5 S	[M+H] ⁺	328.0041	328.003
12	AK-968/15361893	99.29	C19 H20 Cl N O5 S	[M+H] ⁺	410.0823	410.0812
13	AG-205/37175018	87.96	C16 H16 N2 O6	[M+H] ⁺	333.1081	333.1074
14	AF-886/30566062	99.27	C16 H10 Cl N3 O3	[M+H] ⁺	328.0483	328.0475
15	AG-205/37169028	96.02	C18 H12 Cl F N2 O3	[M+H] ⁺	359.0593	359.0613
16	AK-968/11197127	99.33	C19 H16 Cl N3 O3	[M+H] ⁺	370.0953	370.0994
17	AF-962/02757026	93.56	C13 H10 N2 O5 S2	[M+H] ⁺	339.0104	339.0147
18	AM-879/12634007	96.84	C16 H8 Cl3 N O2 S2	[M+H] ⁺	415.9135	415.9108
19	AG-690/36826026	96.37	C11 H8 Br N O3 S2	[M+H] ⁺	345.9202	345.9189
20	AK-968/37202023	98.16	C14 H6 Br2 F3 N3 O2	[M+H] ⁺	463.8852	463.8832

Manuscript 2

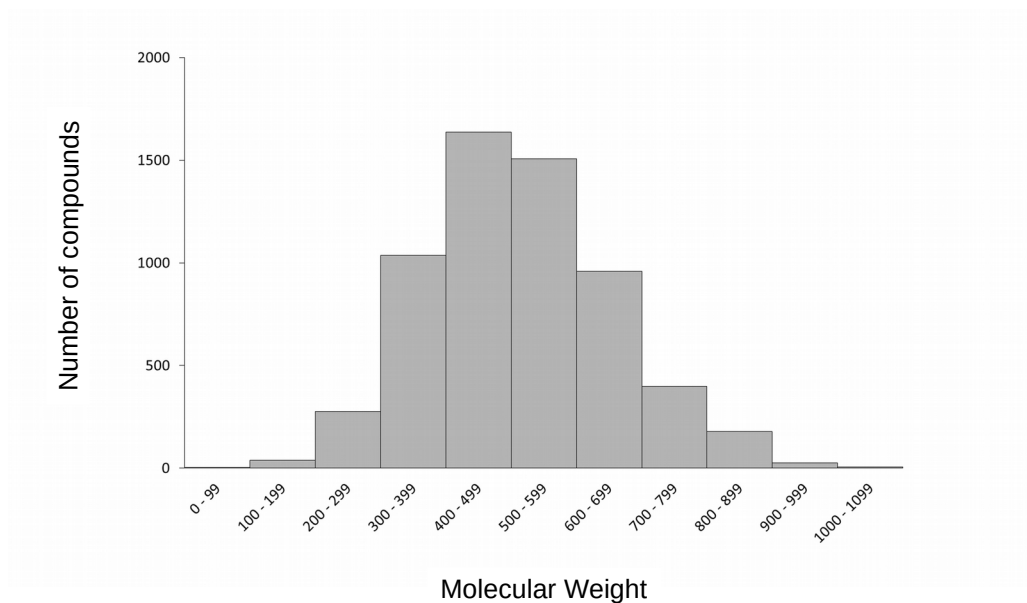


Figure S1. Histogram representation of the molecules used as references to establish a MW range for the MW filter.

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

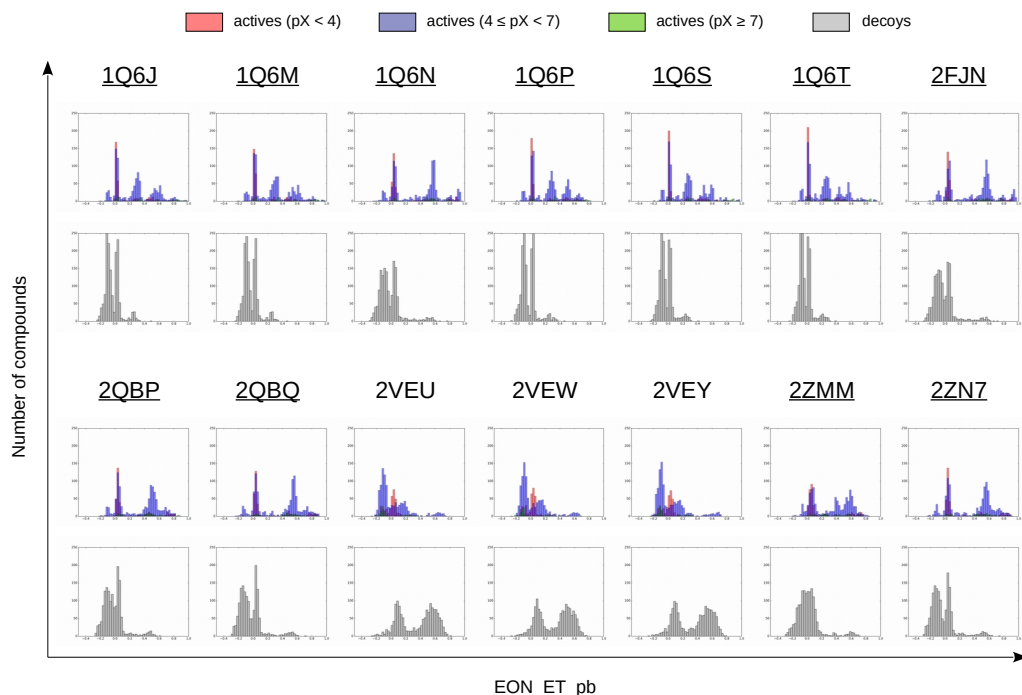
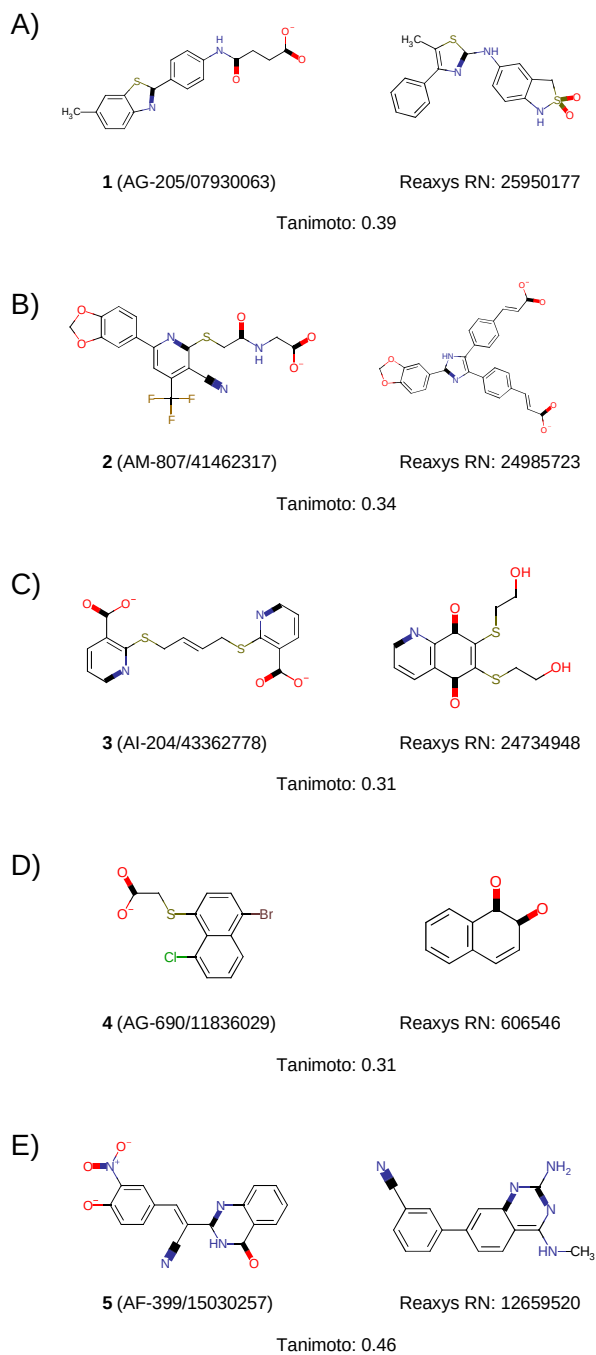


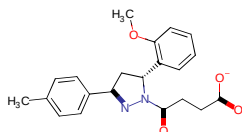
Figure S2. Histogram representation of the highest electrostatic Tanimoto (i.e. EON_ET_pb) values obtained in the comparison of the validation set to each query. For each query, the PDB^[34,35] code of the crystal structure from which it was obtained and two histograms are shown: one corresponding to the actives and one corresponding to the decoys. In the actives histogram, actives with pX lower than 4 are in red, actives with pX from 4 to 7 are in blue and actives with pX higher than 7 are in green. In the decoys histogram, decoys are in gray. The 11 queries selected for the *in silico* validation of the electrostatic similarity analysis are underlined.

Manuscript 2

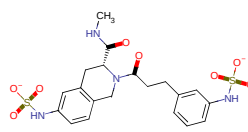


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

F)



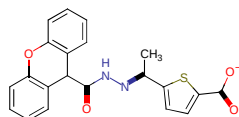
6 (AF-399/15285024)



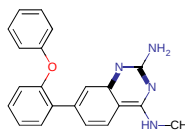
Reaxys RN: 25330632

Tanimoto: 0.25

G)



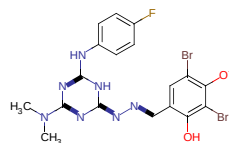
7 (AK-968/41025519)



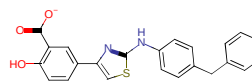
Reaxys RN: 12601206

Tanimoto: 0.27

H)



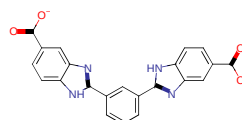
8 (AG-690/40107322)



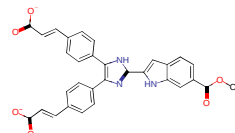
Reaxys RN: 12509120

Tanimoto: 0.30

I)



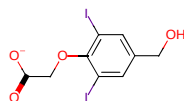
9 (AJ-077/33270018)



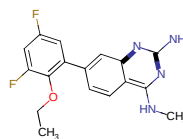
Reaxys RN: 24985689

Tanimoto: 0.55

J)



10 (AH-487/41936905)

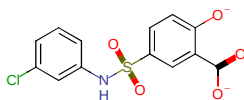


Reaxys RN: 12601276

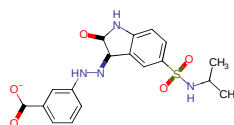
Tanimoto: 0.23

Manuscript 2

K)



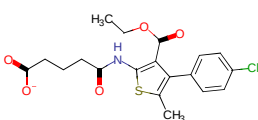
11 (AE-641/42119771)



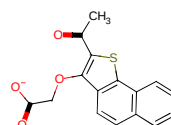
Reaxys RN: 13005572

Tanimoto: 0.43

L)



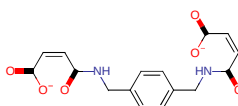
12 (AK-968/15361893)



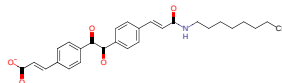
Reaxys RN: 28048260

Tanimoto: 0.39

M)



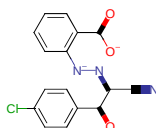
13 (AG-205/37175018)



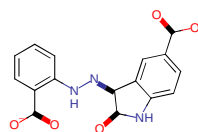
Reaxys RN: 24985673

Tanimoto: 0.39

N)



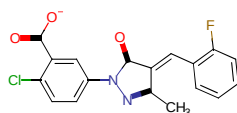
14 (AF-886/30566062)



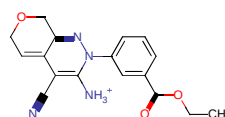
Reaxys RN: 13005538

Tanimoto: 0.48

O)



15 (AG-205/37169028)

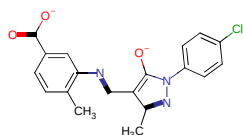


Reaxys RN: 13942891

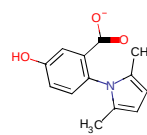
Tanimoto: 0.34

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

P)



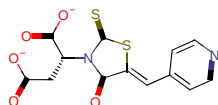
16 (AK-968/11197127)



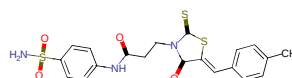
Reaxys RN: 15419640

Tanimoto: 0.34

Q)



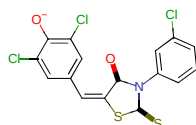
17 (AF-962/02757026)



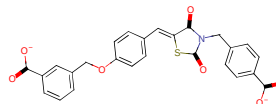
Reaxys RN: 25432438

Tanimoto: 0.31

R)



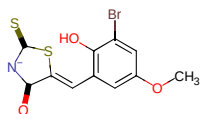
18 (AM-879/12634007)



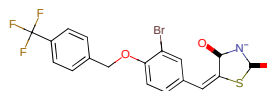
Reaxys RN: 26119874

Tanimoto: 0.34

S)



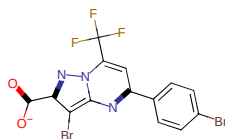
19 (AG-690/36826026)



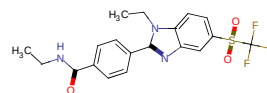
Reaxys RN: 25232528

Tanimoto: 0.31

T)



20 (AK-968/37202023)



Reaxys RN: 15705274

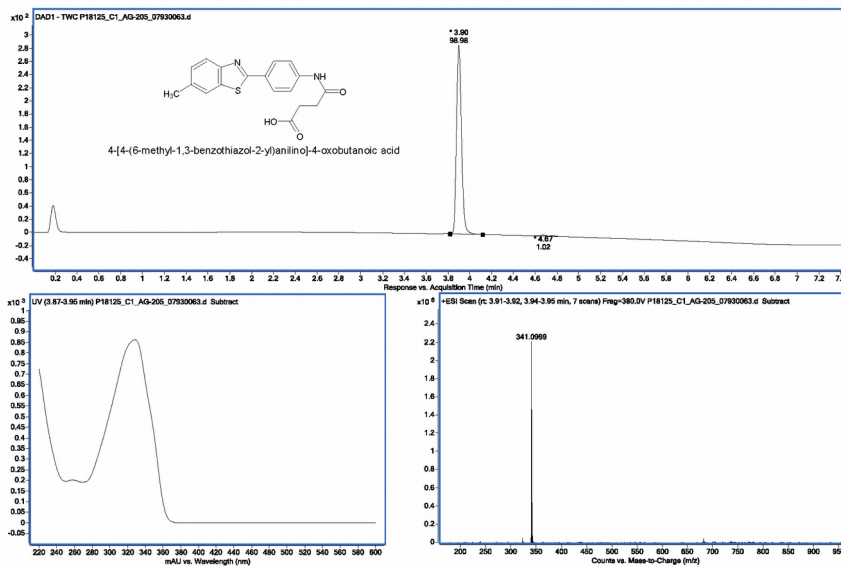
Tanimoto: 0.34

Manuscript 2

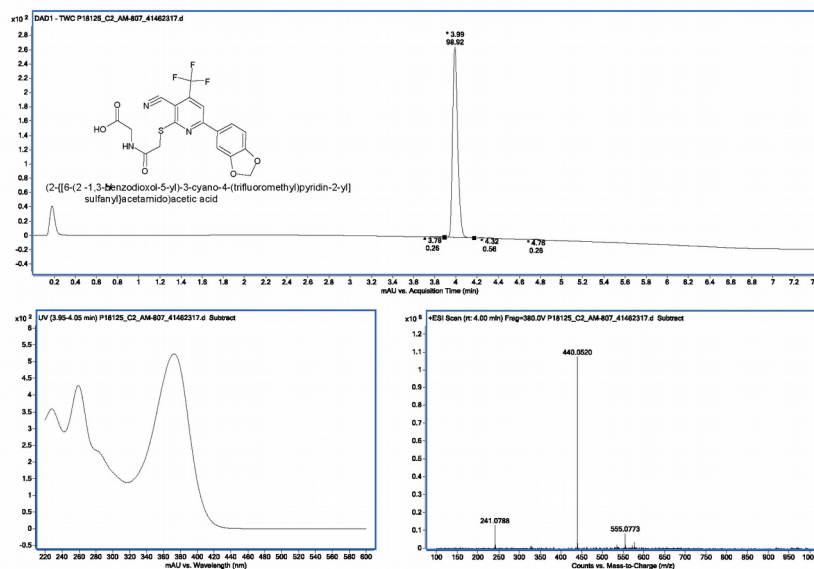
Figure S3. In each panel, a hit compound is represented in 2D, together with the most similar active compound that was used during the clustering. In each case, the most similar active compound is labeled with its Reaxys¹ Registry Number. The Tanimoto similarity value resulting from the comparison of the OpenEyePath FP of both compounds is shown below. MarvinSketch² was used to draw the structures. The protonation state of each compound corresponds to the protonation state of the docked pose selected for that compound.

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

A)

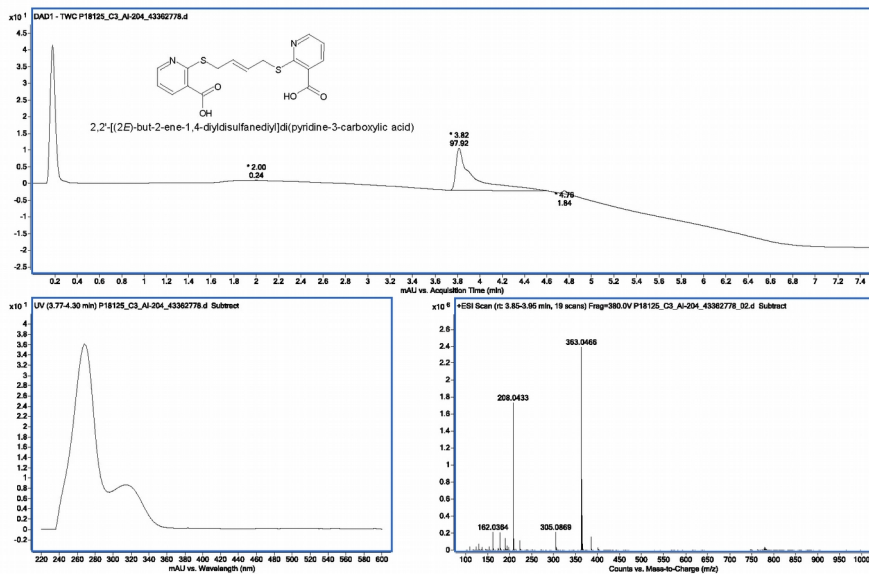


B)

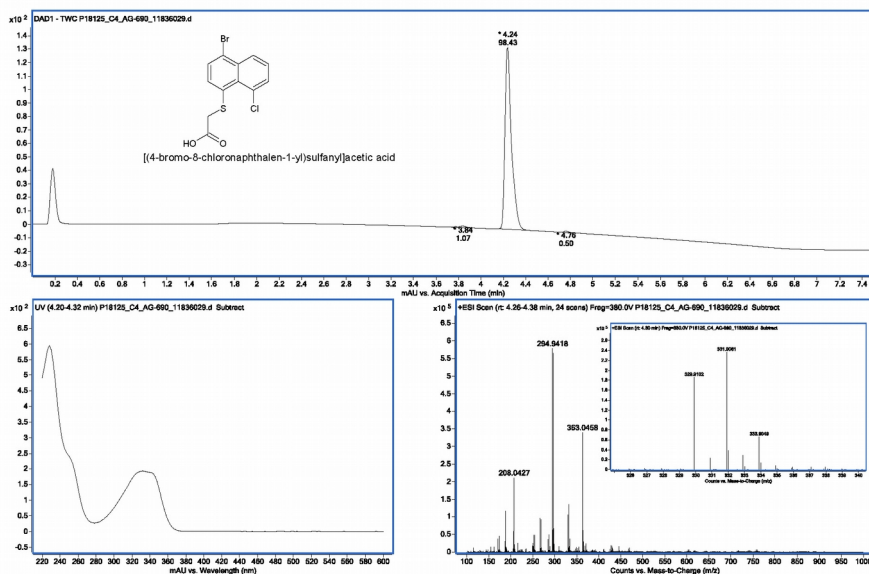


Manuscript 2

C)

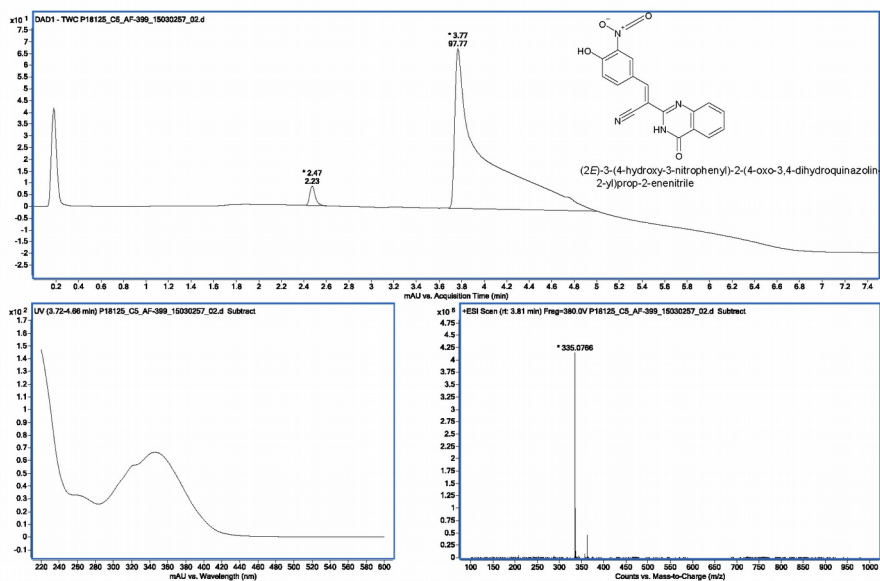


D)

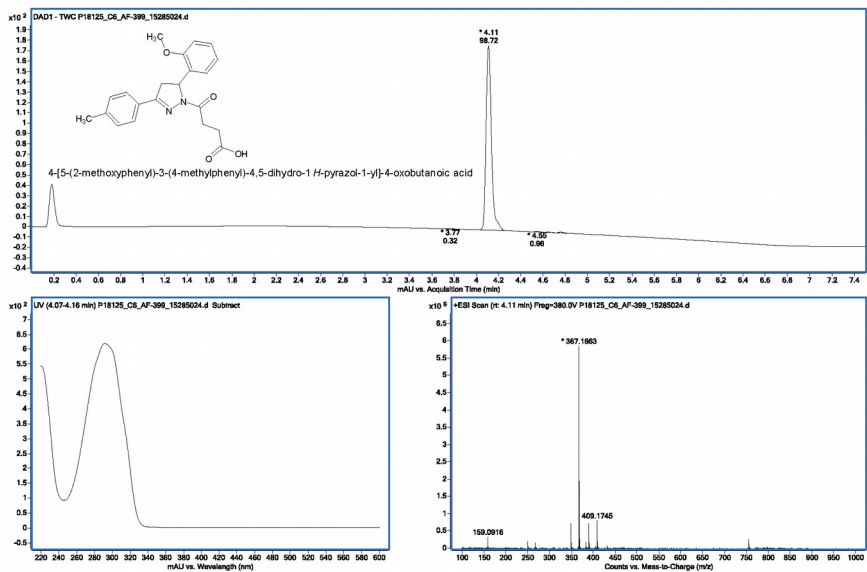


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

E)

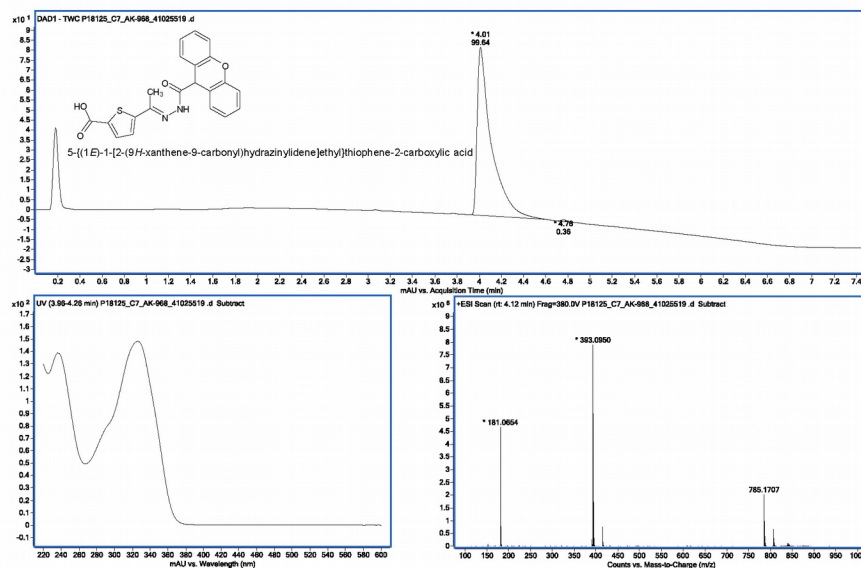


F)

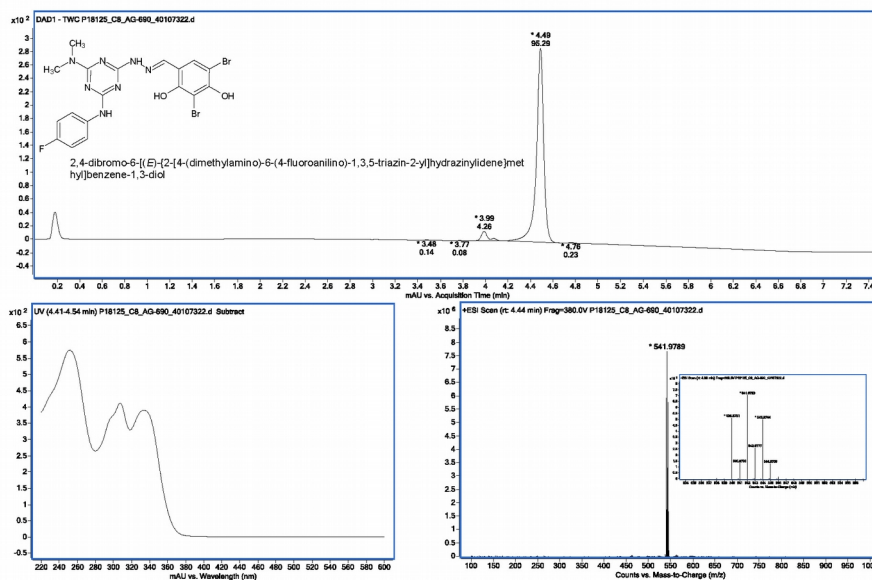


Manuscript 2

G)

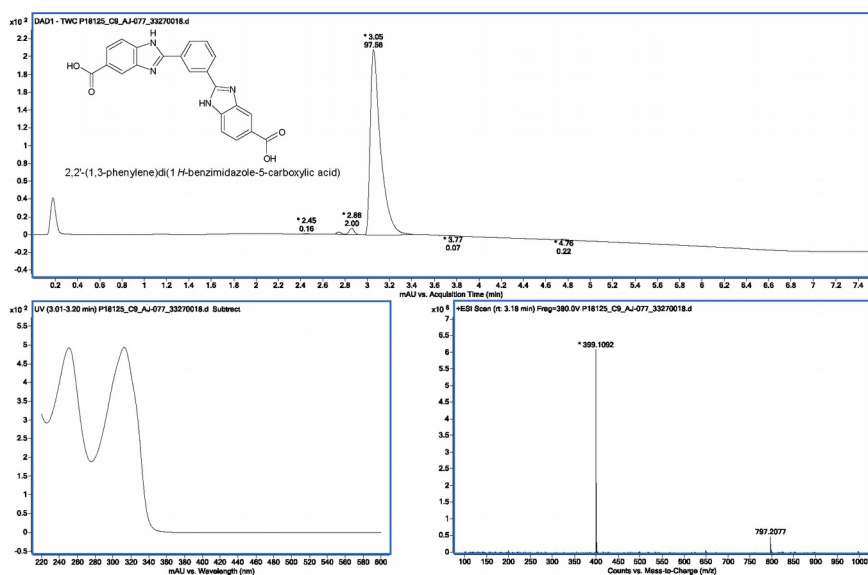


H)

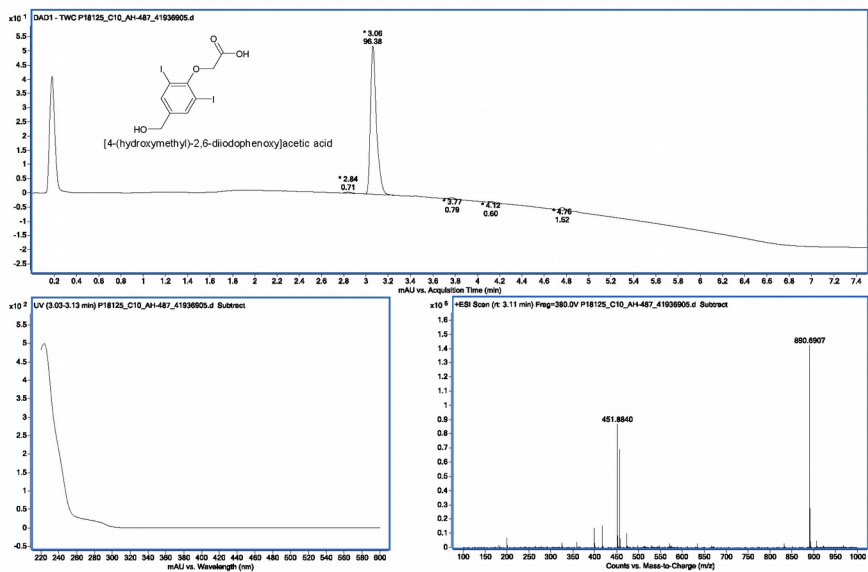


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

I)

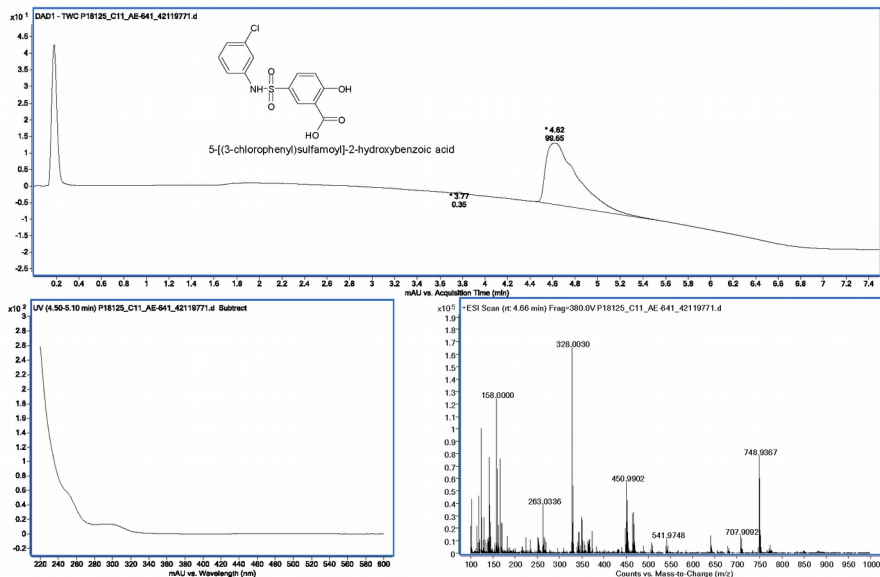


J)

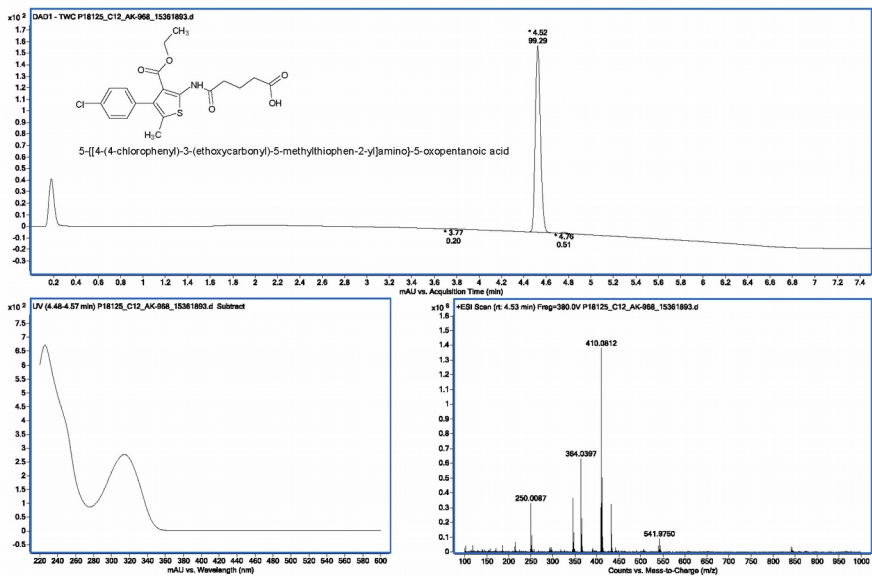


Manuscript 2

K)

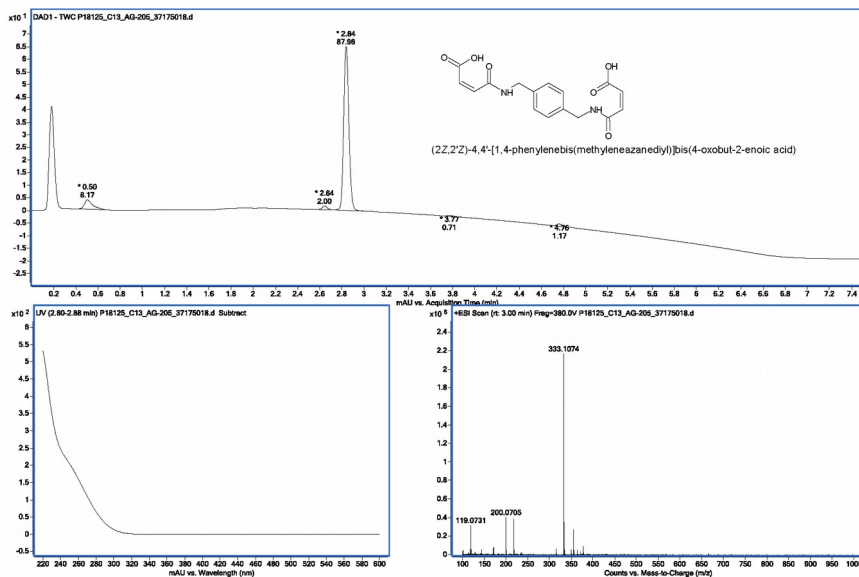


L)

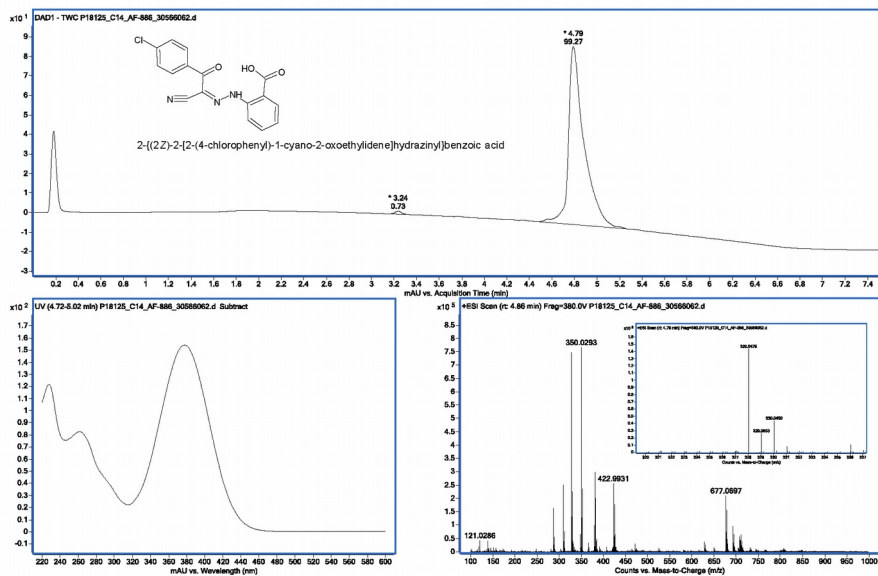


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

M)

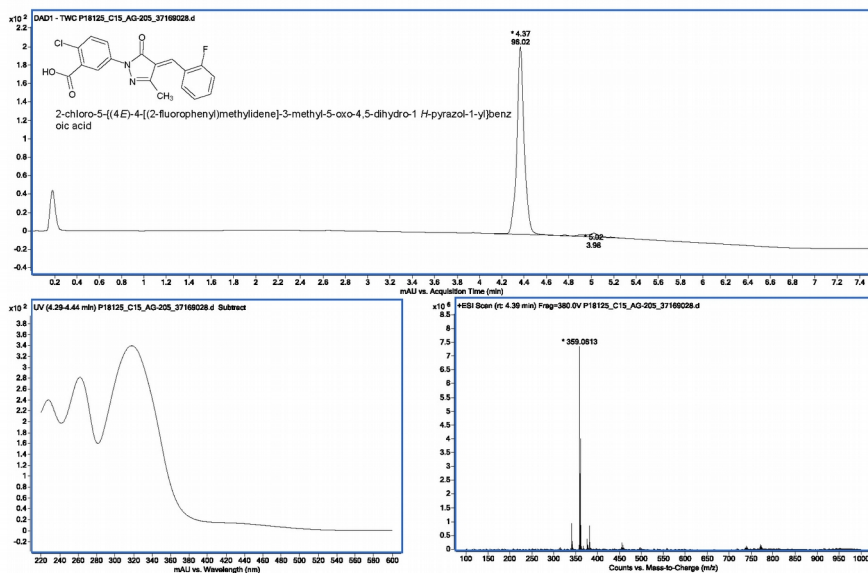


N)

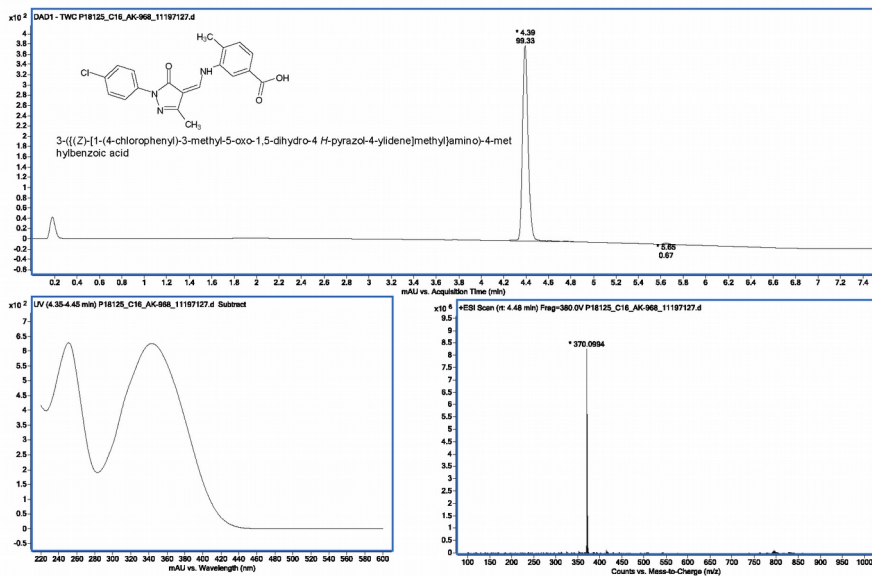


Manuscript 2

O)

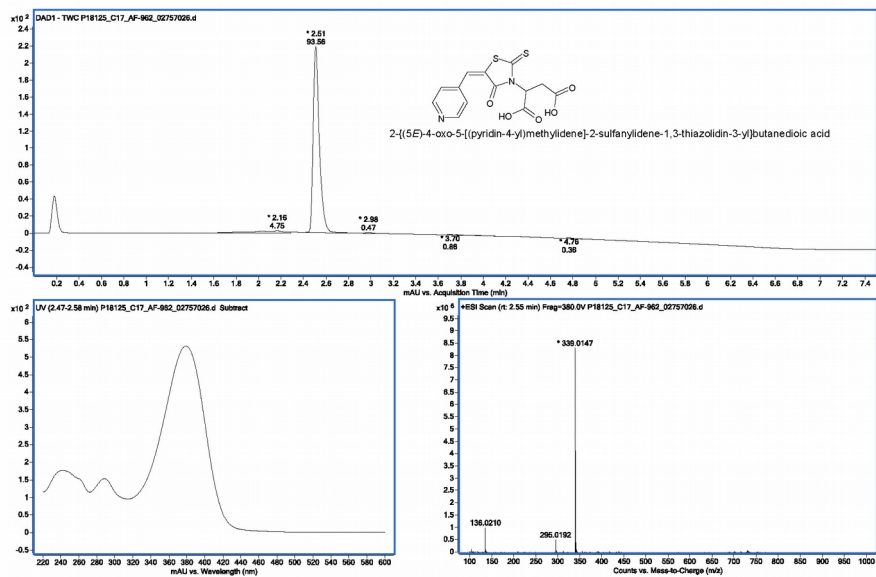


P)

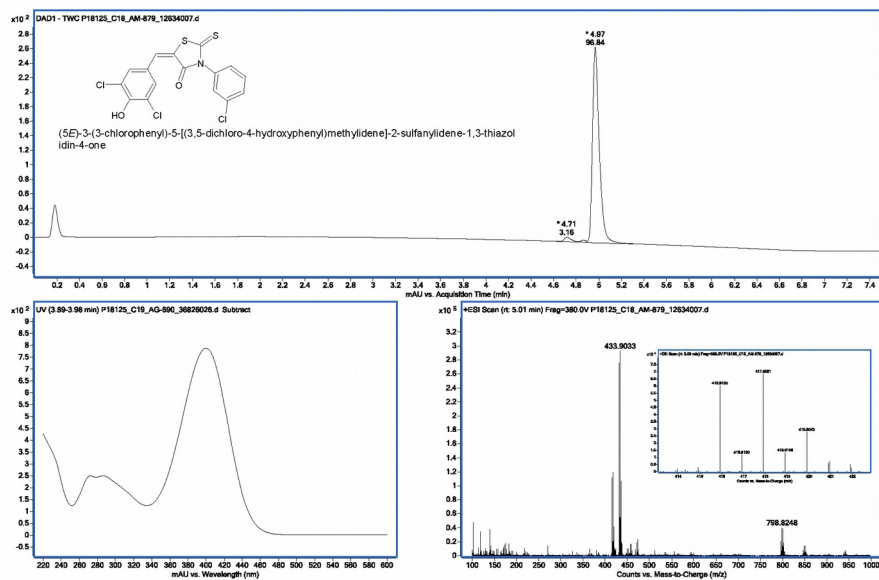


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

Q)

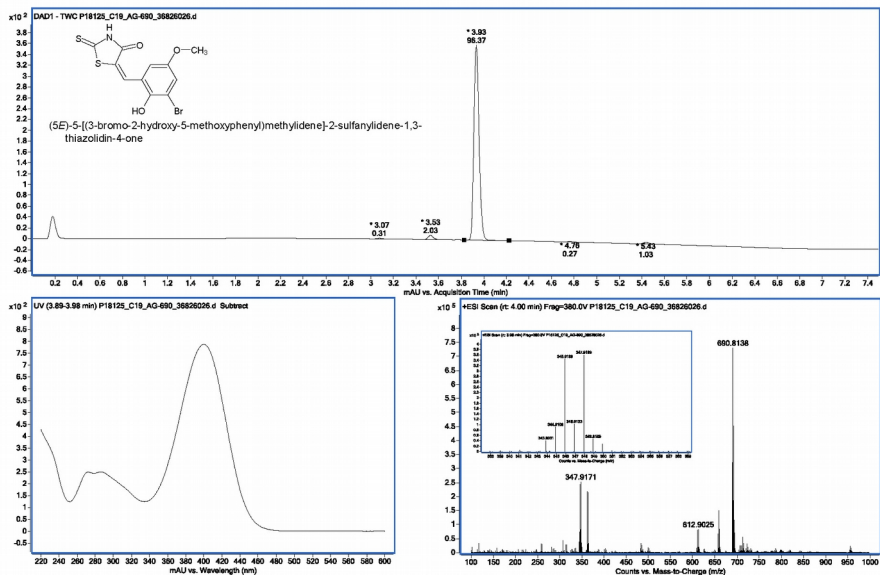


R)

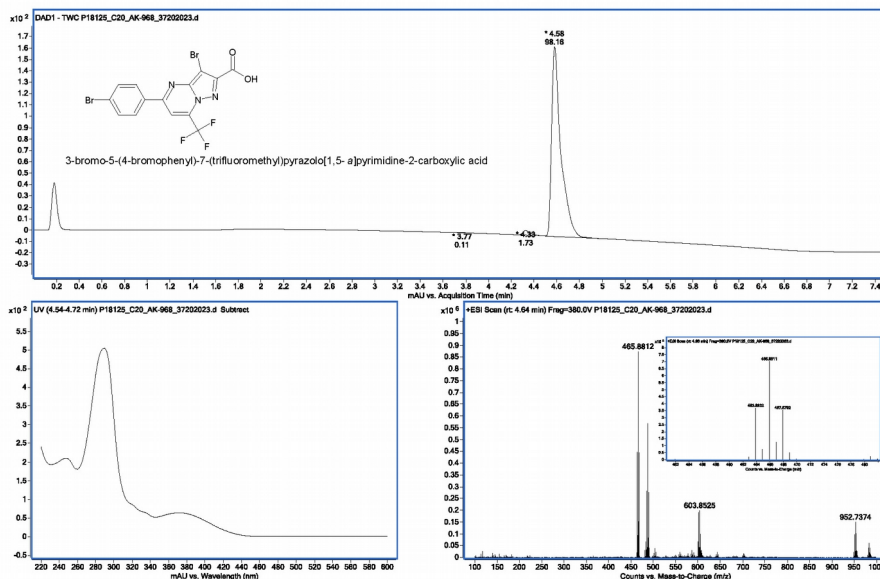


Manuscript 2

S)



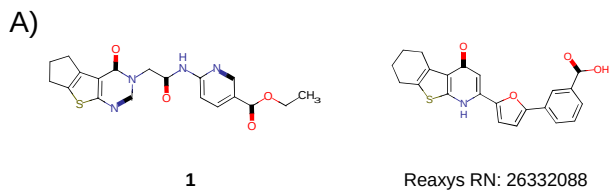
T)



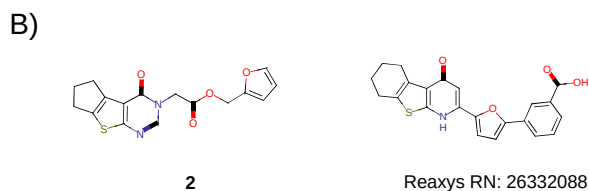
Combined ligand- and receptor-based VS to identify PTP1B inhibitors

Figure S4. This figure shows the LC-UV/Vis and Q-TOF spectra for the 20 compounds whose bioactivities have been tested. Panels A-T refer to compounds **1-20**, respectively. The following information is provided for each tested compound: a) the UV/Vis relative area percentage (calculated without considering the solvent peak) on top of each chromatographic peak (with the corresponding retention time); b) the total wavelength chromatograms (220-600 nm) from the LC-UV/Vis; c) the UV/Vis spectrum from 220 to 600 nm; and d) the mass spectrum from 100 to 1000 m/z.

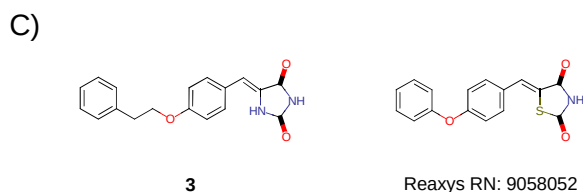
Manuscript 2



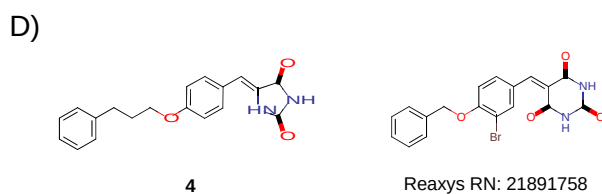
Tanimoto: 0.32



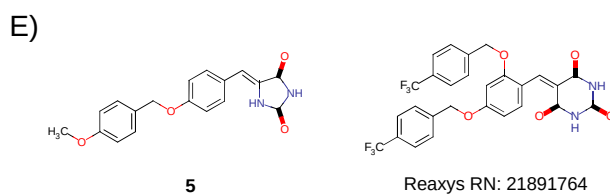
Tanimoto: 0.32



Tanimoto: 0.43



Tanimoto: 0.46



Tanimoto: 0.43

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

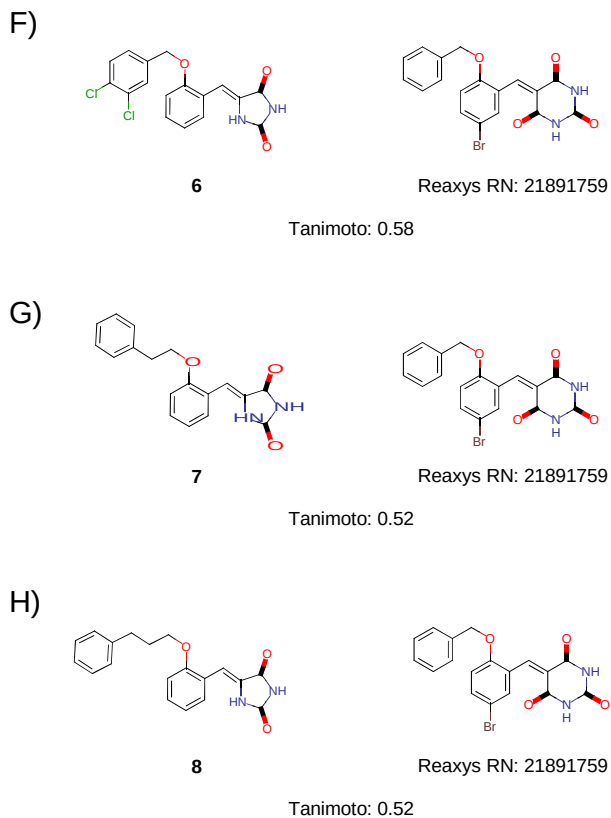
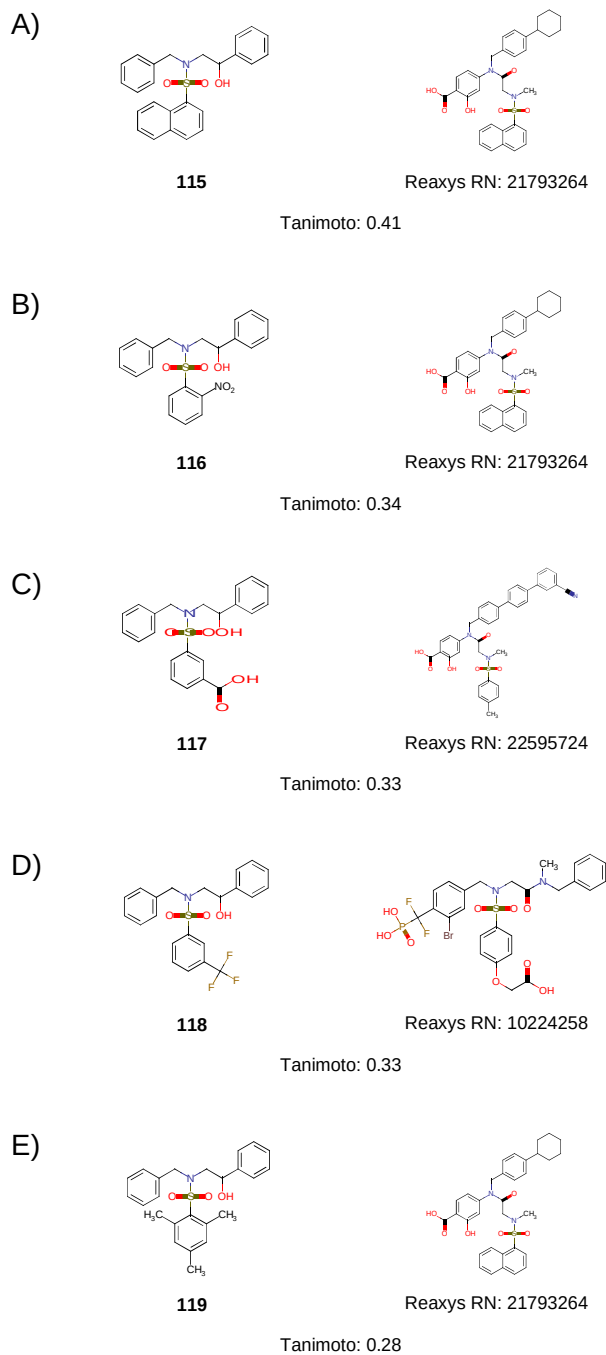


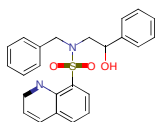
Figure S5. In each panel, a hit compound reported by Ma *et al.*³ is represented in 2D, together with the most similar active compound that was used for clustering in the current study. In each case, the hit compound is labeled with its corresponding code in the original publication and the most similar active compound is labeled with its Reaxys¹ Registry Number. The Tanimoto similarity value resulting from the comparison of the OpenEyePath FP of both compounds is shown below. MarvinSketch² was used to draw the structures.

Manuscript 2

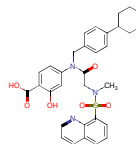


Combined ligand- and receptor-based VS to identify PTP1B inhibitors

F)



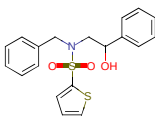
120



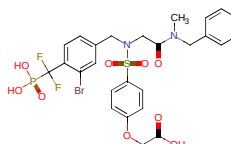
Reaxys RN: 21793265

Tanimoto: 0.41

G)



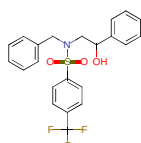
121



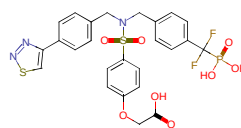
Reaxys RN: 10224258

Tanimoto: 0.25

H)



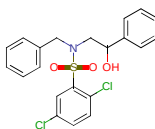
127



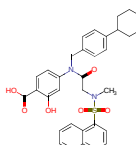
Reaxys RN: 10222316

Tanimoto: 0.38

I)



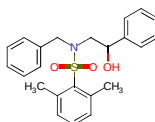
130



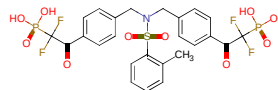
Reaxys RN: 21793264

Tanimoto: 0.34

J)



131



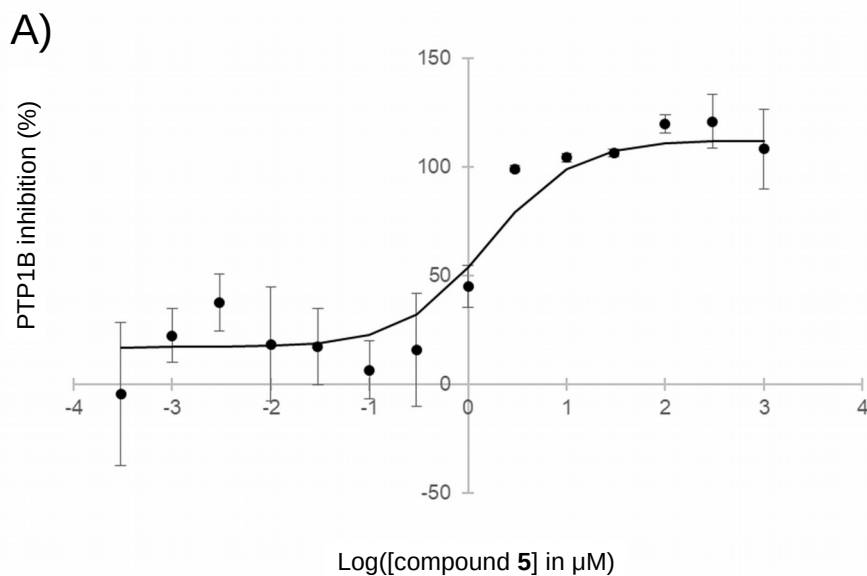
Reaxys RN: 25330566

Tanimoto: 0.31

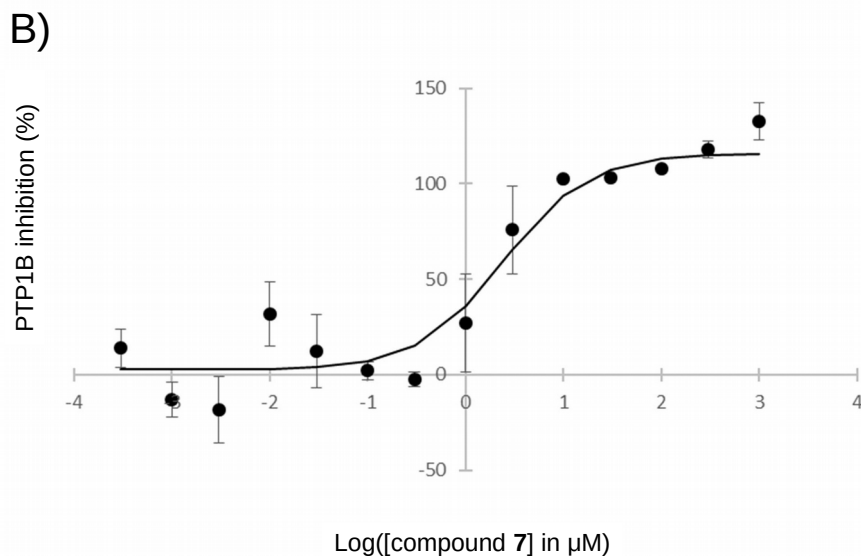
Manuscript 2

Figure S6. In each panel, a hit compound reported by Balaramnavar *et al.*⁴ is represented in 2D, together with the most similar active compound that was used for clustering in the current study. In each case, the hit compound is labeled with its corresponding code in the original publication and the most similar active compound is labeled with its Reaxys¹ Registry Number. The Tanimoto similarity value resulting from the comparison of the OpenEyePath FP of both compounds is shown below. MarvinSketch² was used to draw the structures.

Combined ligand- and receptor-based VS to identify PTP1B inhibitors



[compound 5] in μM	Log([compound 5] in μM)	PTP1B inhibition (%)			Mean \pm standard deviation
		Replicate 1	Replicate 2	Replicate 3	
1000	3	118.7	118.9	87.0	108.2 \pm 18.4
300	2.5	125.2	107.0	130.7	121.0 \pm 12.4
100	2	122.1	115.0	122.7	120.0 \pm 4.3
30	1.5	107.9	106.5	104.6	106.3 \pm 1.6
10	1	103.0	103.2	106.4	104.2 \pm 1.9
3	0.5	100.7	98.6	97.8	99.0 \pm 1.5
1	0	36.1	55.2	43.3	44.9 \pm 9.6
0.3	-0.5	6.9	-4.4	45.2	15.9 \pm 26.0
0.1	-1	8.4	18.9	-7.6	6.6 \pm 13.3
0.03	-1.5	28.6	26.2	-3.1	17.3 \pm 17.6
0.01	-2	-6.7	45.6	16.4	18.4 \pm 26.2
0.003	-2.5	34.8	26.2	51.8	37.6 \pm 13.0
0.001	-3	14.0	36.5	16.4	22.3 \pm 12.4
0.0003	-3.5	-27.6	18.9	n/a	-4.3 \pm 32.9

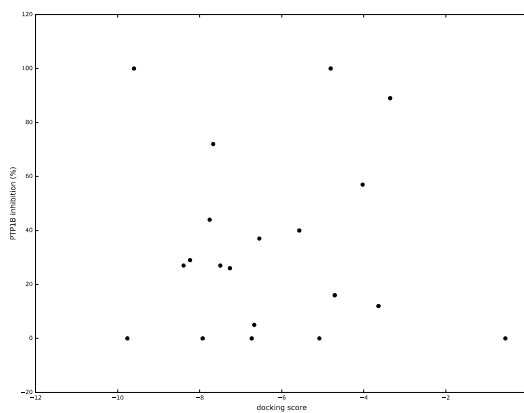


[compound 7] in μM	Log([compound 7] in μM)	PTP1B inhibition (%)			Mean \pm standard deviation
		Replicate 1	Replicate 2	Replicate 3	
1000	3	137.3	121.8	139.2	132.7 \pm 9.5
300	2.5	113.9	117.3	122.3	117.8 \pm 4.2
100	2	108.9	108.6	107.0	108.2 \pm 1.0
30	1.5	103.5	103.2	103.4	103.4 \pm 0.1
10	1	102.5	102.3	102.5	102.4 \pm 0.1
3	0.5	49.6	92.0	86.1	75.9 \pm 23.0
1	0	18.3	7.2	55.7	27.1 \pm 25.4
0.3	-0.5	1.3	-5.9	-2.4	-2.3 \pm 3.6
0.1	-1	7.4	-1.0	-0.0	2.1 \pm 4.6
0.03	-1.5	-9.0	28.0	18.4	12.5 \pm 19.2
0.01	-2	32.4	14.4	47.9	31.6 \pm 16.8
0.003	-2.5	1.8	-30.3	-25.8	-18.1 \pm 17.4
0.001	-3	-17.8	-19.3	-2.7	-13.2 \pm 9.2
0.0003	-3.5	6.7	21.1	n/a	13.9 \pm 10.1

Figure S7. Panels A and B show the IC_{50} curves of compounds **5** and **7**, respectively, and the data used to represent them. In both cases, the percentages of inhibition of PTP1B for 14 different compound concentrations were determined using triplicates. IC_{50} values were calculated using a four parameter logistic regression.

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

A)



B)

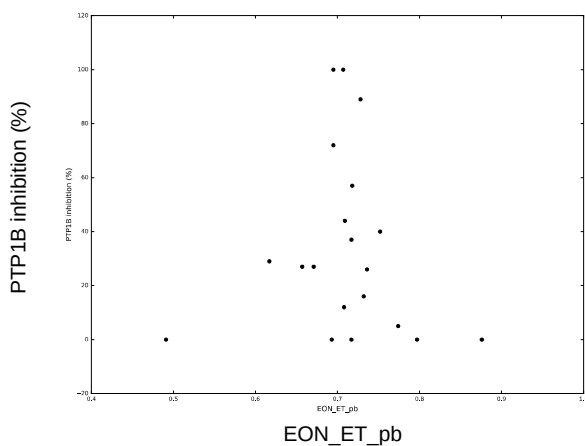
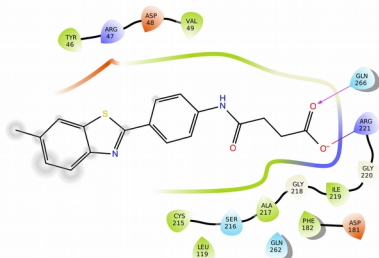


Figure S8. Representations of the docking score (panel A) and EON_ET_pb (panel B) against the percentages of PTP1B inhibition obtained for the 20 hit compounds at the concentration of 100 μ M.

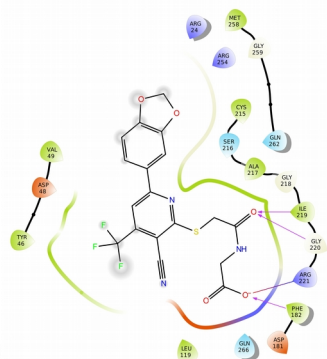
Manuscript 2

A)



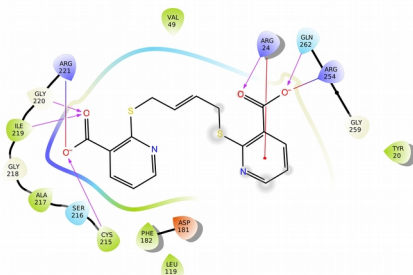
1 (AG-205/07930063)

B)



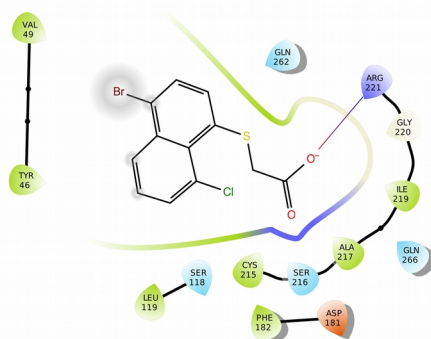
2 (AM-807/41462317)

C)



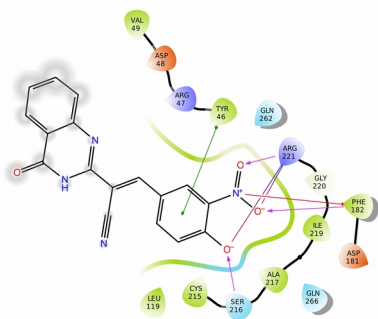
3 (AI-204/43362778)

D)



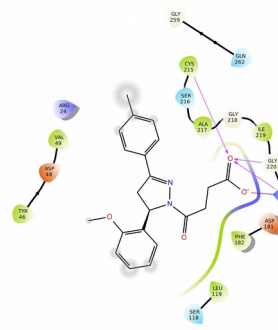
4 (AG-690/11836029)

E)



5 (AF-399/15030257)

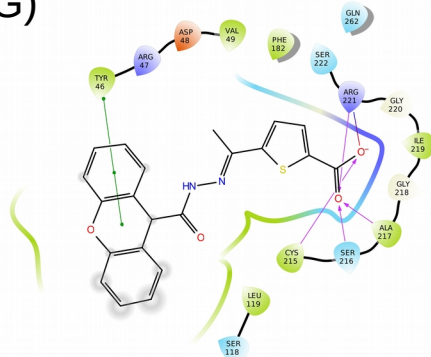
F)



6 (AF-399/15285024)

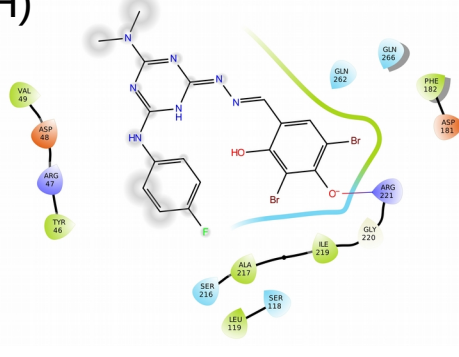
Combined ligand- and receptor-based VS to identify PTP1B inhibitors

G)



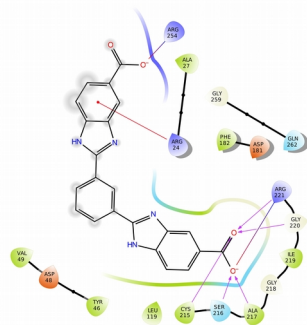
7 (AK-968/41025519)

H)



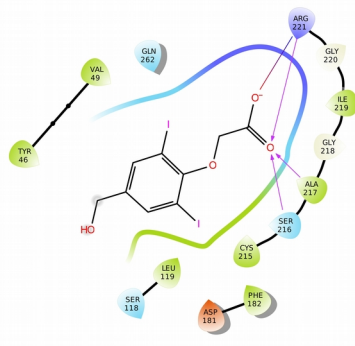
8 (AG-690/40107322)

I)



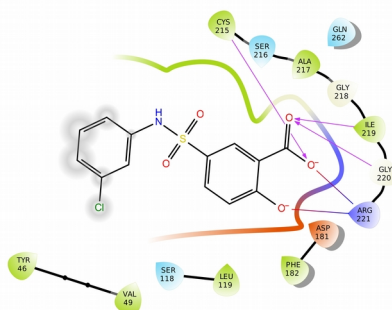
9 (AJ-077/33270018)

J)



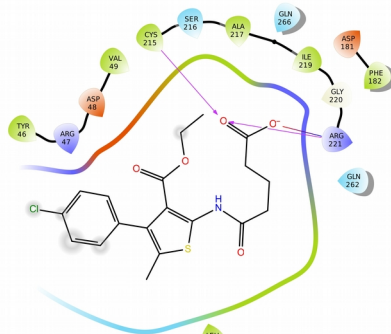
10 (AH-487/41936905)

K)



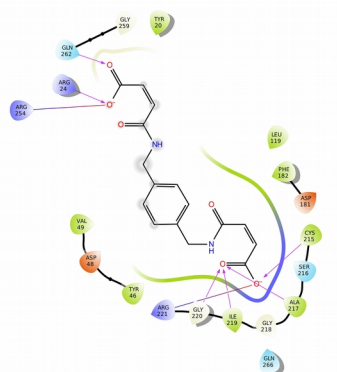
11 (AE-641/42119771)

L)



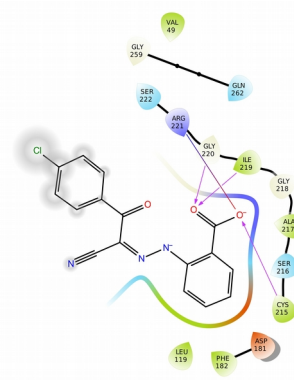
12 (AK-968/15361893)

M)



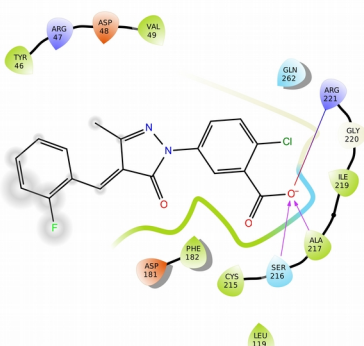
13 (AG-205/37175018)

N)



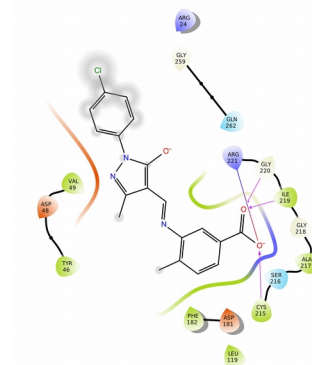
14 (AF-886/30566062)

O)



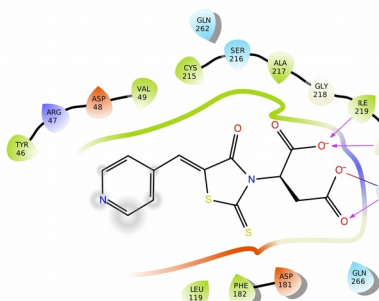
15 (AG-205/37169028)

P)



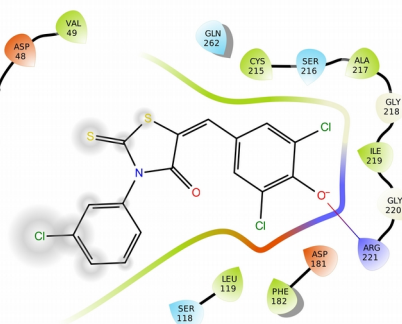
16 (AK-968/11197127)

Q)



17 (AF-962/02757026)

R)



18 (AM-879/12634007)

Combined ligand- and receptor-based VS to identify PTP1B inhibitors

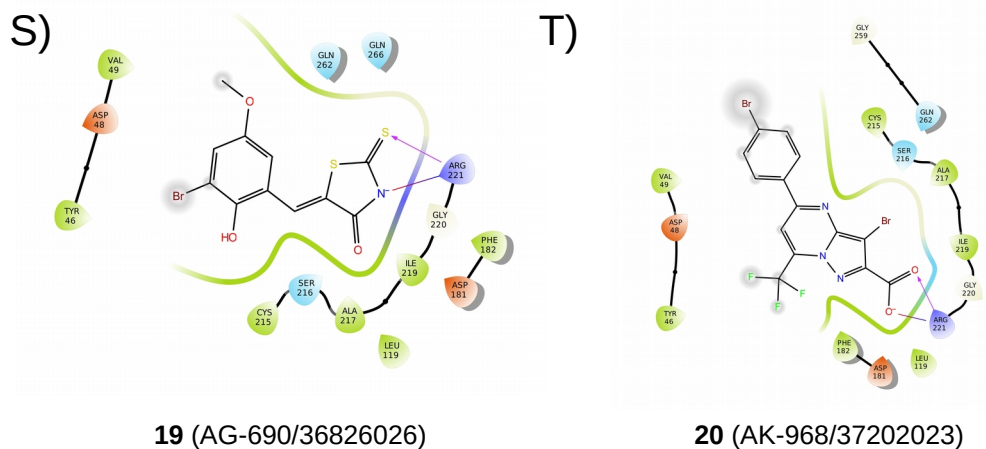


Figure S9. Ligand interaction diagrams of the docked poses with the highest EON_ET_pb values of the 20 hit compounds. Negatively charged residues are colored in red, positively charged residues are colored in blue, hydrophobic residues are colored in green, polar residues are colored in cyan, glycine residues are colored in salmon, hydrogen bonds are represented as fuchsia arrows, salt bridges are represented as red and blue lines, π - π stacking interactions are represented as green lines and π -cation interactions are represented as red lines and atoms exposed to the solvent are circled in gray. The protein surface corresponding to the residues closest to the ligand is represented as a solid line colored by residue according to the previously described color scheme. This representation has been obtained with the help of Maestro⁵ v10.7.

Manuscript 2

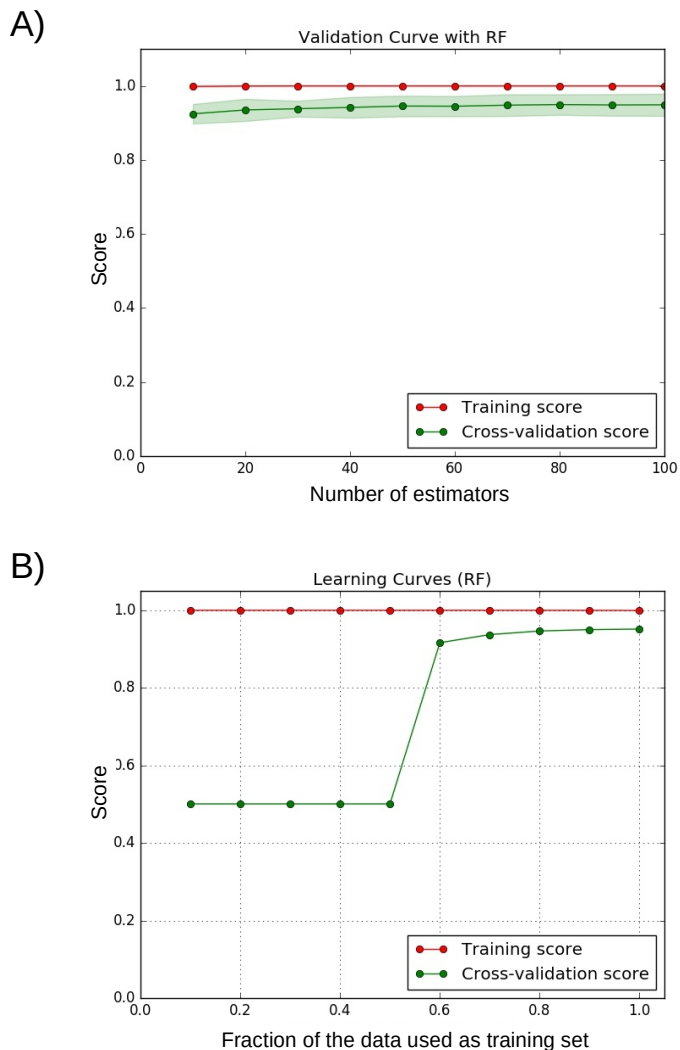


Figure S10. Representations used to define the RF model parameters and evaluate its performance. The training set is plotted in red and the validation set is plotted in green. In panel A, the validation curve is plotted varying the number of trees, showing that the estimator is not overfitted as the training and test scores are similar and that the number of trees chosen is adequate. In panel B, the learning curve is plotted using training and test sets of different sizes. It shows that the training set needs to comprise a high percentage of the data to reduce the error of variance.

References

1. Reaxys. Available at: <https://www.reaxys.com/>.
2. Marvin 16.10.10.0, 2016, ChemAxon. Available at: <http://www.chemaxon.com>.
3. Ma, Y. *et al.* The discovery of a novel and selective inhibitor of PTP1B over TCPTP: 3D QSAR pharmacophore modeling, virtual screening, synthesis, and biological evaluation. *Chem. Biol. Drug Des.* **83**, 697–709 (2014).
4. Balaramnavar, V. M. *et al.* Identification of novel PTP1B inhibitors by pharmacophore based virtual screening, scaffold hopping and docking. *Eur. J. Med. Chem.* **87**, 578–94 (2014).
5. Schrödinger Release 2016-3: Maestro, Schrödinger, LLC, New York, NY, 2016.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Manuscript 3

Understanding the variability of the S1' pocket to improve matrix metalloproteinase inhibitor selectivity profiles

Aleix Gimeno^[a], Raúl Beltrán-Debón^[a], Miquel Mulero^[a], Gerard Pujadas^{[a],[b],*}, Santiago
Garcia-Vallvé^{[a],[b]}

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. e-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Abstract

Matrix metalloproteinases (MMPs) are a family of proteins involved in a wide range of pathologies. Because MMP broad-spectrum inhibition is associated with severe side effects, selectivity has become a priority in the design of MMP inhibitors and it is often achieved by targeting the variable S1' pocket. However, the specific characteristics of the S1' pocket that determine inhibitor selectivity are often not described and, in many cases, challenging to identify. In this review we have inspected the variability of the S1' pocket across the MMP family and we propose explanations for the selectivity of previously described inhibitors. These analyses provide valuable insights into how to design novel inhibitors specific for a given MMP.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

1. Introduction

Matrix metalloproteinases (MMPs) are a family of proteases that degrade various components of the extracellular matrix (ECM) ¹. MMPs are involved in the development of a wide range of diseases, and thus, MMP inhibitors are of interest in different areas of clinical therapy (see Table S1) ²⁻⁶. Early attempts to design MMP inhibitors failed in clinical trials due to the development of musculoskeletal syndrome (MSS) ^{2,4}. Although none of the explanations for the occurrence of MSS proposed over the years have been confirmed, MSS is believed to be a result of broad-spectrum MMP inhibition ^{7,8}. Therefore, MMP inhibitor research now focuses on the identification of selective MMP inhibitors that can prevent these side effects. Moving towards this goal, other regions of the binding site further away from the zinc-binding region have been targeted and, as a result, new insight has been gained into the differences between MMP binding sites ⁹. Concretely, the S1' pocket has been identified as one of the regions with the most variability among the MMPs ^{3,10}. This constitutes an opportunity for the design of selective MMP inhibitors.

The aim of this review is to analyze the differences in the S1' pocket among the different members of the MMP family and shed light on unreported mechanisms of MMP inhibitor selectivity so that they can be rationally exploited in the future to develop novel selective inhibitors for specific MMPs. In this review we will: **(a)** classify the S1' pockets of the MMP experimental structures deposited at the PDB ^{11,12}; **(b)** perform protein-ligand docking simulations to determine whether the differences in selectivity can be attributed to steric hindrances resulting from differences in size and shape of the S1' pocket; and **(c)** compare the electrostatic properties and the hydrophobicity of different S1' pockets in order to explain the selectivity of several inhibitors for a specific MMP. An examination of the recent literature revealed that the most inhibitors have been developed for MMP-2, MMP-9, MMP-12 and MMP-13 (see Tables S2 to S10) and they constitute the main focus of this review. The inhibitors analyzed in this review correspond to those that have been co-crystallized with a given MMP, whose electron

Manuscript 3

density maps (EDMs) are available in the PDB ^{12,13} and whose selectivity against other MMPs has been measured.

2. Moving from the zinc-binding site towards the S1' pocket

Catalytic sites of MMPs are characterized by the presence of a zinc ion required for coordination and catalysis (see Figure S1A). The first approach to the design of MMP inhibitors was the development of peptidomimetic inhibitors (*e.g.*, batimastat and marimastat; see Figure S2A) designed to block the cleavage of collagen by chelating the zinc ion using a hydroxamic acid moiety as the zinc binding group (ZBG) ²⁻⁴.

Next, small molecule inhibitors with different ZBGs were developed (*e.g.*, hydroxamates, carboxylates, thiols and phosphorous-based ZBGs) ². However, the administration of these inhibitors resulted in the development of MSS and failure in clinical trials. Although the reason for this side-effect remains unknown, it has been hypothesized that the low selectivity profiles of these inhibitors may be the cause of this side-effect ^{3,14}. This broad-spectrum inhibition of MMPs is believed to be a consequence of the high structurally conserved zinc-binding region among the MMP family ⁴ (see Figure S1B).

In light of these facts, other types of inhibitors have been developed for the purpose of occupying the S1' pocket, a region in the binding site adjacent to the zinc-binding site which is less conserved among the different enzymes of the MMP family (see Figure S1B). The S1' pocket is accessed through a tunnel created by the wall-forming segment (*i.e.*, residues Pro242, Ile243 and Tyr244; MMP-13 numbering) and it is delimited by a loop of variable length and flexibility called the Ω -loop, which is made up of different residues in different MMPs (see Table S11). Another feature of the S1' pocket is the residue in position 218 (MMP-13 numbering), which is different in some MMPs (Arg in MMP-1, Tyr in MMP-7 and Leu in MMPs -2, -3, -8, -9, -10, -12, -13 and -14). This variability has led to the identification of inhibitors that take advantage of the

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

differences in size and shape of the S1' pocket among the members of the MMP family to achieve selectivity.

Furthermore, several inhibitors that bind to the S1' pocket but that lack a ZBG have been reported for MMP-13¹⁵⁻¹⁸, MMP-8¹⁹ and MMP-12²⁰. Although the presence of a ZBG plays a role in determining the potency of MMP inhibitors^{14,21}, as MMP-13 and MMP-8 have deeper S1' pockets due to the length and flexibility of their Ω -loops, inhibitors that do not possess a ZBG can be designed to occupy the side pocket of S1', which is referred to as the S1'' pocket¹⁸ (see Figure S1A), and therefore achieve inhibitor selectivity against the rest of MMPs¹⁵⁻¹⁹.

Overall, since the S1' pocket is a clear focus of variability among the members of the MMP family, it is the current general belief that it has great potential for the design of selective inhibitors and that future generations of MMP inhibitors should target this pocket given the success of recent inhibitors in achieving selectivity^{5,9,10,14}.

3. S1' pocket classification and docking

The S1' pockets of the different members of the MMP family are generally classified in three categories based on their depth^{3,10}: **(a)** shallow (MMP-1 and MMP-7); **(b)** intermediate (MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14); and **(c)** deep (MMP-3, MMP-10 and MMP-13). Examining the experimental structures available in the PDB^{11,12} reveals that, in many cases, for each MMP, the Ω -loop adopts several conformations in many cases (see Figure 1). For a better understanding of the variability of the S1' pocket, we have classified the different MMPs according to the conformations of their Ω -loop. For this purpose: **a)** all the X-ray structures for which the EDM was available were downloaded from the PDB^{11,12}; **b)** the fitting of the coordinates on the EDM for their binding sites and co-crystallized ligands was validated with VHELIBS²² and Jmol²³ for holo and apo structures, respectively; and **c)** the binding sites that passed this validation were superposed and compared, grouping experimental structures with the same Ω -loop conformation and choosing a

Manuscript 3

representative experimental structure in each case (see Tables S12 to S21). In this comparison, the residue side chains facing to the outside of the S1' pocket were omitted, as they do not have a direct influence on the size and shape of the S1' pocket. In the following paragraphs we discuss, for each MMP, the different Ω -loop conformations that resulted from this classification.

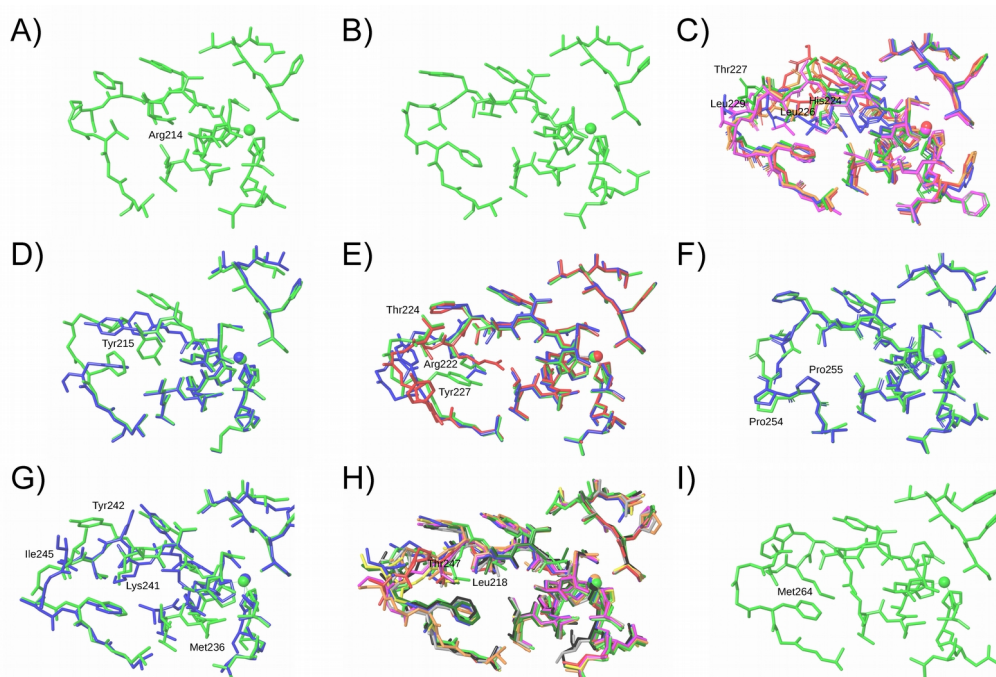


Figure 1. Superposition of the different Ω -loop conformations observed in each MMP. Panels A, B, C, D, E, F, G, H and I show the superposition of representative structures for MMP-1, -2, -3, -7, -8, -9, -12, -13 and -14, respectively. These structures correspond to the representative subunits from Tables S12 to S21. All panels have the same orientation to facilitate comparison. This figure was obtained with Maestro²⁴ v11.

3.1. MMP-1

Of the 11 X-ray experimental structures deposited in the PDB^{11,12} for MMP-1, three demonstrate a good fit between the EDM and the atomic coordinates of the residues of the binding site and have been analyzed (see Table S12). The three experimental structures analyzed for MMP-1 correspond to apo structures and all of them present

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

the same Ω -loop conformation (see Figures 1A and S3). However, the specific conformation of the Ω -loop may not be relevant in the design of MMP-1 inhibitors because the residue Arg214 closes the S1' pocket preventing the interaction of the ligands with the residues of the Ω -loop. This residue should constitute a steric hindrance for those MMP inhibitors that extend to the S1' pocket and it would provide an explanation for the lack of activity against MMP-1 observed in MMP inhibitors such as **1** and **2** reported by Johnson *et al.*¹⁵; **1** and **2** reported by Pochetti *et al.*¹⁹; **21k** reported by Nara *et al.*¹⁸; **15** reported by Taylor *et al.*¹⁷; **1**, **6c** and **7a** reported by Holmes *et al.*²⁵; **1** and **2** reported by Heim-Riether *et al.*²⁶; **12** reported by Shieh *et al.*²⁷; **19v** reported by Becker *et al.*²⁸; **24f** reported by Monovich *et al.*²⁹; **3** reported by Devel *et al.*³⁰; **44** reported by Nara *et al.*³¹; **AM-409** and **MS-560** reported by Tochowicz *et al.*³²; **1** and **3** reported by Nuti *et al.*³³; **3** reported by Camodeca *et al.*³⁴; and **2** reported by Nuti *et al.*³⁵ (see Figure S2B). To illustrate this fact, protein-ligand docking was performed using MMP inhibitors of different sizes that do not show activity on MMP-1 and that were obtained from experimental complexes with other MMPs (see Figure S4). In their respective co-crystallized MMPs, some of these inhibitors occupy the S1' pocket interacting with the Ω -loop (see Figure S4A), others partly occupy the S1' pocket without establishing many interactions with the Ω -loop (see Figure S4B), and others just reach the region that is occupied by Arg214 in MMP-1 (see Figure S4C). Despite their differences, in all cases, the docked poses of these compounds in MMP-1 could not simultaneously occupy the zinc binding region and the S1' pocket due to the blockade by Arg214, resulting in a different predicted binding mode, which would explain their low bioactivity for MMP-1. This supports the idea that selectivity over MMP-1 can be achieved simply by occupying the S1' pocket, preferably by seeking an interaction with the Ω -loop deep in the S1' pocket to ensure that MMP-1 will not be able to accommodate the compound.

3.2. MMP-2

Of the seven X-ray experimental structures deposited in the PDB^{11,12} database for MMP-2, only one has available EDMs and has been analyzed (see Table S13). The Ω -loop in MMP-2 is shorter than in MMP-8 or MMP-13 (see Table S11). This has allowed

Manuscript 3

certain inhibitors to achieve selectivity for these MMPs by aiming to occupy the S1' pocket. This is the case of the MMP-8 inhibitor **2** reported by Pochetti *et al.*¹⁹; the MMP-13 inhibitors **1**, **2** reported by Johnson *et al.*¹⁵; **21k** reported by Nara *et al.*¹⁸ and **15** reported by Taylor *et al.*,¹⁷ which show a low activity against MMP-2 (see Figure S2C). To illustrate this, protein-ligand docking with MMP-2 was performed with these compounds, which were obtained from their respective experimental structures. The docked poses were unable to occupy the S1' pocket as in their original experimental structures, resulting in a different predicted binding mode of the inhibitor and explaining their selectivity over MMP-2 (see Figure S5 for the docking of **2**¹⁵ on MMP-2 compared with its experimental complex at MMP-13). Therefore, the design of MMP-13 and MMP-8 inhibitors with bulky groups in the S1' pocket should give them selectivity over MMP-2. In the case of MMP-8 inhibitor **1** reported by Pochetti *et al.*¹⁹, (see Figure S2C) although it does not reach the S1' pocket in MMP-8, its docked poses on MMP-2 reveal a different binding mode than that observed in MMP-8 (see Figure S6). The different binding mode can be attributed to the difference in length of the Ω -loop between MMP-8 and MMP-2, providing a possible explanation for the selectivity observed for this compound.

3.3. MMP-3

Of the 31 X-ray experimental structures deposited in the PDB^{12,13} for MMP-3, 11 have available EDMs and have been analyzed (see Table S14). The abundance of residues that constitute the Ω -loop of MMP-3 (see Table S11) makes it highly flexible, which is manifested in the wide variety of conformations that this loop adopts in different MMP-3 experimental structures (see Figures 1C and S7). The Ω -loop conformations A and B are observed in the absence of ligands (see Table S14). The C conformation of the Ω -loop is very similar to the D conformation with the exception that the ligand induces an alternate conformation of the residue His224, which faces outwards, thus enlarging the pocket (see Figures S7D and S7E). Owing to the sequence length of the Ω -loop in MMP-3, the S1' pocket of MMP-3 is frequently classified as large, in the same category as MMP-13¹⁰. However, despite the similarity between these two proteins regarding the number of residues that constitute their respective Ω -loops (see Table S11), the

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

residue differences between them allow specific residue conformations to be adopted in each MMP that play an important role in the definition of the shape and size of the S1' pocket. While in all the conformations of the MMP-3 Ω -loop, the side-chain of residue Leu229 (and in some conformations also the side chains of residues Leu226 and Thr227) hinders access to the bottom of the S1' pocket, this is not the case for MMP-13, in which no residue side-chains face the S1' pocket with the exception of the small side chain of Thr247 (see Figure 1H). As a result, the S1' pocket of MMP-13 is bigger than that of MMP-3, enabling it to accommodate larger ligands. Many MMP-13 inhibitors including **1** and **2** reported by Johnson *et al.*¹⁵, **21k** reported by Nara *et al.*¹⁸ and **15** reported by Taylor *et al.*¹⁷ have taken advantage of this larger size to achieve selectivity against MMP-3 (see Figures S2D and S8) and we propose following this strategy to obtain MMP-13 inhibitors that are selective over MMP-3. Even the S1' pocket of MMP-8 may prove to be larger than that of MMP-3 in some cases, despite being classified as intermediate¹⁰. This can occur because the Ω -loop of MMP-8 is able to adopt a particular conformation in which the conformation of the residues from Arg222 to Tyr227 is different, and the different orientations of the side chains of residues Thr224 and Tyr227 allow for a larger S1' pocket cavity (see Figure 1E). Triggering this conformational change may be a useful strategy in developing MMP-8 inhibitors that are selective over MMP-3, as is the case of MMP-8 inhibitor **2** reported by Pochetti *et al.*¹⁹ (see Figure S2D).

3.4. MMP-7

Of the five X-ray experimental structures deposited in the PDB^{12,13} for MMP-7, only two have available EDMs and have been analyzed (see Table S15). They both correspond to holo structures, each of them presenting a different S1' pocket (see Figures 1D and S9). In MMP-7, the residue equivalent to Leu218 of MMP-13 is Tyr215. Because of this particular residue, the S1' pocket of MMP-7 is typically classified as shallow, in the same category as MMP-1¹⁰. Similar to Arg214 in MMP-1, in MMP-7, Tyr215 adopts a conformation that faces the zinc-binding region and prevents the binding of ligands with the S1' pocket. Therefore, this residue should constitute a steric hindrance for those ligands which interact with the Ω -loop and it provides an explanation for the

Manuscript 3

selectivity observed in the majority of ligands that perform these type of interactions. This is illustrated by the docking on MMP-7 of several MMP inhibitors that show selectivity over MMP-7 (*i.e.*, **1** reported by Pochetti *et al.* ¹⁹, **AM-409** reported by Tochowicz *et al.* ³², **20** reported by Mannino *et al.* ³⁶, **1** reported by Holmes *et al.* ²⁵, **36** reported by Devel *et al.* ³⁷, **3** reported by Devel *et al.* ³⁰, **2** reported by Johnson *et al.* ¹⁵, **24f** reported by Monovich *et al.* ²⁹, **1** and **2** reported by Heim-Riether *et al.* ²⁶, **19v** reported by Becker *et al.* ²⁸, **12** reported by Shieh *et al.* ²⁷ and **44** reported by Nara *et al.* ³¹; see Figure S2E), whose docked poses present a different binding mode than that observed in their original MMP (Figure S10 shows the docking of inhibitor **19v** ²⁸). However, an experimental structure of MMP-7 with a large ligand has been reported to show an alternate conformation of residue Tyr215, opening the access to the S1' pocket in order to accommodate the ligand ³⁸ (see Figures 1D and S9C). Nevertheless, although the coordinates of the residues from Gly242 to Asn248 were not included in the experimental structure due to a lack of electron density, as the side-chain of residue Tyr215 in this conformation faces the S1' pocket (see Figures 1D and S9C), it could be argued that the size of the S1' pocket of MMP-7 in this particular conformation may still be significantly smaller than that of other MMPs such as MMP-8 or MMP-13. Thus, the targeting of the S1' pocket by MMP inhibitors directed to these MMPs should provide their selectivity over MMP-7. Actually, inhibitors that target the S1' pocket of MMP-8 and MMP-13 have been shown to present high selectivity over MMP-7, supporting the previous idea. This is the case of the MMP-8 inhibitor **2** reported by Pochetti *et al.*, ¹⁹ MMP-13 inhibitors **1** and **2** reported by Johnson *et al.* ¹⁵, MMP-13 inhibitor **21k** reported by Nara *et al.* ¹⁸ and MMP-13 inhibitor **15** reported by Taylor *et al.* ¹⁷ (see Figure S2E).

3.5. MMP-8

Of the 25 X-ray experimental structures deposited in the PDB ^{12,13} for MMP-8, 16 have available EDMs and have been analyzed (see Table S16). In the most common conformation of the Ω -loop for MMP-8 (conformation A), the side chain of residue Tyr227 faces to the inside of the S1' pocket (see Figures 1E and S11B). Although both apo and holo structures present this Ω -loop conformation (see Table S16), the ligands

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

of those structures do not reach the bottom of the S1' pocket. Conformation B is observed in holo structures with large ligands which are able to establish contacts with the Ω -loop. In this conformation, the overall arrangement of the residues of the omega loop is altered to better accommodate the ligand (*i.e.*, the side chains of Arg222 and Tyr227 face slightly outwards, thus granting the ligand access to the S1' pocket, and the side chain of Thr224, which faced the outside of the S1' pocket in conformation A, faces the inside of the S1' pocket, allowing it to interact with the ligand; see Figures 1E and S11C). Yet another conformation has been observed for the Ω -loop of MMP-8 (conformation C) in which Tyr227 has an open conformation (see Figures 1E and S11D). However, only one structure has been determined with this conformation, in which the 2-(N-morpholino)-ethanesulfonic acid moiety of the ligand is in close proximity to Tyr227, which therefore could be the cause of this conformation. Although the S1' pocket of MMP-8 is often classified as medium-sized¹⁰, the Ω -loop of MMP-8 demonstrates a certain degree of flexibility by adopting a different conformation (conformation B) and being able to accommodate large ligands in the S1' pocket. This conformation should be considered when discussing selectivity against MMP-8 since, in terms of size, it is more comparable to the S1' pocket of MMP-13 than to the S1' pocket of other MMPs with the same Ω -loop length, such as MMP-12 or MMP-14 (see Table S11). Nevertheless, as the S1' pocket of MMP-13 is still larger than that of MMP-8, it is possible for MMP-13 inhibitors that bind to the S1' pocket to achieve selectivity by targeting the S1'' pocket. The protein-ligand docking on MMP-8 of several MMP-13 inhibitors that show selectivity over MMP-8 (see Figure S2F) predicts that these inhibitors are not able to interact with MMP-8 in the same binding mode as they do in MMP-13 (see Figure S12). Although some docked poses for the MMP-13 inhibitors **2** and **21k** reported respectively by Johnson *et al.*¹⁵ and Nara *et al.*¹⁸ reach the S1'' pocket of MMP-8, none of the docked poses are able to place the carboxylic group at the S1'' pocket while extending to the zinc-binding region at the same time (see Figure S12 for the docking of **21k** on the three Ω -loop conformations for MMP-8). This could be attributed to the difference in size and shape between the S1' pockets of these MMPs and may offer an explanation for the selectivity observed in these inhibitors.

3.6. MMP-9

Of the 24 X-ray experimental structures deposited in the PDB ^{11,12} for MMP-9, 17 have available EDMs and have been analyzed (see Table S17). The Ω -loop in MMP-9 is special in the sense that it presents two consecutive proline residues (Pro254 and Pro255, see Table S11). Given the exceptional conformational rigidity of proline and the low spatial variability of the Ω -loop residues in the MMP-9 experimental structures obtained, it can be inferred that the presence of these two proline residues significantly reduces the flexibility of the MMP-9 Ω -loop (see Figure 1F). In fact, most of the structures analyzed present the same conformation, regardless of the presence of a ligand (see Table S17 and Figure S13B). Only in one experimental structure did we observe a slightly different arrangement of the Ω -loop (conformation B; see Figures 1F and S13C). This rigidity of the Ω -loop together with the small size of its S1' pocket allows for an easier identification of inhibitors of MMPs with a long and more flexible Ω -loop that are selective against this protein just by introducing a bulky group in the S1' pocket which can be accommodated by these proteins, but not by MMP-9. Protein-ligand docking on MMP-9 of several MMP inhibitors that are selective over MMP-9 (*i.e.*, **1** and **2** reported by Pochetti *et al.* ¹⁹, **1** and **2** reported by Johnson *et al.* ¹⁵, **21k** reported by Nara *et al.* ¹⁸, **44** reported by Nara *et al.* ³¹ and **15** reported by Taylor *et al.* ¹⁷; see Figure S2G) has been conducted in order to illustrate this fact (Figure S14 shows the docking of inhibitor **2** ¹⁹). Some of these inhibitors occupy the larger S1'' pocket in their original experimental structure (sometimes reaching the S1'' pocket, which is not present in MMP-9), but are not able to do so in MMP-9. Because of the difference in size and shape of the S1' pocket between MMP-9 and the original MMP of these inhibitors, their docked poses fail to reproduce the binding mode observed in the original MMP, which offers an explanation for their selectivity.

3.7. MMP-10

For human MMP-10, only three X-ray experimental structures have been determined by X-ray diffraction and are deposited in the PDB ^{11,12}. All of them have EDMs and have been analyzed (see Table S18). MMP-10 has a very flexible Ω -loop, showing variability

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

among different experimental structures and often low electron density for some of its Ω -loop residues (see Figure S15). For example, the coordinates of residue Phe242 have not been obtained for any of the experimental structures analyzed here due to the lack of electron density in that region, probably indicating the flexibility of that region of the loop. However, despite this flexibility, the S1' pocket of MMP-10 may present differences in size, shape or electrostatic properties when compared to the S1' pockets of MMPs which also present a long Ω -loop, which would explain the high selectivity values of the MMP-13 inhibitors **21k** and **15** reported by Nara *et al.*¹⁸ and Taylor *et al.*¹⁷ against MMP-10 (see Figure S2H). Unfortunately, the lack of electron density for some residues makes it difficult to relate this selectivity to size, shape or electrostatic differences between MMP-10 and MMP-13.

3.8. MMP-12

Of the 64 X-ray experimental structures deposited in the PDB^{11,12} for MMP-12, 50 have available EDMs and have been analyzed. All of them show the same conformation of the Ω -loop (see Table S19 and Figure S16) except for one particular structure (*i.e.*, the B conformation) whose main characteristics are the conformations adopted by residues Tyr242 and Met236 (*i.e.*, Tyr242 faces out of the S1' pocket and Met236 blocks the access to the S1' pocket; see Figures 1G and S15C). Despite the reasonably large amount of residues that conform the Ω -loop in MMP-12 and define an S1' pocket classified as medium-sized compared to other MMPs¹⁰, the low variability of the Ω -loop observed among the experimental structures of MMP-12 indicates that this loop has limited flexibility. Apart from residue Lys241, which seems to adopt a wide range of conformations, the other residues of the Ω -loop do not present alternate conformations. Assuming that conformation A is the only conformation of the MMP-12 Ω -loop that allows the ligand to reach the S1' pocket, achieving selectivity over MMP-12 by inhibitors directed to MMPs with a larger and more flexible Ω -loop should be possible by targeting the S1'' pocket, which is blocked by residue Ile245 in MMP-12. To illustrate this, protein-ligand docking was performed using ligands selective over MMP-12 (*i.e.*, **2** reported by Pochetti *et al.*¹⁹, **1** and **2** reported by Johnson *et al.*¹⁵ and **15** reported by Taylor *et al.*¹⁷; see Figure S2I) obtained from their respective experimental

Manuscript 3

structures (Figure S17 shows the docking of inhibitor **15**¹⁷). In all cases, the docked poses of these compounds in MMP-12 could not occupy the S1' pocket due to its smaller size, therefore resulting in a different hypothetical binding mode of the inhibitors than that observed in their experimental structures, which would explain their selectivity over MMP-12.

3.9. MMP-13

Of the 44 X-ray experimental structures deposited in the PDB^{12,13} for MMP-13, 31 have available EDMs and have been analyzed (see Table S20). The Ω -loop of MMP-13 is composed by 13 residues (see Table S11) and its S1' pocket is classified as deep¹⁰. The large Ω -loop of MMP-13 gives a high degree of flexibility, manifested in several conformations adopted by this loop (see Figures 1H and S18). From the structures classified here, four subunits corresponded to the apo form and 42 subunits corresponded to holo forms. We found that all the apo forms had the same Ω -loop conformation (conformation A) whereas the holo forms show different conformations depending on the type of ligand that binds to MMP-13. In structures where the ligand reaches only the S1' pocket, we saw conformations from A to F. However, some ligands occupy the S1'' pocket. In most of these cases, the conformation of the loop is different (conformations from G to I), meaning that the high flexibility of the Ω -loop allows it to adapt in order to allow ligands to bind to the S1'' pocket.

3.10. MMP-14

Of the nine X-ray experimental structures deposited in the PDB^{11,12} for MMP-14, only one has an available EDM and has been analyzed (see Table S21). The S1' pocket of MMP-14 is classified as intermediate and its Ω -loop has the same length as those of MMP-13 and MMP-8 (see Table S11). However, recent molecular dynamics simulation analyses have showed that the Ω -loop in MMP-14 is less flexible than that of MMP-13 or MMP-8³⁹. A closer look at the side chains of the Ω -loop in MMP-14 that are oriented towards the S1' pocket reveals that the side chain of residue Met264 occupies part of the pocket (see Figures 1I and S19). The low flexibility of the loop might not allow the

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

adoption of a different conformation for this residue side-chain; therefore, seeking a steric clash with this particular residue may provide an anchor for identifying inhibitors selective against MMP-14. When performing docking simulations on MMP-14 with co-crystallized inhibitors selective towards this MMP (*i.e.*, **1** and **2** reported by Pochetti *et al.* ¹⁹, **36** reported by Devel *et al.* ³⁷, **3**, **16** and **18** reported by Devel *et al.* ³⁰, **1** and **2** reported by Johnson *et al.* ¹⁵, **24** reported by Savi *et al.* ⁴⁰, **2** reported by Heim-Riether *et al.* ²⁶, **21k** reported by Nara *et al.* ¹⁸, **44** reported by Nara *et al.* ³¹ and **1** and **15** reported by Taylor *et al.* ¹⁷; see Figure S2J), in some cases residue Met264 represented a steric hindrance for these inhibitors that does not allow them to bind to the S1' pocket as they would in their original MMPs, or at least not in the same fashion (Figure S20 shows the docking of inhibitor **3** ³⁰). This would provide an explanation for the selectivity observed in these inhibitors. Based on these observations, inhibitors that reach the S1' pocket can generally be expected to have selectivity over MMP-14.

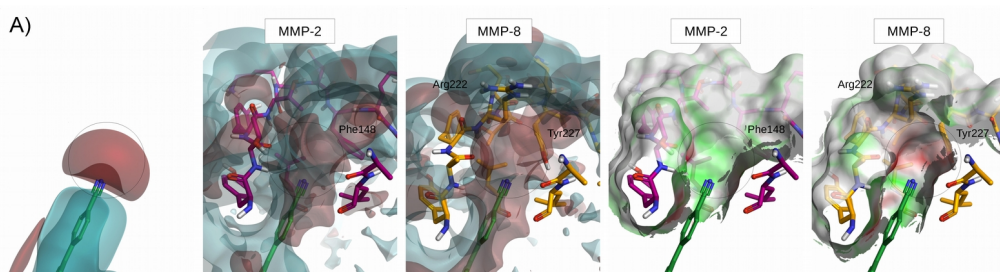
4. Different sequence, different pocket

While the previous section illustrated how the size and shape of the S1' pocket are crucial determining factors in the selectivity of many MMP inhibitors, the residue differences in different Ω -loop sequences may also translate into different hydrogen bond interactions and Van der Waals contacts with the ligand among the different MMPs, as well as different electrostatic and hydrophobic properties of the S1' pocket. These type of differences also constitute an opportunity to achieve inhibitor selectivity. Here we will examine the characteristics of several MMP-2, MMP-9, MMP-12 and MMP-13 inhibitors and propose the possible causes of their selectivity against other MMPs in order to gain insight into the particular characteristics of the S1' pocket of these MMPs.

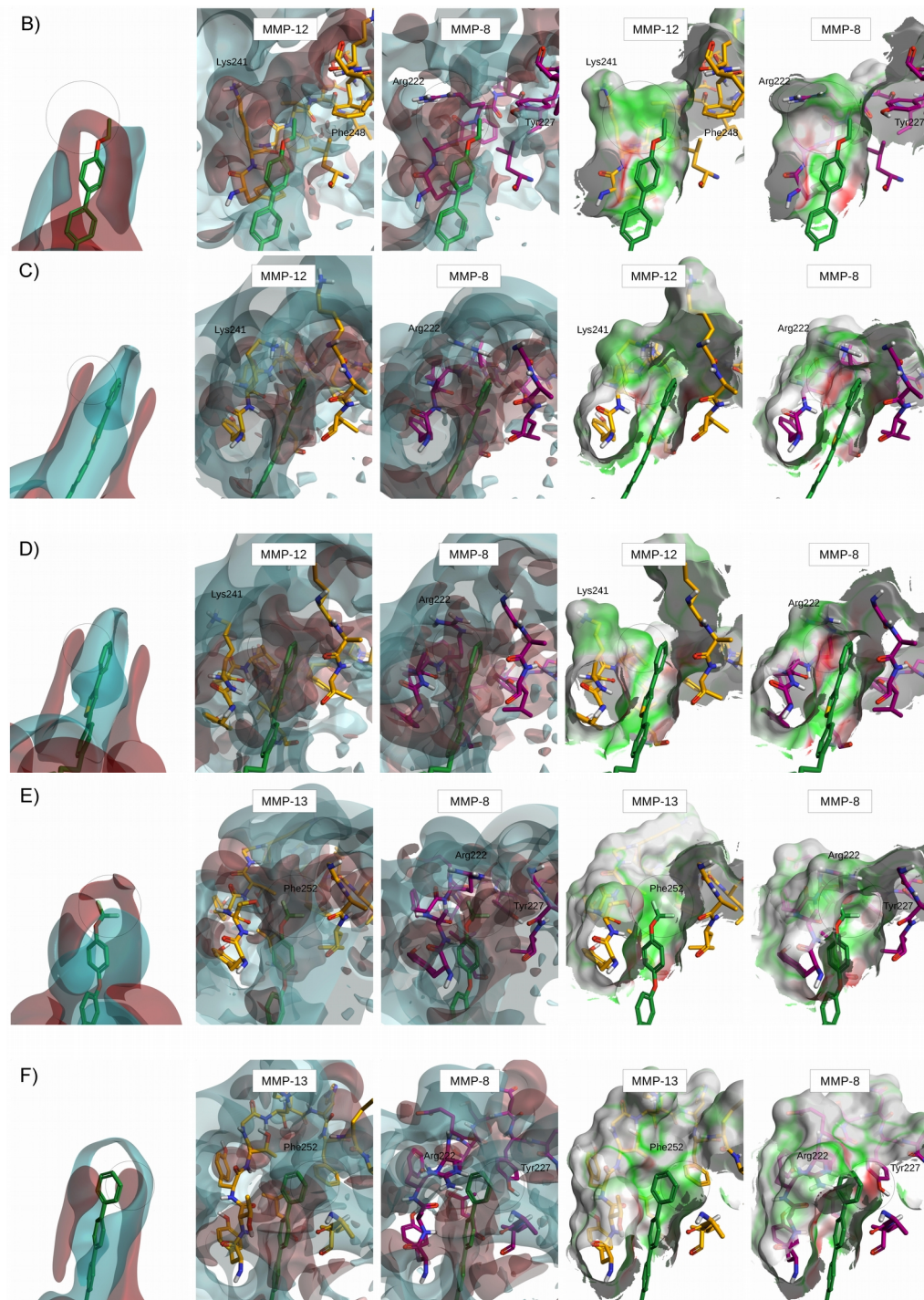
4.1. Selectivity towards MMP-2

A critical aspect in the design of MMP-2 inhibitors for cancer therapy is avoiding the inhibition of anti-targets such as MMP-3, MMP-8, MMP-9 and MMP-14, whose inhibition could be detrimental to cancer prognoses^{41–43}. In the previous section we have already proposed how to achieve selectivity against MMP-14 and here we will focus on how MMP-2 inhibitors have been able to achieve selectivity against MMP-8 and MMP-9. As MMP-2 inhibitors that spare MMP-3 have not been co-crystallized with either protein, we have not been able to analyze this particular case.

Regarding selectivity against MMP-8, the MMP-8 inhibitor **2b** co-crystallized with MMP-8 by Campestre *et al.*⁴⁴ was found to be more active for MMP-2 than for MMP-8 (see Figure S2K). Although the authors attribute the increase in MMP-2 affinity to the possible release of the torsional strain of the two rings of the biphenyl group in the wider S1' pocket of MMP-2⁴⁴, this increased activity could also be related to the different electrostatic nature of the two S1' pockets. In this regard, Figure 2A shows that the negative nitro group of the ligand has a better electrostatic complementarity with the S1' pocket of MMP-2 than with the S1' pocket of MMP-8. Therefore, the inclusion of negative groups in MMP-2 inhibitors targeting this region of the S1' pocket should provide them selectivity against MMP-8.



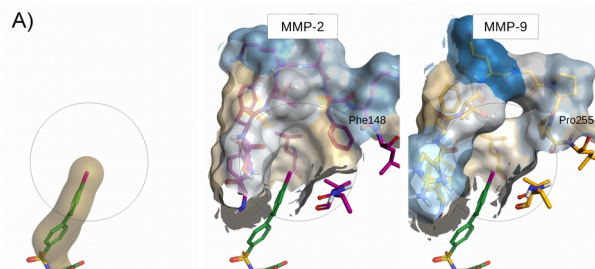
Understanding the variability of the S1' pocket to improve MMPi selectivity profiles



Manuscript 3

Figure 2. Inhibitors selective against MMP-8. The five screen captions that form each panel have the same orientation to facilitate comparison and, from left to right, show: (i) the electrostatic potential around the ligand; (ii) the electrostatic potential around the binding site of the MMP that is compared with MMP-8; (iii) the electrostatic potential around the binding site of MMP-8; (iv) the electrostatic complementarity between the ligand and the MMP that is compared with MMP-8; and (v) the electrostatic complementarity between the ligand and MMP-8. In the first three screen captions of each panel, the positive and negative electrostatic potentials are shown in blue and red, respectively. In the last two screen captions of each panel, the molecular surface of the protein is colored by the electrostatic complementarity between the ligand and the protein: green and red represent the areas of good and bad complementarity, respectively. The protein backbone of the MMP that has not been crystallized with the inhibitor (in purple) has been superposed to the experimental complex (in orange) to determine its relative position to the co-crystallized ligand (in green). The co-crystallized inhibitors shown in panels from A to F are **2b**,⁴⁴ **PD-0359601**,²¹ **16**,³⁰ **19v**²⁸ and **18**³⁰ (which were co-crystallized with MMP-8, MMP-12, MMP-12, MMP-12, MMP-13 and MMP-13, respectively). A circle indicates the region (either from the corresponding MMP or from the ligand) involved in differences in the bioactivity of the same ligand relative to the two MMP that are compared in each panel. This figure was obtained with Flare⁴⁵ v2.0.

MMP-9 is another clear anti-target usually considered in the development of MMP-2 inhibitors due to its reported anti-angiogenic and anti-tumorigenic properties⁴². Interestingly, Tochowicz *et al.*³² reported two inhibitors co-crystallized in MMP-9, **An-1** and **MS-560**, which displayed higher activity for MMP-2 than for MMP-9 (see Figure S2K). The authors hypothesize that the mobility of the side-chain of residue Arg249 in MMP-9 is responsible for this difference in activity³². This residue is replaced with the less bulky residue Thr224 in MMP-2, which would result in much weaker hindering by its side-chain³². Nevertheless, given the rigidity of the MMP-9 Ω -loop, whose conformation is conserved among almost all the reported crystal structures (see Table S17 and Figure S13), another possible explanation could be that residue Phe148 in MMP-2 may be able to get closer to the ligand and establish better hydrophobic interactions than residue Pro255 in MMP-9 (see Figure 3).



Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

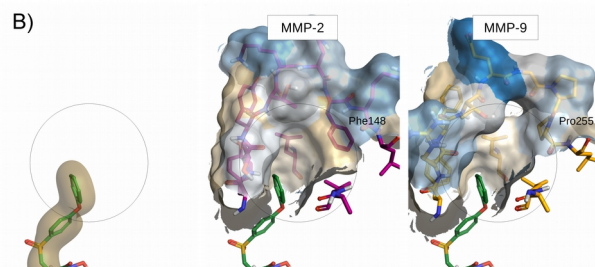


Figure 3. Inhibitors selective against MMP-9. The three screen captions that form each panel have the same orientation to facilitate comparison and, from left to right, they show: (a) the hydrophobic area around the ligand; (b) the hydrophobicity of the MMP-2 binding site; (c) the hydrophobicity of the MMP-9 binding site. In the first screen caption of each panel, the hydrophobic surface of the ligand is shown in beige. In the next two screen captions of each panel, the molecular surface of the protein is colored by its polarity: while the hydrophilic areas are shown in blue, the hydrophobic areas are in beige. The inhibitors shown in panels A and B are **An-1**³² and **MS-560**,³² respectively, which were both co-crystallized with MMP-9. A circle indicates the region (either from the corresponding MMP or from the ligand) involved in differences in the bioactivity of the same ligand relative to the two MMP that are compared in each panel.

4.2. Selectivity towards MMP-9

Interestingly, while searching for MMP-12 inhibitors, Morales *et al.*²⁰ obtained the high-throughput screening hit **CP-271485**, which has higher activity for MMP-9 than for MMP-12 (see Figure S2L). While we could not relate this difference in activity to electrostatics or hydrophobics, given the smaller size of this inhibitor compared to the other ones obtained and the narrower shape of the S1' pocket of MMP-9 compared to that of MMP-12, the authors hypothesized that the higher activity of this inhibitor for MMP-9 could be explained by its better accommodation in the MMP-9 S1' pocket²⁰.

4.3. Selectivity towards MMP-12

The S1' pocket of MMP-12 is mostly characterized by its high hydrophobicity²⁰. The combination of a series of hydrophobic residues (*i.e.*, Ala234, Val235, Phe237, Lys241, Val243 and Phe248), provide the S1' pocket with a more hydrophobic environment compared to other MMPs²⁰. The selectivity of many MMP-12 inhibitors has been attributed to this hydrophobic nature. Morales *et al.*²⁰ identified the MMP-12 inhibitors **PD-0359601** and **PF-00356231**, which are highly selective against MMP-2 (see Figure S2M). The authors claim that these ligands are less stabilized by hydrophobic

Manuscript 3

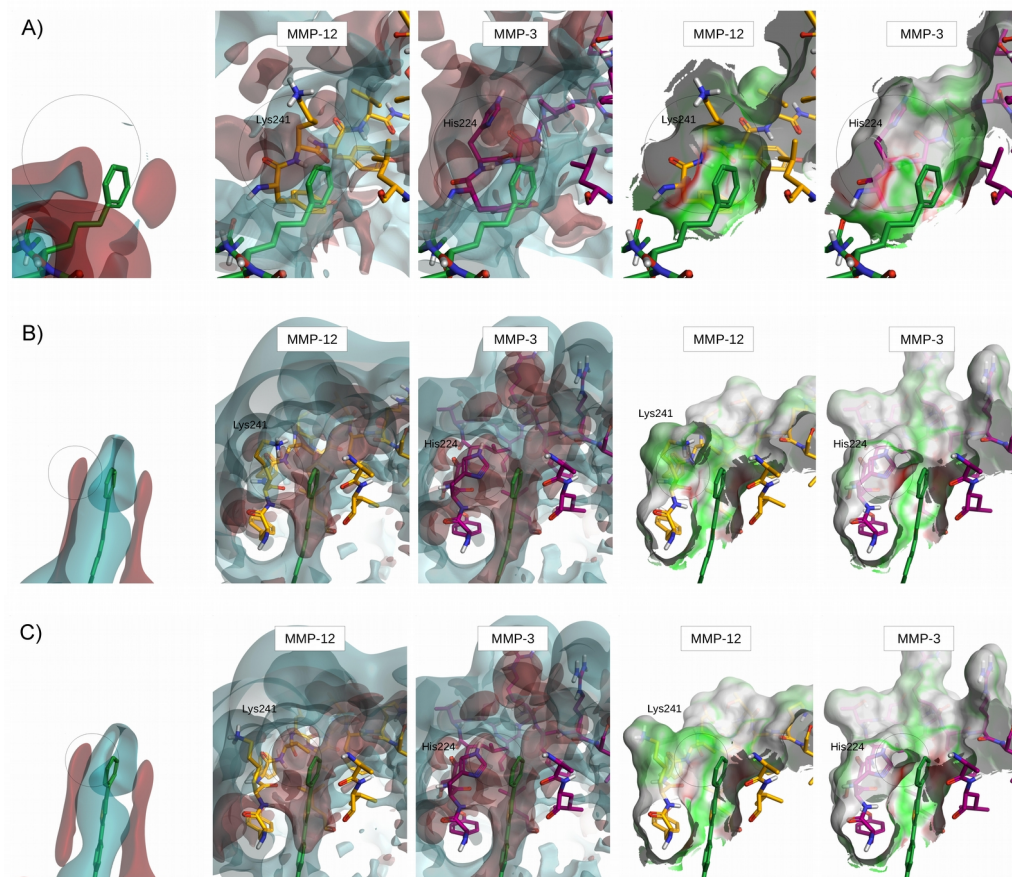
interactions in the more open MMP-2 S1' pocket than in that of MMP-12²⁰. These inhibitors were also selective against MMP-3, which the authors attribute to the fact that MMP-3 has a larger S1' pocket and the ligands are not able to occupy its volume to the same degree as in MMP-12²⁰. These hypotheses are in agreement with our docking simulations showing that the lipophilic interactions of these two ligands with the protein are stronger for MMP-12 than for MMP-2, and MMP-3 (see Table S22). Likewise, our docking simulations of compounds **3** and **16** reported by Devel *et al.*³⁰ show that lipophilic interactions of these ligands with the protein are significantly stronger for MMP-12 than for MMP-2, -3, -8, -9 and 13, MMPs against which these inhibitors are selective (see Table S22 and Figure S2M). A similar situation may occur in the case of compounds **1**, **6c** and **7a** reported by Holmes *et al.*²⁵ and compound **36** reported by Devel *et al.*³⁷ (see Table S22 and Figure S2M), since these inhibitors introduce hydrophobic moieties in the S1' pocket and are also selective against other MMPs.

Overall, the predominance of hydrophobic residues in the S1' pocket of MMP-12, makes its hydrophobic environment an attractive characteristic to target in the design of selective inhibitors. Furthermore, electrostatics can also play a part in determining selectivity for MMP-12. For instance, in the case of **PD-0359601**, the ethoxy group of the ligand (see Figure S2M) places a negative electrostatic surface in the S1' pocket of MMP-12 that results in a better electrostatic complementarity with MMP-12 relative to MMP-8 (see Figure 2B). This higher electrostatic complementarity would explain why this ligand displays a higher bioactivity for MMP-12 than for MMP-8. A similar situation occurs in the case of the MMP-12 inhibitors **16** and **3** reported by Devel *et al.*³⁰ (see Figures 2C, 2D and S2M), as the π system of the phenyl ring of these ligands projects a negative electrostatic surface to both sides of the ring resulting in a better electrostatic complementarity with MMP-12 than with MMP-8.

Differences in the electrostatics of the S1' pocket can also be exploited to achieve selectivity against MMP-3. In the case of the MMP-12 inhibitor **1** reported by Holmes *et al.*²⁵ (see Figure S2M), the negative electrostatic surface of the π system of the ligand's phenyl ring makes the ligand more suitable to interact with the S1' pocket of

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

MMP-12 than with the S1' pocket of MMP-3 (see Figure 4A). This is in agreement with the lower affinity of this ligand for MMP-3 than for MMP-12 (see Figure S2M). Similarly, the MMP-12 inhibitors **16** and **3** reported by Devel *et al.*³⁰, which are also selective against MMP-3 (see Figure S2M) present a negative electrostatic surface projecting from their thiophene and phenyl rings towards the S1' pocket of this protein (see Figures 4B and 4C). In all these cases, the negative electrostatic surface of the ligand shows better electrostatic complementarity with MMP-12 than with MMP-3 and this could explain the selectivity of these ligands against MMP-3 (see Figure S2M).



Manuscript 3

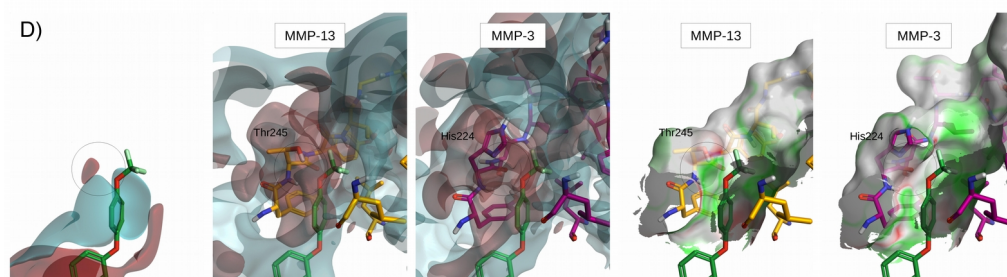


Figure 4. Inhibitors selective against MMP-3. This figure has been prepared using the same guidelines as Figure 2. The co-crystallized inhibitors shown in panels from A to D are **1**,²⁵ **16**,³⁰ **3**³⁰ and **19v**²⁸ (which were co-crystallized with MMP-12, MMP-12, MMP-12 and MMP-13, respectively).

4.4. Selectivity towards MMP-13

Although as discussed in the previous section, the more secure approach to selectively inhibiting MMP-13 is considered to be the targeting of the S1'' pocket, some MMP-13 inhibitors have shown selectivity against other MMPs without reaching this subpocket. This is for example the case of the MMP-13 inhibitor **19v** reported by Becker *et al.*²⁸, which presents selectivity against both MMP-3 and MMP-8 (see Figure S2N). The analysis of the potential electrostatic interactions between the ligand and the two proteins reveals that if the binding mode observed by the inhibitor in MMP-13 were conserved in the binding sites of MMP-3 and MMP-8, the electrostatic complementarity of this compound and the corresponding binding site would be worse for MMP-3 and MMP-8 relative to MMP-13 (see Figures 4D and 2E, respectively). Therefore, both of these situations would result in a decrease in bioactivity for either MMP relative to MMP-13. These observations correlate with the observed selectivity of this MMP-13 inhibitor against both of these MMPs (see Figure S2N). The MMP-13 inhibitor **18** reported by Devel *et al.*³⁰ (see Figure S2N) also places a negative electrostatic surface in the S1' pocket of MMP-8 through the π system of its last phenyl ring, therefore decreasing the affinity of the ligand for this MMP (see Figure 2F). Interestingly though, the MMP-8 inhibitor **17** also reported by Devel *et al.*³⁰ can in its turn take advantage its higher electrostatic complementarity with the S1' pocket of MMP-8 relative to the S1' pockets of MMP-13, MMP-2 or MMP-3 to achieve selectivity against these MMPs (see

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Figure S2O) by placing a methyl group bearing a positive electrostatic surface in this region of the S1' pocket (see Figure 5).

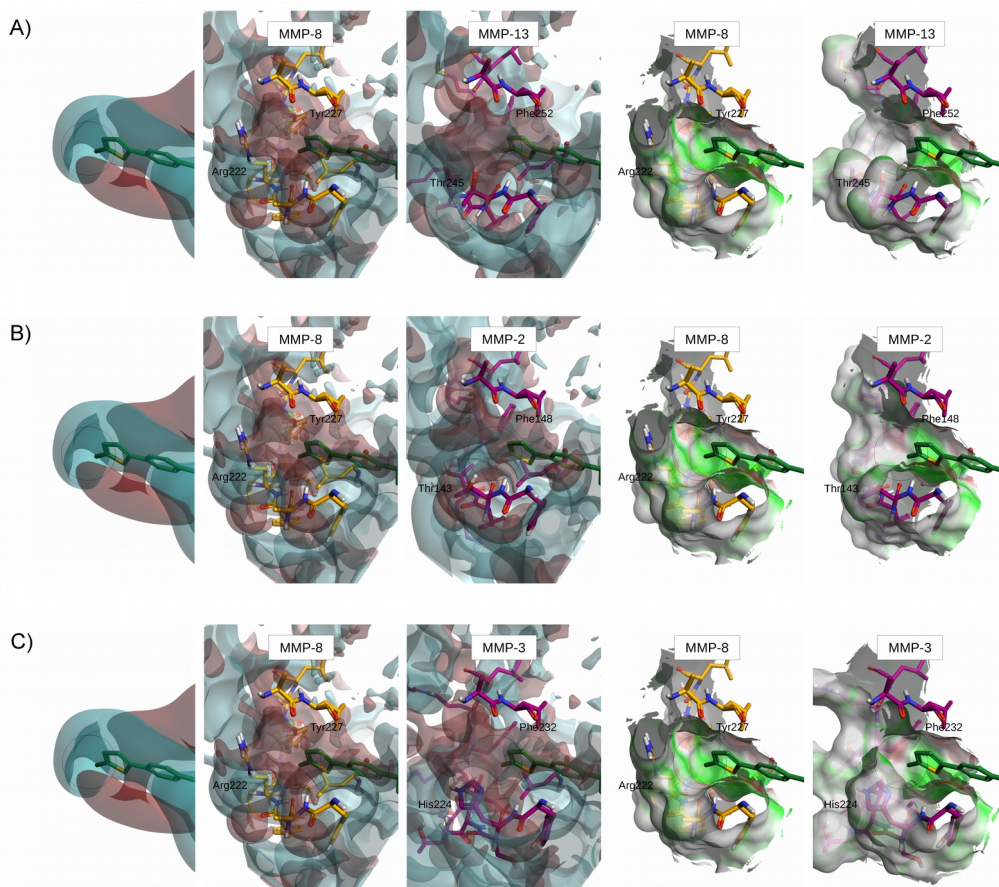


Figure 5. Inhibitor 17,³⁰ selective for MMP-8. This figure has been prepared using the same guidelines as Figure 2. Panels A, B and C show the comparison between the binding site of MMP-8 and the binding sites of MMP-13, MMP-2 and MMP-3, respectively.

Nevertheless, targeting the S1'' pocket in MMP-13 offers the advantage that selectivity against MMP-1, -2, -3, -7, -8, -9, -12 and -14 is generally achieved as these proteins either lack this subpocket or have a smaller one and this constitutes a steric hindrance for MMP-13 inhibitors to bind to these proteins. However, given the large size of the MMP-8 S1' pocket, the selectivity of some MMP-13 inhibitors should not be a consequence of steric hindrances as the inhibitors should fit well in the MMP-8 binding

Manuscript 3

site. This is for instance the case of the MMP-13 inhibitor **15** reported by Taylor *et al.*¹⁷ (see Figure S2N). The electrostatic analysis of this inhibitor in the S1' pocket of MMP-13 reveals that its negatively-charged carboxylic acid group located at the S1'' pocket shows a high electrostatic complementarity with that region of the protein, which has a positive electrostatic potential caused partly by the basic residue Lys140 in MMP-13, which is not present in MMP-8 (see Figure 6). Thus, the presence of this negative group of the ligand boosts the activity of this inhibitor for MMP-13. Moreover, in MMP-8, an acidic residue (*i.e.*, Asp115) is present in this region of the pocket, making the electrostatic environment of MMP-8 less suitable for a negatively charged group (see Figure 6). Therefore, this feature of the ligand may also be responsible for the selectivity observed against MMP-8 (see Figure S2N). This feature is also present in other co-crystallized MMP-13 inhibitors that extend deeply in the S1'' pocket, such as inhibitor **2** reported by Johnson *et al.*¹⁵, inhibitor **29b** reported by Schnute *et al.*¹⁶ and inhibitor **21k** reported by Nara *et al.*¹⁸ (see Figure S2N). All of these inhibitors have been reported to be highly potent and selective against all the other members of the MMP family for which bioactivities have been measured.

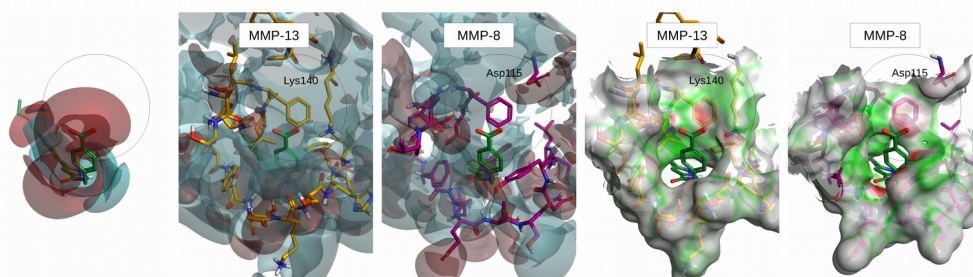


Figure 6. MMP-13 inhibitor **15**,¹⁷ selective over MMP-8. This figure has been prepared using the same guidelines as Figure 2 and it compares the binding site of MMP-13 and the binding site of MMP-8.

Based on this analysis, we can propose two mechanisms for obtaining potency and selectivity for MMP-13 given two characteristic features of this target: **a)** extension to the S1'' pocket, as MMP-13 is the MMP with the deepest S1' pocket; and **b)** incorporation of a negatively charged group in the compound able to establish an electrostatic interaction with the residue Lys140, not present in other MMPs.

4.5. Selectivity against MMP-3 and MMP-8

Overall, in this section, apart from establishing several criteria for finding inhibitors able to attain higher potency for a certain MMP compared to others, we have also acquired novel valuable information on how to avoid the preference of specific targets by inhibitors of other MMPs. In order to avoid the inhibition of MMP-3 and MMP-8, the inhibitor should be able to project a negative electrostatic surface to the S1' pocket of both proteins, so that a repulsive interaction can occur due to their negative potential. In the case of MMP-3, its S1' pocket has been shown to have a more negative electrostatic region than MMPs such as MMP-12 or MMP-13 (see Figure 4). In the case of MMP-8, its S1' pocket has been shown to have a more negative electrostatic character than MMPs such as MMP-2, MMP-12 and MMP-13 (see Figure 2). Therefore, based on these observations, it should be possible to take advantage of the electrostatic characteristics of the S1' pockets of these MMPs by designing MMP inhibitors which incorporated functional groups that would generate a repulsive electrostatic interaction in these environments while maintaining a high affinity for the targeted MMP.

5. Conclusions

In this review we have proposed several mechanisms for MMP inhibitors to achieve selectivity against other MMPs which have been summarized in Table 1. We have shown how to take advantage of differences in the size and shape of the S1' pocket, as is the case for the shallow pockets of MMP-1 and MMP-7, and the deeper pockets of MMP-13 and MMP-8, as well as how to exploit the differences in residues between S1' pockets. More importantly, we have showed that the variability in the S1' pocket characterizes each MMP in terms of hydrophobicity and electrostatic properties and therefore these aspects need to be considered as they constitute an opportunity for new MMP inhibitors to achieve better selectivity profiles.

Manuscript 3

Table 1. Summary of means of achieving selectivity for MMP-2, MMP-9, MMP-12 and MMP-13 against other MMPs. Different colors refer to different means of achieving selectivity.

		Targeted MMP			
		MMP-2	MMP-9	MMP-12	MMP-13
Untargeted MMP	MMP-1	occupy S1' pocket (Figure S2B)	occupy S1' pocket (Figures S2B, S4C)	occupy S1' pocket (Figures S2B, S4B)	occupy S1' pocket (Figures S2B, S4A)
	MMP-2	-	-	maximize hydrophobic contacts in S1' pocket (Figure S2M, Table S22)	extend to S1'' pocket (Figure S2C)
	MMP-3	-	-	maximize hydrophobic contacts in S1' pocket (Figure S2M, Table S22)	extend to S1'' pocket (Figures S2D, S8)
	MMP-7	occupy S1' pocket	occupy S1' pocket (Figure S2E)	occupy S1' pocket (Figure S2E)	occupy S1' pocket (Figures S2E, S10) extend to S1'' pocket (Figure S2E)
	MMP-8	partial negative charge in S1' pocket (Figures S2K, 2A)	-	maximize hydrophobic contacts in S1' pocket (Figure S2M, Table S22) partial negative charge in S1' pocket (Figures S2M, 2B, 2C, 2D)	extend to S1'' pocket (Figures S2F, S12) negative charge in S1'' pocket (Figure 6) partial negative charge in S1' pocket (Figures S2N, 2E, 2F)
	MMP-9	interact with Q-loop (Figure S2K)	-	maximize hydrophobic contacts in S1' pocket (Figure S2M, Table S22)	extend to S1'' pocket (Figure S2G)
	MMP-12	-	smaller inhibitor size (Figure S2L)	-	extend to S1'' pocket (Figures S2I, S17)
	MMP-13	-	-	maximize hydrophobic contacts in S1' pocket (Figure S2M, Table S22)	-
	MMP-14	occupy S1' pocket	occupy S1' pocket	occupy S1' pocket (Figures S2J, S20)	occupy S1' pocket (Figure S2J)

Acknowledgments

This study was supported by research grants 2016PFR-URV-B2-67 and 2017PFR-URV-B2-69 from our university. AG's contract is supported by grant 2015FI_B00655 from the Government of Catalonia. MM is a Serra Hunter research fellow. We thank Cresset BioMolecular Discovery Ltd. for kindly providing us with a software bursary for using their programs. This manuscript has been edited by the English language service of our university.

References

- (1) Tallant, C.; Marrero, A.; Gomis-Rüth, F. X. Matrix Metalloproteinases: Fold and Function of Their Catalytic Domains. *Biochim. Biophys. Acta - Mol. Cell Res.* **2010**, *1803* (1), 20–28.
- (2) Cathcart, J. M.; Cao, J. MMP Inhibitors: Past, Present and Future. *Front. Biosci. (Landmark Ed.)* **2015**, *20*, 1164–1178.
- (3) Vandembroucke, R. E.; Libert, C. Is There New Hope for Therapeutic Matrix Metalloproteinase Inhibition? *Nat. Rev. Drug Discov.* **2014**, *13* (12), 904–927.
- (4) Pulkoski-Gross, A. E. Historical Perspective of Matrix Metalloproteases. *Front. Biosci.* **2015**, *7* (JUNE 2015), 125–149.
- (5) Li, N.-G.; Tang, Y.-P.; Duan, J.-A.; Shi, Z.-H. Matrix Metalloproteinase Inhibitors: A Patent Review (2011 - 2013). *Expert Opin. Ther. Pat.* **2014**, *24* (9), 1039–1052.
- (6) Maradni, A.; Khoshnevisan, A.; Mousavi, S. H.; Emamirazavi, S. H.; Noruzijavidan, A. Role of Matrix Metalloproteinases (MMPS) and MMP Inhibitors on Intracranial Aneurysms: A Review Article. *Med. J. Islam. Repub. Iran* **2013**, *27* (4), 249–254.
- (7) Coussens, L. M.; Fingleton, B.; Matrisian, L. M. Matrix Metalloproteinase Inhibitors and Cancer: Trials and Tribulations. *Science* **2002**, *295* (5564), 2387–2392.
- (8) Fingleton, B. MMPs as Therapeutic Targets--Still a Viable Option? *Semin. Cell Dev. Biol.* **2008**, *19* (1), 61–68.
- (9) Pirard, B. Insight into the Structural Determinants for Selective Inhibition of Matrix Metalloproteinases. *Drug Discov. Today* **2007**, *12* (15–16), 640–646.
- (10) Fabre, B.; Ramos, A.; De Pascual-Teresa, B. Targeting Matrix Metalloproteinases: Exploring the Dynamics of the S1' Pocket in the Design of Selective, Small Molecule Inhibitors. *J. Med. Chem.* **2014**, *57* (24), 10205–10219.
- (11) RCSB PDB <http://www.rcsb.org>.
- (12) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (13) PDB_REDO http://www.cmbi.ru.nl/pdb_redo/.
- (14) Jacobsen, J. A.; Major Jourden, J. L.; Miller, M. T.; Cohen, S. M. To Bind Zinc or Not to Bind Zinc: An Examination of Innovative Approaches to Improved Metalloproteinase Inhibition. *Biochim. Biophys. Acta - Mol. Cell Res.* **2010**, *1803* (1), 72–94.
- (15) Johnson, A. R.; Pavlovsky, A. G.; Ortwine, D. F.; Prior, F.; Man, C. F.; Bornemeier, D. A.; Banotai, C. A.; Mueller, W. T.; McConnell, P.; Yan, C.; et al. Discovery and Characterization of a Novel Inhibitor of Matrix Metalloprotease-13 That Reduces Cartilage Damage in Vivo without Joint Fibroplasia Side Effects. *J. Biol. Chem.* **2007**, *282* (38), 27781–27791.
- (16) Schnute, M. E.; O'Brien, P. M.; Nahra, J.; Morris, M.; Howard Roark, W.; Hanau, C. E.; Ruminski, P. G.; Scholten, J. A.; Fletcher, T. R.; Hamper, B. C.; et al. Discovery of (Pyridin-4-Yl)-2H-Tetrazole as

Manuscript 3

- a Novel Scaffold to Identify Highly Selective Matrix Metalloproteinase-13 Inhibitors for the Treatment of Osteoarthritis. *Bioorg. Med. Chem. Lett.* **2010**, *20* (2), 576–580.
- (17) Taylor, S. J.; Abeywardane, A.; Liang, S.; Muegge, I.; Padyana, A. K.; Xiong, Z.; Hill-Drzewi, M.; Farmer, B.; Li, X.; Collins, B.; et al. Fragment-Based Discovery of Indole Inhibitors of Matrix Metalloproteinase-13. *J. Med. Chem.* **2011**, *54* (23), 8174–8187.
- (18) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Discovery of Novel, Highly Potent, and Selective Quinazoline-2- Carboxamide-Based Matrix Metalloproteinase (MMP)-13 Inhibitors without a Zinc Binding Group Using a Structure-Based Design Approach. *J. Med. Chem.* **2014**, *57* (21), 8886–8902.
- (19) Pochetti, G.; Montanari, R.; Gege, C.; Chevrier, C.; Taveras, A. G.; Mazza, F. Extra Binding Region Induced by Non-Zinc Chelating Inhibitors into the S1' Subsite of Matrix Metalloproteinase 8 (MMP-8). *J. Med. Chem.* **2009**, *52* (4), 1040–1049.
- (20) Morales, R.; Perrier, S.; Florent, J. M.; Beltra, J.; Dufour, S.; De Mendez, I.; Manceau, P.; Tertre, A.; Moreau, F.; Compere, D.; et al. Crystal Structures of Novel Non-Peptidic, Non-Zinc Chelating Inhibitors Bound to MMP-12. *J. Mol. Biol.* **2004**, *341* (4), 1063–1076.
- (21) Rouanet-Méhouas, C.; Czarny, B.; Beau, F.; Cassar-Lajeunesse, E.; Stura, E. A.; Dive, V.; Devel, L.; Rouanet-Mehouas, C.; Czarny, B.; Beau, F.; et al. Zinc-Metalloproteinase Inhibitors: Evaluation of the Complex Role Played by the Zinc-Binding Group on Potency and Selectivity. *J. Med. Chem.* **2017**, *60* (1), 403–414.
- (22) Cereto-Massagué, A.; Ojeda, M. J.; Joosten, R. P.; Valls, C.; Mulero, M.; Salvado, M. J.; Arola-Arnal, A.; Arola, L.; Garcia-Vallvé, S.; Pujadas, G. The Good, the Bad and the Dubious: VHELIBS, a Validation Helper for Ligands and Binding Sites. *J. Cheminform.* **2013**, *5* (1), 36.
- (23) Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>.
- (24) Schrödinger Release 2018-1: Maestro, Schrödinger, LLC, New York, NY, 2018.
- (25) Holmes, I. P. P.; Gaines, S.; Watson, S. P. P.; Lorthioir, O.; Walker, A.; Baddeley, S. J. J.; Herbert, S.; Egan, D.; Convery, M. A. M. A. M. A. A.; Singh, O. M. P. M. P.; et al. The Identification of β -Hydroxy Carboxylic Acids as Selective MMP-12 Inhibitors. *Bioorganic Med. Chem. Lett.* **2009**, *19* (19), 5760–5763.
- (26) Heim-Riether, A.; Taylor, S. J.; Liang, S.; Gao, D. A.; Xiong, Z.; Michael August, E.; Collins, B. K.; Farmer, B. T.; Haverty, K.; Hill-Drzewi, M.; et al. Improving Potency and Selectivity of a New Class of Non-Zn-Chelating MMP-13 Inhibitors. *Bioorganic Med. Chem. Lett.* **2009**, *19* (18), 5321–5324.
- (27) Shieh, H. S.; Tomasselli, A. G.; Mathis, K. J.; Schnute, M. E.; Woodard, S. S.; Caspers, N.; Williams, J. M.; Kiefer, J. R.; Munie, G.; Wittwer, A.; et al. Structure Analysis Reveals the Flexibility of the ADAMTS-5 Active Site. *Protein Sci.* **2011**, *20* (4), 735–744.
- (28) Becker, D. P.; Barta, T. E.; Bedell, L. J.; Boehm, T. L.; Bond, B. R.; Carroll, J.; Carron, C. P.; Decrescenzo, G. A.; Easton, A. M.; Freskos, J. N.; et al. Orally Active MMP-1 Sparing α -Tetrahydropyryl and α -Piperidinyl Sulfone Matrix Metalloproteinase (MMP) Inhibitors with Efficacy in Cancer, Arthritis, and Cardiovascular Disease. *J. Med. Chem.* **2010**, *53* (18), 6653–6680.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

- (29) Monovich, L. G.; Tommasi, R. A.; Fujimoto, R. A.; Blancuzzi, V.; Clark, K.; Cornell, W. D.; Doti, R.; Doughty, J.; Fang, J.; Farley, D.; et al. Discovery of Potent, Selective, and Orally Active Carboxylic Acid Based Inhibitors of Matrix Metalloproteinase-13. *J Med Chem* **2009**, *52* (11), 3523–3538.
- (30) Devel, L.; Beau, F.; Amoura, M.; Vera, L.; Cassar-Lajeunesse, E.; Garcia, S.; Czarny, B.; Stura, E. A.; Dive, V. Simple Pseudo-Dipeptides with a P2' Glutamate: A Novel Inhibitor Family of Matrix Metalloproteases and Other Metzincins. *J. Biol. Chem.* **2012**, *287* (32), 26647–26656.
- (31) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Thieno[2,3-d]Pyrimidine-2-Carboxamides Bearing a Carboxybenzene Group at 5-Position: Highly Potent, Selective, and Orally Available MMP-13 Inhibitors Interacting with the S1'' Binding Site. *Bioorganic Med. Chem.* **2014**, *22* (19), 5487–5505.
- (32) Tochowicz, A.; Maskos, K.; Huber, R.; Oltenfreiter, R.; Dive, V.; Yiotakis, A.; Zanda, M.; Bode, W.; Goettig, P. Crystal Structures of MMP-9 Complexes with Five Inhibitors: Contribution of the Flexible Arg424 Side-Chain to Selectivity. *J. Mol. Biol.* **2007**, *371* (4), 989–1006.
- (33) Nuti, E.; Cantelmo, A. R.; Gallo, C.; Bruno, A.; Bassani, B.; Camodeca, C.; Tuccinardi, T.; Vera, L.; Orlandini, E.; Nencetti, S.; et al. N-O-Isopropyl Sulfonamido-Based Hydroxamates as Matrix Metalloproteinase Inhibitors: Hit Selection and in Vivo Antiangiogenic Activity. *J. Med. Chem.* **2015**, *58* (18), 7224–7240.
- (34) Camodeca, C.; Nuti, E.; Tepshi, L.; Boero, S.; Tuccinardi, T.; Stura, E. A.; Poggi, A.; Zocchi, M. R.; Rossello, A. Discovery of a New Selective Inhibitor of A Disintegrin and Metalloprotease 10 (ADAM-10) Able to Reduce the Shedding of NKG2D Ligands in Hodgkin's Lymphoma Cell Models. *Eur. J. Med. Chem.* **2016**, *111*, 193–201.
- (35) Nuti, E.; Cuffaro, D.; D'Andrea, F.; Rosalia, L.; Tepshi, L.; Fabbi, M.; Carbotti, G.; Ferrini, S.; Santamaria, S.; Camodeca, C.; et al. Sugar-Based Arylsulfonamide Carboxylates as Selective and Water-Soluble Matrix Metalloproteinase-12 Inhibitors. *ChemMedChem* **2016**, *11* (15), 1626–1637.
- (36) Mannino, C.; Nievo, M.; Machetti, F.; Papakyriakou, A.; Calderone, V.; Fragai, M.; Guarna, A. Synthesis of Bicyclic Molecular Scaffolds (BTAA): An Investigation towards New Selective MMP-12 Inhibitors. *Bioorganic Med. Chem.* **2006**, *14* (22), 7392–7403.
- (37) Devel, L.; Garcia, S.; Czarny, B.; Beau, F.; Lajeunesse, E.; Vera, L.; Georgiadis, D.; Stura, E.; Dive, V. Insights from Selective Non-Phosphinic Inhibitors of MMP-12 Tailored to Fit with an S1' Loop Canonical Conformation. *J. Biol. Chem.* **2010**, *285* (46), 35900–35909.
- (38) Edman, K.; Furber, M.; Hemsley, P.; Johansson, C.; Pairedeau, G.; Petersen, J.; Stocks, M.; Tervo, A.; Ward, A.; Wells, E.; et al. The Discovery of MMP7 Inhibitors Exploiting a Novel Selectivity Trigger. *ChemMedChem* **2011**, *6* (5), 769–773.
- (39) Mahasenan, K. V.; Bastian, M.; Gao, M.; Frost, E.; Ding, D.; Zorina-Lichtenwalter, K.; Jacobs, J.; Suckow, M. A.; Schroeder, V. A.; Wolter, W. R.; et al. Exploitation of Conformational Dynamics in Imparting Selective Inhibition for Related Matrix Metalloproteinases. *ACS Med. Chem. Lett.* **2017**, *8* (6), 654–659.
- (40) Savi, C. De; Morley, A. D.; Ting, A.; Nash, I.; Karabelas, K.; Wood, C. M.; James, M.; Norris, S. J.; Karoutchi, G.; Rankine, N.; et al. Selective Non Zinc Binding Inhibitors of MMP13. *Bioorganic Med. Chem. Lett.* **2011**, *21* (14), 4215–4219.

Manuscript 3

- (41) Alcantara, M. B.; Dass, C. R. Pigment Epithelium-Derived Factor as a Natural Matrix Metalloproteinase Inhibitor: A Comparison with Classical Matrix Metalloproteinase Inhibitors Used for Cancer Treatment. *J. Pharm. Pharmacol.* **2014**, *66* (7), 895–902.
- (42) Adhikari, N.; Halder, A. K.; Mallick, S.; Saha, A.; Saha, K. D.; Jha, T. 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. *Bioorg. Med. Chem.* **2016**, *24* (18), 4291–4309.
- (43) Hu, J.; Van den Steen, P. E.; Sang, Q.-X. a; Opdenakker, G. Matrix Metalloproteinase Inhibitors as Therapy for Inflammatory and Vascular Diseases. *Nat. Rev. Drug Discov.* **2007**, *6* (6), 480–498.
- (44) Campestre, C.; Agamennone, M.; Tortorella, P.; Preziuso, S.; Biasone, A.; Gavuzzo, E.; Pochetti, G.; Mazza, F.; Hiller, O.; Tschesche, H.; et al. N-Hydroxyurea as Zinc Binding Group in Matrix Metalloproteinase Inhibition: Mode of Binding in a Complex with MMP-8. *Bioorganic Med. Chem. Lett.* **2006**, *16* (1), 20–24.
- (45) Flare, v2.0 , Cresset®, Litlington, Cambridgeshire, UK, ; <http://www.cresset-group.com/flare/>; Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.*

Supplementary material for

***Understanding the variability of the S1' pocket to
improve matrix metalloproteinase inhibitor
selectivity profiles***

Aleix Gimeno^[a], Raúl Beltrán-Debón^[a], Miquel Mulero^[a], Gerard Pujadas^{[a],[b].*}, Santiago
Garcia-Vallvé^{[a],[b]}

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S1. MMPs, their alternative names, functional classification, main substrates and some of the pathologies in which they are involved.

MMP	Alternative names	Functional classification	Main substrates	Related Pathologies
MMP-1	- Interstitial collagenase - Collagenase 1 - Vertebrate collagenase	- Collagenases	- Collagen types I, II and III	- Rheumatoid arthritis ⁴ - Cancer ^{1,8}
MMP-2	- Gelatinase A - 72 kDa gelatinase - Type IV collagenase	- Gelatinases	- Collagen type IV - Gelatin - Fibrinogen	- Asthma ^{4,9} - Cancer ¹⁰ - Cardiovascular diseases ⁴ - Heart failure ⁴ - Liver fibrosis ¹¹
MMP-3	- Stromelysin 1 - Proteoglycanase - Transin	- Stromelysins	- Proteoglycan - Fibronectin - Collagen types I, III, IV, V and IX	- Atherosclerosis ¹³ - Coronary artery disease ¹⁴ - Inflammatory bowel disease ⁴ - Periodontitis ¹²
MMP-7	- Matrilysin - PUMP-1 - Putative metalloproteinase-1 - Uterine metalloendopeptidase	- Matrilysins	- Elastin - Fibronectin - Casein - Laminin	- Cancer ¹⁰ - Inflammatory bowel disease ⁵ - Lung fibrosis ⁴
MMP-8	- Neutrophil collagenase	- Collagenases	- Collagen types I and III	- Asthma ⁴ - Cancer ¹ - Periodontitis ⁴ - Rheumatoid arthritis ¹³
MMP-9	- Gelatinase B - 92 kDa gelatinase - Type IV collagenase - Macrophage gelatinase	- Gelatinases	- Collagen types I, III, IV, V and XI - Gelatin types I and V - Laminin	- Asthma ⁴ - Cancer ¹¹ - Heart failure ⁴ - Inflammatory bowel disease ¹² - Rheumatoid arthritis ⁹ - Liver fibrosis ¹⁵

Manuscript 3

MMP-10	- Stromelysin 2 - Transin 2	- Stromelysins	- Fibronectin - Proteoglycan - Gelatin types I, III, IV, V - Collagen types I, III, IV and V	- Atherothrombosis ¹⁷ - Chronic obstructive pulmonary disease ¹⁶
MMP-12	- Macrophage elastase - Metalloelastase	- Metalloelastases	- Elastin - Casein - Fibronectin - Gelatin - Laminin - Collagen type IV	- Chronic obstructive pulmonary disease ¹⁸ - Neurological diseases ⁴
MMP-13	- Collagenase 3	- Collagenases	- Collagen types I, II, III - Fibrillin types 1 and 2	- Cancer ¹ - Inflammatory bowel disease ⁴ - Osteoarthritis ⁵ - Rheumatoid arthritis ² - Obesity ³
MMP-14	- Membrane-type matrix metalloproteinase-1	- Membrane-type MMPs	- Collagen type I - Fibronectin - Laminin	- Aortic aneurysm ⁷ - Cancer ⁶

This information has been obtained from the BRENDA,¹⁹ KEGG,²⁰ ExPASy²¹ and MEROPS²² databases.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S2. Summary of manuscripts reporting MMP inhibitors since 2010.

MMP	Number of manuscripts identifying inhibitors for each MMP since 2010	Number of manuscripts in which MMP bioactivities are measured									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
MMP-1	2	2	1	1	0	0	1	0	1	1	0
MMP-2	16	6	16	3	4	6	11	1	3	3	3
MMP-3	0	0	0	0	0	0	0	0	0	0	0
MMP-7	1	1	1	0	1	0	0	0	1	1	1
MMP-8	1	0	1	0	0	1	0	0	0	0	0
MMP-9	10	1	7	0	0	0	10	1	0	0	1
MMP-10	1	0	0	0	0	0	0	1	0	1	0
MMP-12	9	5	7	4	4	5	6	2	9	5	4
MMP-13	25	13	19	14	11	14	14	7	12	25	17
MMP-14	2	2	2	0	0	1	2	2	0	0	2

Table S3. Manuscripts reporting compounds with inhibitory activity against MMP-1 since 2010.^{23,24}

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
29270028	2018	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
23353736	2013	✓	✓	✓	✗	✗	✓	✗	✓	✓	✗

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Manuscript 3

Table S4. Manuscripts reporting compounds with inhibitory activity against MMP-2 since 2010.²⁵⁻⁴⁰

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
30342958	2018	x	✓	x	x	✓	x	x	x	x	x
29614516	2018	x	✓	x	x	x	x	x	x	x	x
29674965	2018	✓	✓	x	x	x	✓	x	x	✓	x
27452283	2016	✓	✓	x	x	✓	✓	x	✓	x	✓
27455162	2016	x	✓	x	x	x	✓	x	x	x	x
27038494	2016	x	✓	x	x	x	✓	x	x	x	x
26061284	2015	x	✓	x	x	x	x	x	x	x	x
26346367	2015	x	✓	x	x	x	✓	x	x	x	x
25418204	2015	x	✓	x	x	x	x	x	✓	x	x
25907368	2015	x	✓	x	x	x	✓	x	x	x	x
24246091	2014	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
24028490	2013	✓	✓	✓	✓	✓	✓	x	x	x	✓
23989288	2013	x	✓	x	x	x	✓	x	x	x	x
23873724	2013	✓	✓	x	✓	✓	x	x	x	✓	x
23395821	2013	✓	✓	✓	✓	✓	✓	x	x	x	x
21552627	2011	x	✓	x	x	x	✓	x	x	x	x

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Table S5. Manuscripts reporting compounds with inhibitory activity against MMP-7 since 2010.⁴¹

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
21520417	2011	✓	✓	x	✓	x	x	x	✓	✓	✓

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S6. Manuscripts reporting compounds with inhibitory activity against MMP-9 since 2010.^{35,38,40,42-48}

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
30079817	2018	x	x	x	x	x	✓	x	x	x	x
29473954	2018	✓	✓	x	x	x	✓	✓	x	x	✓
28623487	2017	x	x	x	x	x	✓	x	x	x	x
27317634	2016	x	✓	x	x	x	✓	x	x	x	x
27455162	2016	x	✓	x	x	x	✓	x	x	x	x
26346367	2015	x	✓	x	x	x	✓	x	x	x	x
25907368	2015	x	✓	x	x	x	✓	x	x	x	x
24473069	2014	x	x	x	x	x	✓	x	x	x	x
22520332	2012	x	✓	x	x	x	✓	x	x	x	x
22248361	2012	x	✓	x	x	x	✓	x	x	x	x

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Table S7. Manuscripts reporting compounds with inhibitory activity against MMP-10 since 2010.⁴⁹

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
28953404	2017	x	x	x	x	x	x	✓	x	✓	x

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Manuscript 3

Table S8. Manuscripts reporting compounds with inhibitory activity against MMP-12 since 2010.⁵⁰⁻⁵⁸

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
29727184	2018	✓	✓	✗	✗	✗	✓	✗	✓	✗	✓
29660282	2018	✗	✓	✗	✗	✓	✓	✗	✓	✓	✗
27356908	2016	✓	✓	✗	✗	✗	✓	✗	✓	✗	✓
26351407	2015	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
22771631	2012	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗
22153340	2012	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗
21111619	2012	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
21532196	2011	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
20817735	2010	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S9. Manuscripts reporting compounds with inhibitory activity against MMP-13 since 2010.^{54,59-82}

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
28814230	2017	-	-	-	-	-	-	-	-	✓	-
28653849	2017	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
27966948	2017	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
27981835	2017	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓
27825552	2016	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
27362887	2016	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
26938528	2016	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
26653735	2016	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
25330343	2014	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
25264600	2014	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
25192810	2014	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
23894097	2013	✗	✓	✗	✗	✗	✗	✗	✓	✓	✓
23810497	2013	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓
22771631	2012	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗
22175799	2012	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
22153941	2012	✗	✓	✗	✗	✗	✓	✗	✓	✓	✓
22088955	2012	✓	✓	✓	✗	✗	✗	✗	✗	✓	✓
22017539	2011	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
22018790	2011	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓
21937229	2011	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗
21669521	2011	✗	✓	✗	✗	✗	✓	✗	✓	✓	✓
21661113	2011	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗
20675133	2010	✗	✓	✓	✗	✓	✗	✓	✗	✓	✗
19902332	2010	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
20005097	2010	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Manuscript 3

Table S10. Manuscripts reporting compounds with inhibitory activity against MMP-14 since 2010.^{43,83}

PubMed Identifier	Year	Measured activities									
		MMP-1	MMP-2	MMP-3	MMP-7	MMP-8	MMP-9	MMP-10	MMP-12	MMP-13	MMP-14
29473954	2018	✓	✓	✗	✗	✗	✓	✓	✗	✗	✓
28087697	2017	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓

MMPs for which bioactivities have been measured have been labeled with a tick, and MMPs for which bioactivities have not been measured have been labeled with a cross.

Table S11. Ω -loop length and residue sequence for MMP-1, -2, -3, -7, -8, -9, -10, -12, -13, -14.

MMP	Number of residues in Ω -loop	Ω -loop sequence
MMP-1	11	238 – PSYTFSGDVQL – 248
MMP-2	11	140 – PIYTYTKNFRL – 150
MMP-3	14	221 – PLYHSLDLTRFRL – 234
MMP-7	13	239 – PTYGNGDPQNFKL – 251
MMP-8	13	217 – PNYAFRETSNYSL – 229
MMP-9	11	246 – PMYRFTEGPPL – 256
MMP-10	14	237 – PLYNSFTELAQFRL – 250
MMP-12	13	238 – PTYKYVDINTFRL – 250
MMP-13	13	242 – PIYTYTGKSHFML – 254
MMP-14	13	259 – PFYQWMDTENFVL – 271

The residues conserved among all MMPs are highlighted. The numbers before and after each sequence segment indicate the location of the first and last segment residue for each MMP.

Table S12. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-1 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-1 crystal structure subunits
A	2J0T (A)	2CLT (B), 2J0T (A), 2J0T (B), 2J0T (C), 3SHI (A), 3SHI (G), 3SHI (M)

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. All the structures correspond to apo structures.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S13. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-2 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-2 crystal structure subunits
A	3AYU (A)	3AYU (A)

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. All the structures correspond to apo structures.

Table S14. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-3 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-3 crystal structure subunits
A	<u>1HFS</u> (A)	<u>1C3I</u> (B), <u>1C8T</u> (B), <u>1HFS</u> (A), 1SLM (A), 1SLN (A), 4DPE (A), 4JA1 (A)
B	4G9L (A)	1C3I (A), 4G9L (A)
C	2D1O (A)	2D1O (A), 2D1O (B)
D	4DPE (B)	<u>4DPE</u> (B), <u>4G9L</u> (B), <u>4JA1</u> (B)
E	1C8T (A)	1C8T (A)
n/a	-	<u>3OHL</u> (A), <u>3OHO</u> (A)

The chain label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. Subunits that belong to the n/a set lack coordinates for some of their Ω -loop residues in the corresponding PDB file and, therefore, they cannot be classified by the conformation of that loop (see also Figure S7G). The subunits that correspond to holo forms have been underlined.

Manuscript 3

Table S15. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-7 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-7 crystal structure subunits
A	<u>2Y6C</u> (A)	<u>2Y6C</u> (A)
B	<u>2Y6D</u> (A)	<u>2Y6D</u> (A)

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. All subunits have been underlined indicating that they correspond to holo forms.

Table S16. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-8 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-8 crystal structure subunits
A	1I73 (A)	<u>1BZS</u> (A), <u>1I73</u> (A), <u>1I76</u> (A), <u>1JH1</u> (A), <u>1JJ9</u> (A), <u>1ZP5</u> (A), <u>1ZS0</u> (A), <u>1ZVX</u> (A), <u>2OY2</u> (A), <u>2OY2</u> (F), <u>2OY4</u> (A), <u>2OY4</u> (F), <u>3TT4</u> (A), <u>4QKZ</u> (A)
B	3DNG (A)	<u>3DNG</u> (A), <u>3DPE</u> (A), <u>3DPF</u> (A), <u>3DPF</u> (B)
C	5H8X (A)	5H8X (A)

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. The subunits that correspond to holo forms have been underlined.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S17. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-9 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-9 crystal structure subunits
A	4WZV (A)	<u>1GKC (A)</u> , <u>1GKC (B)</u> , <u>1GKD (A)</u> , <u>1GKD (B)</u> , <u>2OVX (A)</u> , <u>2OVX (B)</u> , <u>2OVZ(A)</u> , <u>2OVZ(B)</u> , <u>2OW0 (A)</u> , <u>2OW0 (B)</u> , <u>2OW1 (A)</u> , <u>2OW1 (B)</u> , <u>2OW2 (A)</u> , <u>4H3X(A)</u> , <u>4H3X(B)</u> , <u>4H82 (A)</u> , <u>4H82 (B)</u> , <u>4H82 (C)</u> , <u>4H82 (D)</u> , <u>4HMA(B)</u> , <u>4HMA (A)</u> , <u>4JIJ (A)</u> , <u>4JIJ (B)</u> , <u>4JQG (A)</u> , <u>4JQG (B)</u> , <u>4WZV (A)</u> , <u>4WZV (B)</u> , <u>4XCT (A)</u> , <u>5CUH (A)</u> , <u>5CUH (B)</u> , <u>5I12 (A)</u>
B	<u>4H2E (A)</u>	<u>4H2E (A)</u> , <u>4H2E (B)</u>

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. The subunits that correspond to holo forms have been underlined.

Table S18. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-10.

Ω -loop conformation	Representative subunit	MMP-10 crystal structure subunits
n/a	-	<u>1Q3A (A)</u> , <u>1Q3A (B)</u> , <u>1Q3A (C)</u> , <u>3V96 (B)</u> , <u>4ILW (D)</u> , <u>4ILW (F)</u>

All of these structures lack coordinates for some of their Ω -loop residues in the corresponding PDB file and, therefore, they cannot be classified by the conformation of that loop (see also Figure S15). The subunit label is shown in parentheses next to the PDB ID of the structure. The subunits that correspond to holo forms have been underlined.

Manuscript 3

Table S19. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-12 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-12 crystal structure subunits
A	1JIZ (A)	1JIZ (A), 1JIZ (B), <u>1OS2 (A)</u> , 1OS2 (B), 1OS2 (C), <u>1OS2 (D)</u> , 1OS2 (E), 1OS2 (F), 1OS9 (A), 1OS9 (B), 1OS9 (C), 1OS9 (D), 1OS9 (E), 1OS9 (F), <u>1ROS (A)</u> , <u>1ROS (B)</u> , <u>1RMZ (A)</u> , <u>1UTT (A)</u> , <u>1UTZ (A)</u> , <u>1UTZ (B)</u> , <u>1Y93 (A)</u> , <u>2HU6 (A)</u> , 2OXU (A), 2OXW (A), 2OXZ (A), 2W0D (A), 2W0D (B), 2W0D (C), 2W0D (D), <u>2WO8 (A)</u> , <u>2WO8 (B)</u> , <u>2WO8 (C)</u> , <u>2WO8 (D)</u> , <u>2WO9 (A)</u> , <u>2WO9 (B)</u> , <u>2WO9 (D)</u> , <u>2WOA (B)</u> , <u>2WOA (C)</u> , <u>2WOA (D)</u> , <u>3EHX (A)</u> , <u>3EHY (A)</u> , <u>3F1A (A)</u> , <u>3F15 (A)</u> , <u>3F16 (A)</u> , <u>3F17 (A)</u> , <u>3F18 (A)</u> , <u>3F19 (A)</u> , <u>3LIK (A)</u> , <u>3LIL (A)</u> , <u>3LIR (A)</u> , <u>3LJG (A)</u> , <u>3LK8 (A)</u> , <u>3LKA (A)</u> , <u>3N2U (A)</u> , <u>3N2V (A)</u> , <u>3NX7 (A)</u> , <u>3RTS (A)</u> , <u>3RTT (A)</u> , <u>3TS4 (A)</u> , <u>3TSK (A)</u> , <u>3UVC (A)</u> , <u>3UVC (B)</u> , <u>4EFS (A)</u> , <u>4GR0 (A)</u> , <u>4GR3 (A)</u> , <u>4GR8 (A)</u> , <u>4GUY (A)</u> , 4H30 (A), <u>4H30 (B)</u> , 4H49 (A), <u>4H49 (B)</u> , <u>4H49 (C)</u> , 4H49 (D), 4I03 (A), 4IJO (A), <u>4H76 (A)</u> , <u>4H84 (A)</u> , <u>4H84 (B)</u>
B	3BA0 (A)	3BA0 (A)

The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. The subunits that correspond to holo forms have been underlined.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Table S20. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-13 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-13 crystal structure subunits
A	4FU4 (B)	<u>4FU4 (A)</u> , <u>4FU4 (B)</u> , <u>4FVL (A)</u> , <u>4FVL (B)</u> , <u>4G0D (A)</u> , <u>4G0D (B)</u> , <u>4G0D (C)</u> , <u>4G0D (D)</u>
B	2YIG (B)	<u>2YIG (B)</u> , <u>3ELM (B)</u> , <u>3WV3 (B)</u> , <u>3ZXH (B)</u> , <u>4JP4 (B)</u> , <u>4JPA (B)</u> , <u>4L19 (A)</u> , <u>5B5O (B)</u> , <u>5B5P (B)</u> , <u>5BOT (B)</u> , <u>5BOY (B)</u> , <u>5BPA (A)</u> , <u>5BPA (B)</u>
C	<u>1YOU (A)</u>	<u>1YOU (A)</u> , <u>1YOU (B)</u>
D	2D1N (B)	<u>2D1N (A)</u> , <u>2D1N (B)</u> , <u>3TVC (A)</u> , <u>4A7B (A)</u>
E	<u>3KRY (A)</u>	<u>3KRY (A)</u> , <u>3KRY (B)</u>
F	<u>3KRY (C)</u>	<u>3KRY (C)</u>
G	3WV1 (B)	<u>2OW9 (A)</u> , <u>2OW9 (B)</u> , <u>3KEC (A)</u> , <u>3KEC (B)</u> , <u>3KEJ (A)</u> , <u>3KEJ (B)</u> , <u>3KEK (A)</u> , <u>3KEK (B)</u> , <u>3WV1 (A)</u> , <u>3WV1 (B)</u>
H	<u>2OZR (A)</u>	<u>2OZR (A)</u> , <u>2OZR (B)</u> , <u>2OZR (C)</u> , <u>2OZR (D)</u>
I	<u>2OZR (F)</u>	<u>2OZR (F)</u> , <u>2OZR (G)</u>
n/a	-	<u>2OZR (E)</u> , <u>2OZR (H)</u> , <u>2YIG (A)</u> , <u>3ELM (A)</u> , <u>3I7G (A)</u> , <u>3I7G (B)</u> , <u>3I7I (A)</u> , <u>3I7I (B)</u> , <u>3KRY (D)</u> , <u>3LJZ (B)</u> , <u>3LJZ (C)</u> , <u>3LJZ (D)</u> , <u>3O2X (A)</u> , <u>3O2X (B)</u> , <u>3O2X (C)</u> , <u>3WV2 (B)</u> , <u>3WV3 (A)</u> , <u>3ZXH (A)</u> , <u>4JP4 (A)</u> , <u>4JPA (A)</u> , <u>5B5O (A)</u> , <u>5B5P (A)</u> , <u>5BOT (A)</u> , <u>5BOY (A)</u>

Subunits that belong to the n/a set lack coordinates for some residues of their Ω -loop in the corresponding PDB file and, therefore, they cannot be classified by the conformation of that loop (see also Figure S18K). The subunit label is shown in parentheses next to the PDB ID of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. The subunits that correspond to holo forms have been underlined.

Table S21. Subunits of experimental structures deposited in the PDB^{84,85} with available EDMs for human MMP-14 classified by the conformation of their Ω -loop.

Ω -loop conformation	Representative subunit	MMP-14 crystal structure subunits
A	3MA2 (A)	3MA2 (A), 3MA2 (D)

The subunit label is shown in parentheses next to the PDB^{84,85} code of the structure. A representative subunit presenting the specified conformation of the Ω -loop is shown. All the structures correspond to apo structures.

Manuscript 3

Table S22. *LipophilicEVdW* term obtained from protein-ligand docking using GlideXP^{84,85} for different compounds against different MMPs.

Compound	<i>LipophilicEVdW</i>					
	MMP-2	MMP-3	MMP-8	MMP-9	MMP-12	MMP-13
3	-5.52	-5.99	-6.48	-5.66	-7.19 ^c	-5.7
16	-5.25	-5.53	-5.63	-5.37	-6.67 ^a	-5.41
PD-0359601	-5.91	-5.95	-5.46	-5.37	-6.54 ^b	-5.46
CP-271485	-4.98	-5.39	-4.96	-5.38	-5.41	-4.21
PF-00356231	-5.92	-6.49	-6.57	-6.08	-7.00 ^c	-6.29
1	-5.21	-5.26	-5.72	-4.78	-5.41	-5.58
6c	-5.30	-5.83	-5.35	-4.92	-6.17	-5.40
7a	-5.26	-5.62	-5.79	-4.91	-6.53 ^c	-5.39
36	-5.69	-6.37	-6.54	-5.80	-6.88	-6.48

LipophilicEVdW corresponds to a term from the GlideXP scoring function which is derived from the hydrophobic grid potential at the hydrophobic ligand atoms. The MMP-2, -3, -8, -9 and -13 structures used for docking were the structures with PDB⁸⁶ codes 3AYU (A),^{84,85} 1HFS (A),⁸⁷ 1I73 (A),⁸⁸ 4WZV (A)⁸⁹ and 2YIG (A),⁹⁰ respectively. The MMP-12 structure used for docking in each case corresponded to the structure with which the compound was co-crystallized (i.e., structures with the PDB IDs 3TS4 (A),⁷¹ 4EFS (A),⁹¹ 1ROS (A),⁹¹ 1UTT (A),⁹⁰ 1UTZ (A),⁹⁰ 2WO8 (A),⁹⁰ 2WO9 (A),⁹² 2WOA (C)⁹² and 3LIK (A)⁹² for inhibitors 3,⁵⁸ 16,⁹¹ PD-0359601,⁹¹ CP-271485,⁹⁰ PF-00356231,⁹⁰ 1,⁹⁰ 6c,⁹² 7a⁹² and 36,⁹² respectively.). ^{a, b, c} Statistically significant outlier values determined by Dixon's Q test at 99%, 90% and 80% confidence, respectively.

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

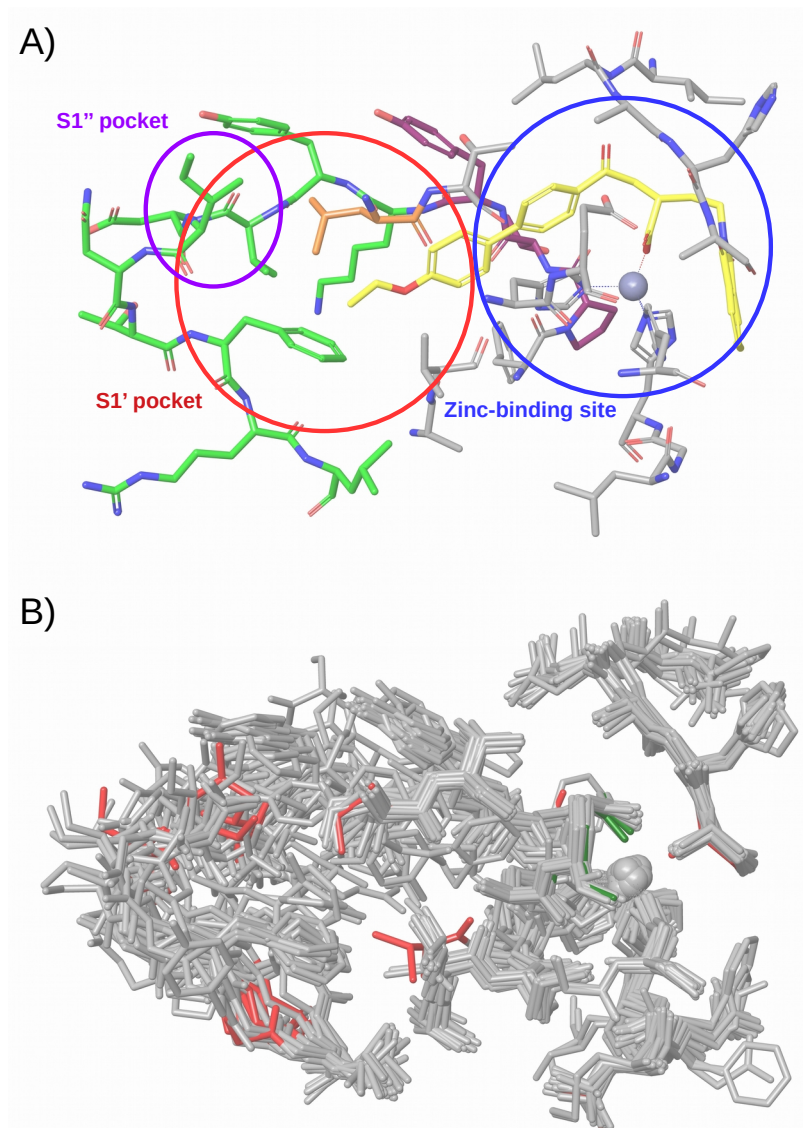
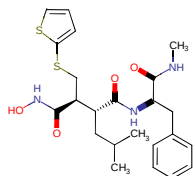


Figure S1. Binding site of MMPs. In Panel A, the binding site of MMP-12 is represented (structure with PDB⁵⁸ ID 1ROS^{84,85}), showing the different regions of the binding site of MMPs. The residues of the Ω -loop are in green, the residues of the wall-forming segment are in purple and other residues of the binding site are in gray. The ligand is in yellow and the catalytic zinc ion is represented in spacefill format. In Panel B, a superposition of different structures of MMP-1, -2, -3, -7, -8, -9, -12, -13 and -14 is shown. These structures correspond to the representative subunits from Tables S12 to S21. Residues for which the electron density has been considered insufficient to ensure the correctness of their conformation are shown in red. Residues corresponding to non-natural mutations are in green. Both panels show the same orientation to facilitate comparison. This figure was obtained with Maestro⁹⁰ v11.

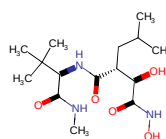
Manuscript 3

A)



Batimastat

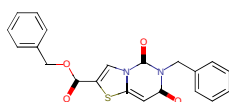
MMP-1 IC₅₀ = 0.4 nM MMP-8 IC₅₀ = 0.7 nM
MMP-2 IC₅₀ = 1 nM MMP-9 IC₅₀ = 0.2 nM
MMP-3 IC₅₀ = 0.65 nM MMP-13 IC₅₀ = 1 nM
MMP-7 IC₅₀ = 2.4 nM MMP-14 IC₅₀ = 2.8 nM



Marimastat

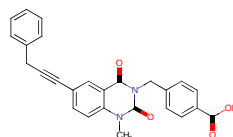
MMP-1 IC₅₀ = 0.76 nM MMP-8 IC₅₀ = 0.53 nM
MMP-2 IC₅₀ = 0.5 nM MMP-9 IC₅₀ = 1.5 nM
MMP-3 IC₅₀ = 4.4 nM MMP-12 IC₅₀ < 5 nM
MMP-7 IC₅₀ = 2 nM MMP-13 IC₅₀ = 0.74 nM

B) Johnson *et al.* (PubMed Identifier: 17623656)



1

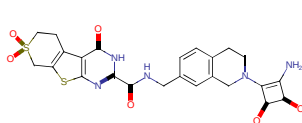
MMP-13 IC₅₀ = 30 nM
MMP-1 IC₅₀ > 100,000 nM



2

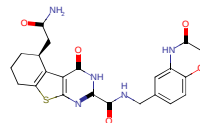
MMP-13 IC₅₀ = 0.67 nM
MMP-1 IC₅₀ > 30,000 nM

Pochetti *et al.* (PubMed Identifier: 19173605)



1

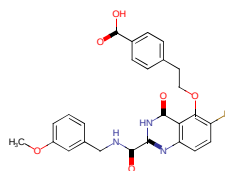
MMP-8 IC₅₀ = 57 nM
MMP-1 IC₅₀ > 10,000 nM



2

MMP-8 IC₅₀ = 7.4 nM
MMP-1 IC₅₀ > 10,000 nM

Nara *et al.* (PubMed Identifier: 25264600)

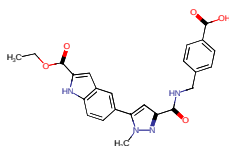


21k

MMP-13 IC₅₀ = 0.0039 nM
MMP-1 IC₅₀ > 10,000 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

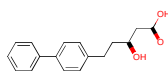
Taylor *et al.* (PubMed Identifier: 22017539)



15

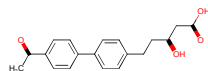
MMP-13 IC₅₀ = 1 nM
MMP-1 IC₅₀ > 22,000 nM

Holmes *et al.* (PubMed Identifier: 19703773)



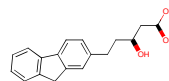
1

MMP-12 IC₅₀ = 520 nM
MMP-1 IC₅₀ > 98,000 nM



6c

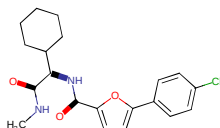
MMP-12 IC₅₀ = 62 nM
MMP-1 IC₅₀ > 98,000 nM



7a

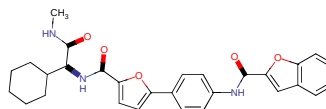
MMP-12 IC₅₀ = 1,150 nM
MMP-1 IC₅₀ > 98,000 nM

Heim-Riether *et al.* (PubMed Identifier: 19692239)



1

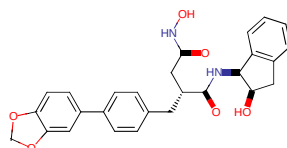
MMP-13 IC₅₀ = 430 nM
MMP-1 IC₅₀ > 21,500 nM



2

MMP-13 IC₅₀ = 620 nM
MMP-1 IC₅₀ > 18,600 nM

Shieh *et al.* (PubMed Identifier: 21370305)

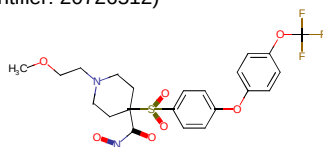


12

MMP-13 IC₅₀ = 7.3 nM
MMP-1 IC₅₀ = 7,560 nM

Manuscript 3

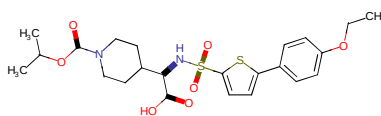
Becker *et al.* (PubMed Identifier: 20726512)



19v

MMP-13 IC₅₀ < 0.1 nM
MMP-1 IC₅₀ > 10,000 nM

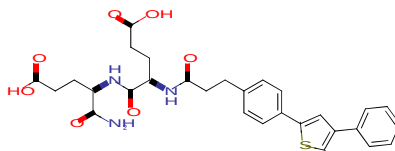
Monovich *et al.* (PubMed Identifier: 19422229)



24f

MMP-13 IC₅₀ = 0.5 nM
MMP-1 IC₅₀ > 10,000 nM

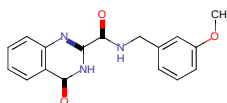
Devel *et al.* (PubMed Identifier: 22689580)



3

MMP-12 IC₅₀ = 1.9 nM
MMP-1 IC₅₀ > 10,000 nM

Nara *et al.* (PubMed Identifier: 25192810)

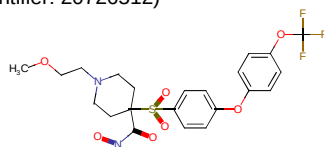


44

MMP-13 IC₅₀ = 12 nM
MMP-1 IC₅₀ > 10,000 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

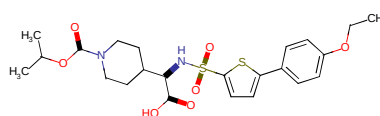
Becker *et al.* (PubMed Identifier: 20726512)



19v

MMP-13 IC_{50} < 0.1 nM
MMP-1 IC_{50} > 10,000 nM

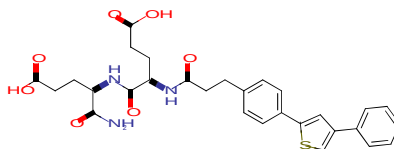
Monovich *et al.* (PubMed Identifier: 19422229)



24f

MMP-13 IC_{50} = 0.5 nM
MMP-1 IC_{50} > 10,000 nM

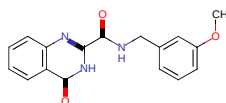
Devel *et al.* (PubMed Identifier: 22689580)



3

MMP-12 IC_{50} = 1.9 nM
MMP-1 IC_{50} > 10,000 nM

Nara *et al.* (PubMed Identifier: 25192810)

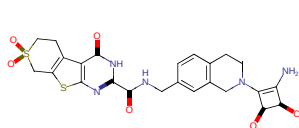


44

MMP-13 IC_{50} = 12 nM
MMP-1 IC_{50} > 10,000 nM

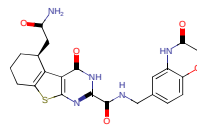
Manuscript 3

C) Pochetti *et al.* (PubMed Identifier: 19173605)



1

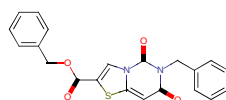
MMP-8 IC₅₀ = 57 nM
MMP-2 IC₅₀ > 2,500 nM



2

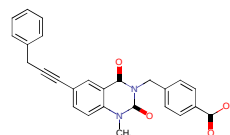
MMP-8 IC₅₀ = 7.4 nM
MMP-2 IC₅₀ > 10,000 nM

Johnson *et al.* (PubMed Identifier: 17623656)



1

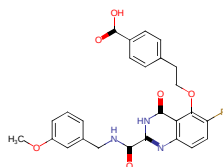
MMP-13 IC₅₀ = 30 nM
MMP-2 IC₅₀ > 100,000 nM



2

MMP-13 IC₅₀ = 0.67 nM
MMP-2 IC₅₀ > 30,000 nM

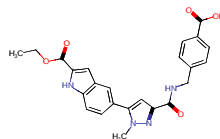
Nara *et al.* (PubMed Identifier: 25264600)



21k

MMP-13 IC₅₀ = 0.0039 nM
MMP-2 IC₅₀ = 5,300 nM

Taylor *et al.* (PubMed Identifier: 22017539)

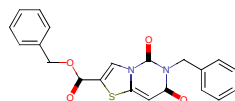


15

MMP-13 IC₅₀ = 1 nM
MMP-2 IC₅₀ = 18,000 nM

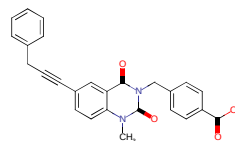
Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

D) Johnson *et al.* (PubMed Identifier: 17623656)



1

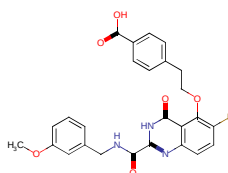
MMP-13 IC₅₀ = 30 nM
MMP-3 IC₅₀ > 100,000 nM



2

MMP-13 IC₅₀ = 0.67 nM
MMP-3 IC₅₀ > 30,000 nM

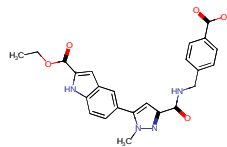
Nara *et al.* (PubMed Identifier: 25264600)



21k

MMP-13 IC₅₀ = 0.0039 nM
MMP-3 IC₅₀ = 4,000 nM

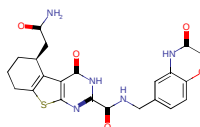
Taylor *et al.* (PubMed Identifier: 22017539)



15

MMP-13 IC₅₀ = 1 nM
MMP-3 IC₅₀ > 22,000 nM

Pochetti *et al.* (PubMed Identifier: 19173605)

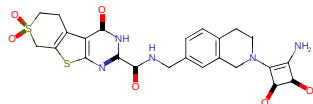


2

MMP-8 IC₅₀ = 7.4 nM
MMP-3 IC₅₀ > 2,500 nM

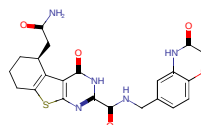
Manuscript 3

E) Pochetti *et al.* (PubMed Identifier: 19173605)



1

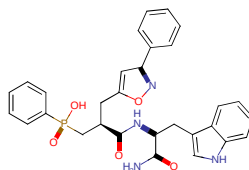
MMP-8 IC₅₀ = 57 nM
MMP-7 IC₅₀ > 10,000 nM



2

MMP-8 IC₅₀ = 7.4 nM
MMP-7 IC₅₀ > 10,000 nM

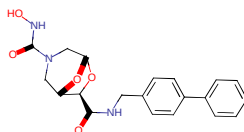
Tochowicz *et al.* (PubMed Identifier: 17599356)



AM-409

MMP-9 K_i = 13 nM
MMP-7 K_i > 2,500 nM

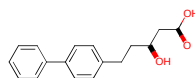
Mannino *et al.* (PubMed Identifier: 16899369)



20

MMP-12 IC₅₀ = 149,000 nM
MMP-7 IC₅₀ = 1,510,000 nM

Holmes *et al.* (PubMed Identifier: 19703773)

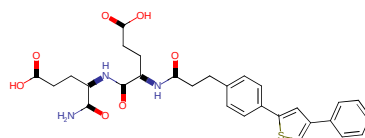


1

MMP-12 IC₅₀ = 520 nM
MMP-7 IC₅₀ > 98,000 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

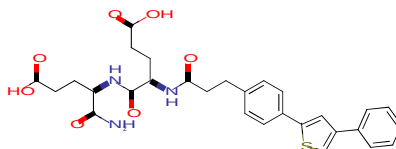
Devel *et al.* (PubMed Identifier: 20817735)



36

MMP-12 $K_i = 1$ nM
MMP-7 $K_i = 1,047$ nM

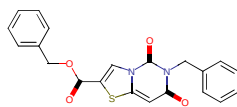
Devel *et al.* (PubMed Identifier: 22689580)



3

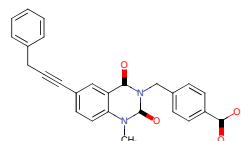
MMP-12 $K_i = 1.9$ nM
MMP-7 $K_i > 1,000$ nM

Johnson *et al.* (PubMed Identifier: 17623656)



1

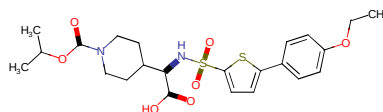
MMP-13 $IC_{50} = 30$ nM
MMP-7 $IC_{50} > 100,000$ nM



2

MMP-13 $IC_{50} = 0.67$ nM
MMP-7 $IC_{50} > 30,000$ nM

Monovich *et al.* (PubMed Identifier: 19422229)

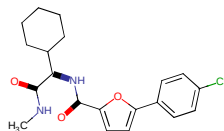


24f

MMP-13 $IC_{50} = 0.5$ nM
MMP-7 $IC_{50} = 3,025$ nM

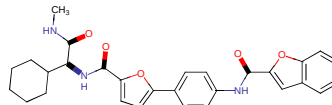
Manuscript 3

Heim-Riether *et al.* (PubMed Identifier: 19692239)



1

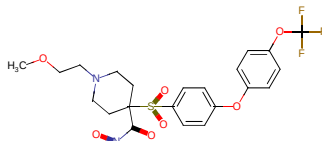
MMP-13 IC_{50} = 430 nM
MMP-7 IC_{50} > 21,500 nM



2

MMP-13 IC_{50} = 620 nM
MMP-7 IC_{50} > 18,600 nM

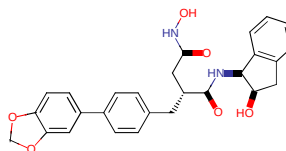
Becker *et al.* (PubMed Identifier: 20726512)



19v

MMP-13 IC_{50} < 0.1 nM
MMP-7 IC_{50} = 7,000 nM

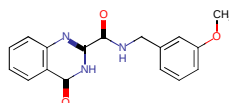
Shieh *et al.* (PubMed Identifier: 21370305)



12

MMP-13 K_i = 7.3 nM
MMP-7 K_i = 622 nM

Nara *et al.* (PubMed Identifier: 25192810)

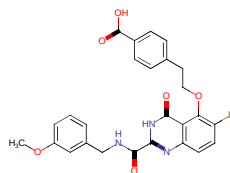


44

MMP-13 IC_{50} = 12 nM
MMP-7 IC_{50} > 10,000 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

Nara *et al.* (PubMed Identifier: 25264600)

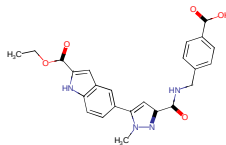


21k

MMP-13 IC_{50} = 0.0039 nM

MMP-7 IC_{50} > 10,000 nM

Taylor *et al.* (PubMed Identifier: 22017539)

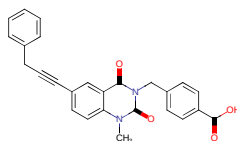


15

MMP-13 IC_{50} = 1 nM

MMP-7 IC_{50} > 22,000 nM

F) Johnson *et al.* (PubMed Identifier: 17623656)

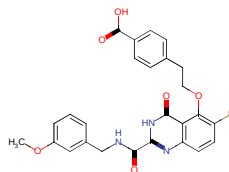


2

MMP-13 IC_{50} = 0.67 nM

MMP-8 IC_{50} > 100,000 nM

Nara *et al.* (PubMed Identifier: 25264600)



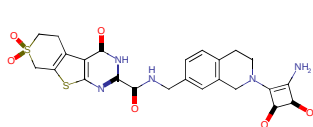
21k

MMP-13 IC_{50} = 0.0039 nM

MMP-8 IC_{50} = 720 nM

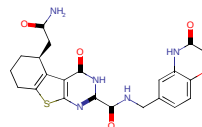
Manuscript 3

G) Pochetti *et al.* (PubMed Identifier: 19173605)



1

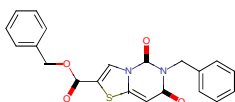
MMP-8 IC_{50} = 57 nM
MMP-9 IC_{50} > 10,000 nM



2

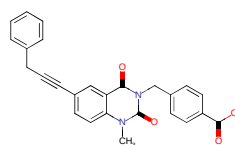
MMP-8 IC_{50} = 7.4 nM
MMP-9 IC_{50} > 10,000 nM

Johnson *et al.* (PubMed Identifier: 17623656)



1

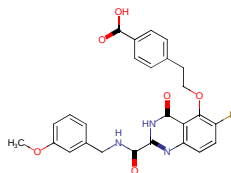
MMP-13 IC_{50} = 30 nM
MMP-9 IC_{50} > 100,000 nM



2

MMP-13 IC_{50} = 0.67 nM
MMP-9 IC_{50} > 100,000 nM

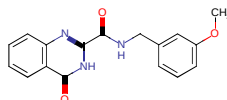
Nara *et al.* (PubMed Identifier: 25264600)



21k

MMP-13 IC_{50} = 0.0039 nM
MMP-9 IC_{50} = 10,000 nM

Nara *et al.* (PubMed Identifier: 25192810)

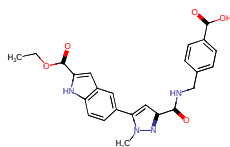


44

MMP-13 IC_{50} = 12 nM
MMP-9 IC_{50} > 10,000 nM

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

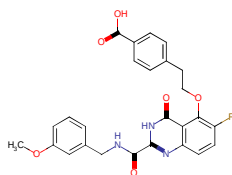
Taylor *et al.* (PubMed Identifier: 22017539)



15

MMP-13 IC₅₀ = 1 nM
MMP-9 IC₅₀ = 8,900 nM

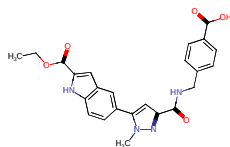
H) Nara *et al.* (PubMed Identifier: 25264600)



21k

MMP-13 IC₅₀ = 0.0039 nM
MMP-10 IC₅₀ = 160 nM

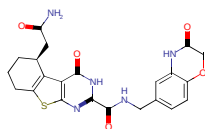
Taylor *et al.* (PubMed Identifier: 22017539)



15

MMP-13 IC₅₀ = 1 nM
MMP-10 IC₅₀ = 16,000 nM

I) Pochetti *et al.* (PubMed Identifier: 19173605)

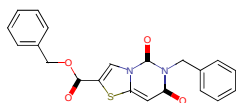


2

MMP-8 IC₅₀ = 7.4 nM
MMP-12 IC₅₀ > 10,000 nM

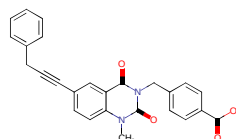
Manuscript 3

Johnson *et al.* (PubMed Identifier: 17623656)



1

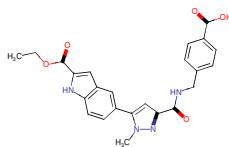
MMP-13 IC₅₀ = 30 nM
MMP-12 IC₅₀ > 100,000 nM



2

MMP-13 IC₅₀ = 0.67 nM
MMP-12 IC₅₀ > 100,000 nM

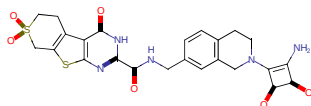
Taylor *et al.* (PubMed Identifier: 22017539)



15

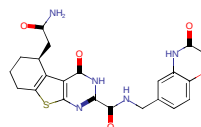
MMP-13 IC₅₀ = 1 nM
MMP-12 IC₅₀ > 22,000 nM

J) Pochetti *et al.* 19173605



1

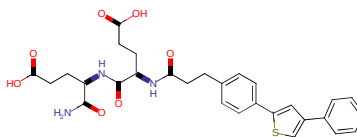
MMP-8 IC₅₀ = 57 nM
MMP-14 IC₅₀ > 10,000 nM



2

MMP-8 IC₅₀ = 7.4 nM
MMP-14 IC₅₀ > 10,000 nM

Devel *et al.* (PubMed Identifier: 20817735)

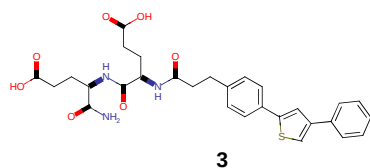


36

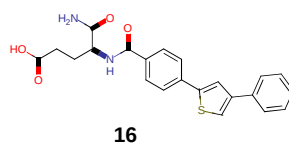
MMP-12 K_i = 1 nM
MMP-14 K_i = 1,568 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

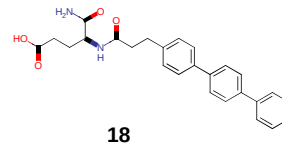
Devel *et al.* (PubMed Identifier: 22689580)



MMP-12 $K_i = 1.9$ nM
MMP-14 $K_i = 3,010$ nM

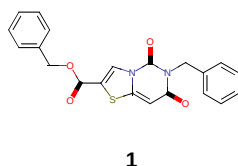


MMP-12 $K_i = 2.5$ nM
MMP-14 $K_i = 968$ nM

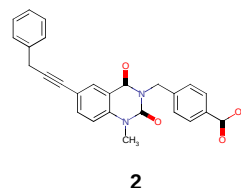


MMP-13 $K_i = 18$ nM
MMP-14 $K_i = 1,054$ nM

Johnson *et al.* (PubMed Identifier: 17623656)

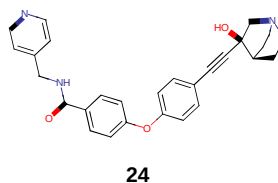


MMP-13 $IC_{50} = 30$ nM
MMP-14 $IC_{50} > 100,000$ nM



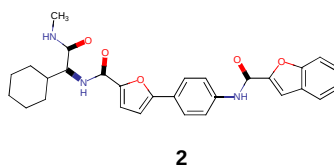
MMP-13 $IC_{50} = 0.67$ nM
MMP-14 $IC_{50} > 30,000$ nM

Savi *et al.* (PubMed Identifier: 21669521)



MMP-13 $IC_{50} = 80$ nM
MMP-14 $IC_{50} > 10,800$ nM

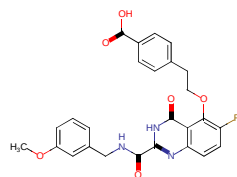
Heim-Riether *et al.* (PubMed Identifier: 19692239)



MMP-13 $IC_{50} = 620$ nM
MMP-14 $IC_{50} > 62,000$ nM

Manuscript 3

Nara *et al.* (PubMed Identifier: 25264600)

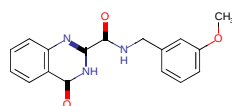


21k

MMP-13 IC_{50} = 0.0039 nM

MMP-14 IC_{50} > 10,000 nM

Nara *et al.* (PubMed Identifier: 25192810)

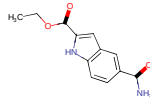


44

MMP-13 IC_{50} = 12 nM

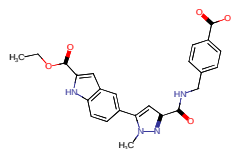
MMP-14 IC_{50} > 10,000 nM

Taylor *et al.* (PubMed Identifier: 22017539)



1

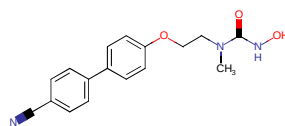
MMP-13 IC_{50} = 390 nM
MMP-14 IC_{50} > 500,000 nM



15

MMP-13 IC_{50} = 1 nM
MMP-14 IC_{50} = 8,300 nM

K) Campestre *et al.* (PubMed Identifier: 16242329)

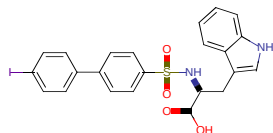


2b

MMP-2 IC_{50} = 52,000 nM
MMP-8 IC_{50} = 1,200,000 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

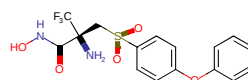
Tochowicz *et al.* (PubMed Identifier: 17599356)



An-1

MMP-2 IC₅₀ = 9.3 nM

MMP-9 IC₅₀ = 201 nM

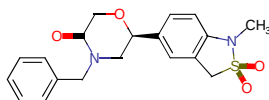


MS-560

MMP-2 IC₅₀ = 0.01 nM

MMP-9 IC₅₀ = 1 nM

L) Morales *et al.* (PubMed Identifier: 15289103)

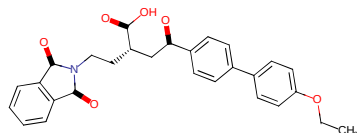


CP-271485

MMP-9 IC₅₀ = 5,100 nM

MMP-12 IC₅₀ > 100,000 nM

M) Morales *et al.* (PubMed Identifier: 15289103)



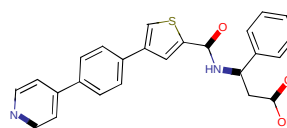
PD-0359601

MMP-12 IC₅₀ = 1.7 nM

MMP-2 IC₅₀ = 6,600 nM

MMP-3 IC₅₀ = 3.2 nM

MMP-8 IC₅₀ = 160 nM



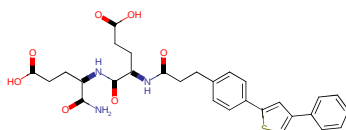
PF-00356231

MMP-12 IC₅₀ = 14 nM

MMP-2 IC₅₀ > 100,000 nM

MMP-3 IC₅₀ = 390 nM

Devel *et al.* (PubMed Identifier: 22689580)



3

MMP-12 K_i = 1.9 nM

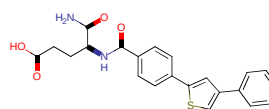
MMP-8 K_i = 410 nM

MMP-2 K_i = 1,060 nM

MMP-9 K_i = 9,890 nM

MMP-3 K_i = 3,880 nM

MMP-13 K_i = 684 nM



16

MMP-12 K_i = 2.5 nM

MMP-8 K_i = 671 nM

MMP-2 K_i = 127 nM

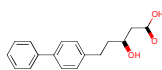
MMP-9 K_i = 2,232 nM

MMP-3 K_i = 5,819 nM

MMP-13 K_i = 501 nM

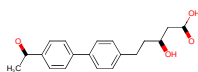
Manuscript 3

Holmes *et al.* (PubMed Identifier: 19703773)



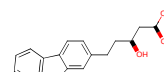
1

MMP-12 IC₅₀ = 520 nM MMP-7 IC₅₀ > 98,000 nM
MMP-1 IC₅₀ > 98,000 nM MMP-9 IC₅₀ = 16,500 nM
MMP-2 IC₅₀ = 4,520 nM MMP-13 IC₅₀ = 12,000 nM
MMP-3 IC₅₀ > 98,000 nM MMP-14 IC₅₀ = 43,500 nM



6c

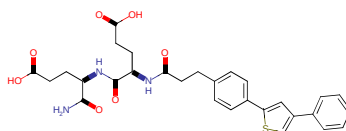
MMP-12 IC₅₀ = 62 nM
MMP-1 IC₅₀ > 98,000 nM
MMP-13 IC₅₀ = 970 nM



7a

MMP-12 IC₅₀ = 1,150 nM
MMP-1 IC₅₀ > 98,000 nM
MMP-13 IC₅₀ = 26,100 nM

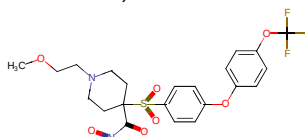
Devel *et al.* (PubMed Identifier: 20817735)



36

MMP-12 K_i = 1.92 nM MMP-3 K_i = 3,880 nM MMP-9 K_i > 10,000 nM
MMP-1 K_i > 10,000 nM MMP-7 K_i = 2,010 nM MMP-13 K_i = 684 nM
MMP-2 K_i = 1,060 nM MMP-8 K_i = 410 nM MMP-14 K_i = 3,010 nM

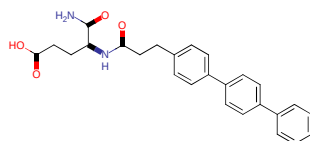
N) Becker *et al.* (PubMed Identifier: 20726512)



19v

MMP-13 IC₅₀ < 0.1 nM
MMP-3 IC₅₀ = 28.7 nM
MMP-8 IC₅₀ = 1.7 nM

Devel *et al.* (PubMed Identifier: 22689580)

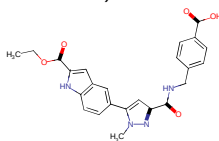


18

MMP-13 K_i = 18 nM
MMP-8 K_i = 205 nM

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

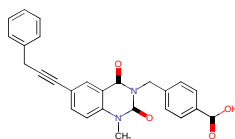
Taylor *et al.* (PubMed Identifier: 22017539)



15

MMP-13 IC ₅₀ = 1 nM	MMP-3 IC ₅₀ > 22,000 nM	MMP-9 IC ₅₀ = 8,900 nM	
MMP-1 IC ₅₀ > 22,000 nM	MMP-7 IC ₅₀ > 22,000 nM	MMP-10 IC ₅₀ = 16,000 nM	MMP-14 IC ₅₀ = 8,300 nM
MMP-2 IC ₅₀ = 18,000 nM	MMP-8 IC ₅₀ > 22,000 nM	MMP-12 IC ₅₀ > 22,000 nM	

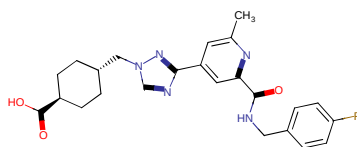
Johnson *et al.* (PubMed Identifier: 17623656)



2

MMP-13 IC ₅₀ = 0.67 nM	MMP-3 IC ₅₀ > 30,000 nM	MMP-9 IC ₅₀ > 100,000 nM	
MMP-1 IC ₅₀ > 30,000 nM	MMP-7 IC ₅₀ > 30,000 nM	MMP-12 IC ₅₀ > 100,000 nM	MMP-17 IC ₅₀ > 30,000 nM
MMP-2 IC ₅₀ > 30,000 nM	MMP-8 IC ₅₀ > 100,000 nM	MMP-14 IC ₅₀ > 30,000 nM	

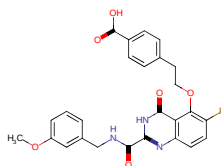
Schnute *et al.* (PubMed Identifier: 20005097)



29b

MMP-13 K _i = 4.4 nM	MMP-7 K _i > 25,000 nM	MMP-14 K _i > 25,000 nM	
MMP-1 K _i > 25,000 nM	MMP-8 K _i > 25,000 nM	MMP-15 K _i > 25,000 nM	MMP-25 K _i > 25,000 nM
MMP-2 K _i > 25,000 nM	MMP-9 K _i > 25,000 nM	MMP-16 K _i > 25,000 nM	MMP-26 K _i > 25,000 nM
MMP-3 K _i > 25,000 nM	MMP-12 K _i > 25,000 nM	MMP-24 K _i > 25,000 nM	

Nara *et al.* (PubMed Identifier: 25264600)

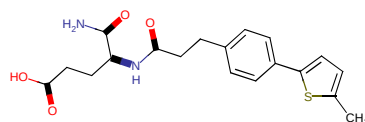


21k

MMP-13 IC ₅₀ = 0.0039 nM	MMP-3 IC ₅₀ = 4,000 nM	MMP-9 IC ₅₀ > 10,000 nM	
MMP-1 IC ₅₀ > 10,000 nM	MMP-7 IC ₅₀ > 10,000 nM	MMP-10 IC ₅₀ = 160 nM	
MMP-2 IC ₅₀ = 5,300 nM	MMP-8 IC ₅₀ = 720 nM	MMP-14 IC ₅₀ > 10,000 nM	

Manuscript 3

O) Devel *et al.* (PubMed Identifier: 22689580)



17

MMP-8 K_i = 5.3 nM

MMP-2 K_i = 57 nM

MMP-3 K_i = 2,164 nM

MMP-13 K_i = 338 nM

Figure S2. Structures and activities of MMP inhibitors.^{52,58,60,61,67,71,75,91–103} Panel A: peptidomimetic inhibitors; Panels from B to J: inhibitors sparing MMP-1, -2, -3, -7, -8, -9, -10, -12 and -14, respectively. Panels from K to O: Inhibitors selective towards MMP-2, -9, -12, -13 and -8, respectively. The highest MMP activity of the inhibitor is shown at the top (except in panel A) and the rest of MMP activities are sorted in alphanumeric order, based on the name of the corresponding MMP. Marvin v16.10¹⁰⁴ was used to draw the 2D structures of the inhibitors. The bioactivity data shown in panel A was obtained from Reaxys.¹⁰⁵

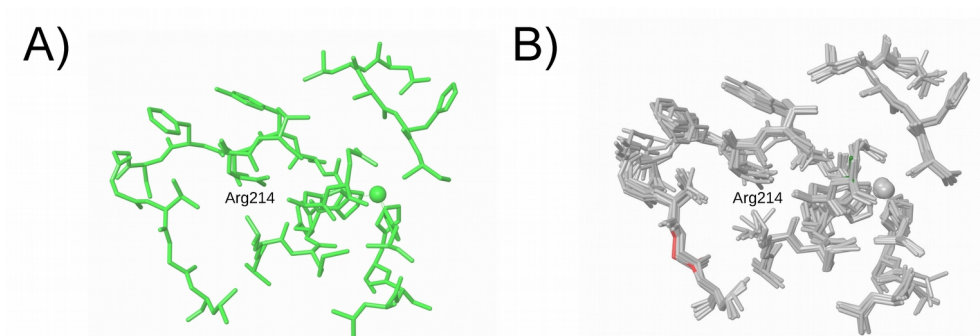


Figure S3. Ω -loop conformations of MMP-1. Panel A shows a representative structure of the Ω -loop conformation of MMP-1 (see Table S12). Panel B shows the superposition of the subunits with available EDMs that present this conformation (see Table S12). Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. Non-natural mutations are in green. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. This figure was obtained with Maestro¹⁰⁵ v11.

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

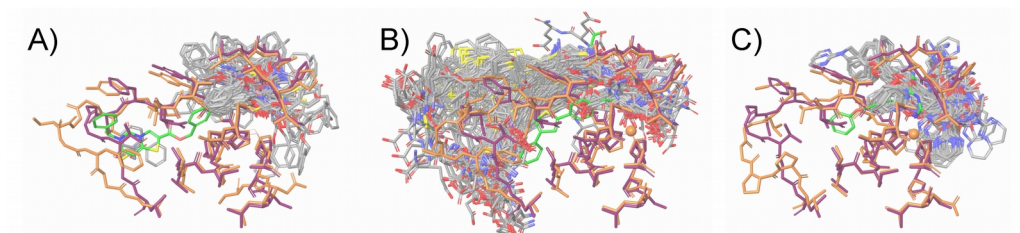


Figure S4. Docking of **1¹⁰⁶** (panel A), **36¹⁰³** (panel B) and **RO-206-0222⁵⁸** (panel C) on MMP-1 compared with their respective experimental complexes at MMP-13 [PDB⁹⁸ ID and subunit: 2OW9 (A)^{84,85}], MMP-12 [(PDB¹⁰³ ID and subunit: 3LIK (A)^{84,85}] and MMP-9 [PDB⁵⁸ ID and subunit: 2OVX (A)^{84,85}]. The MMP-1 structure used for docking [PDB⁹⁸ ID and subunit: 2JOT (A)^{84,85}] is shown in purple, the docked poses for **1¹⁰⁷**, **36¹⁰³** and **RO-206-0222⁵⁸** are shown in gray and the ligands and MMP structures from the experimental complexes are shown in green and orange, respectively. All panels are in the same orientation to facilitate comparison. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro⁹⁸ v11.

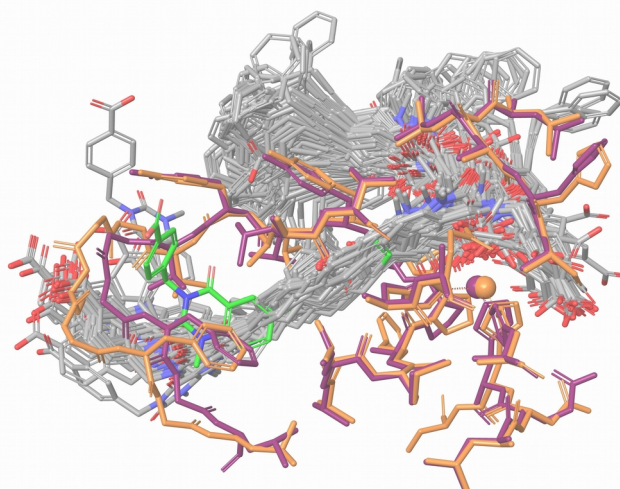


Figure S5. Docking of **2¹⁰⁶** on MMP-2 compared with its experimental complex at MMP-13 [PDB¹⁰³ ID and subunit: 2OZR (A)^{84,85}]. The MMP-2 structure used for docking [PDB¹⁰³ ID and subunit: 3AYU (A)^{84,85}] is shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro⁸⁷ v11.

Manuscript 3

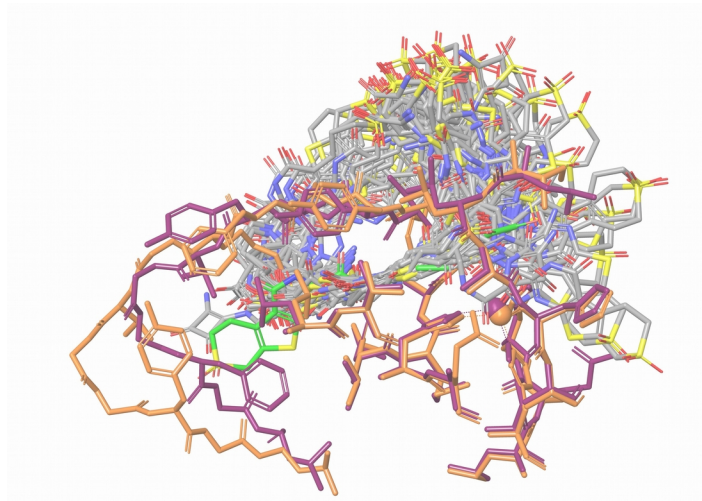


Figure S6. Docking of **1¹⁰⁶** on MMP-2 compared with its experimental complex at MMP-8 [PDB⁹³ ID and subunit: 3DPE (A)^{84,85}]. The MMP-2 structure used for docking [PDB⁹³ ID and subunit: 3AYU (A)^{84,85}] is shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. For clarity reasons, only the docked poses in which the ring system of the ligand occupies the S1' pocket have been represented. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro⁸⁷ v11.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

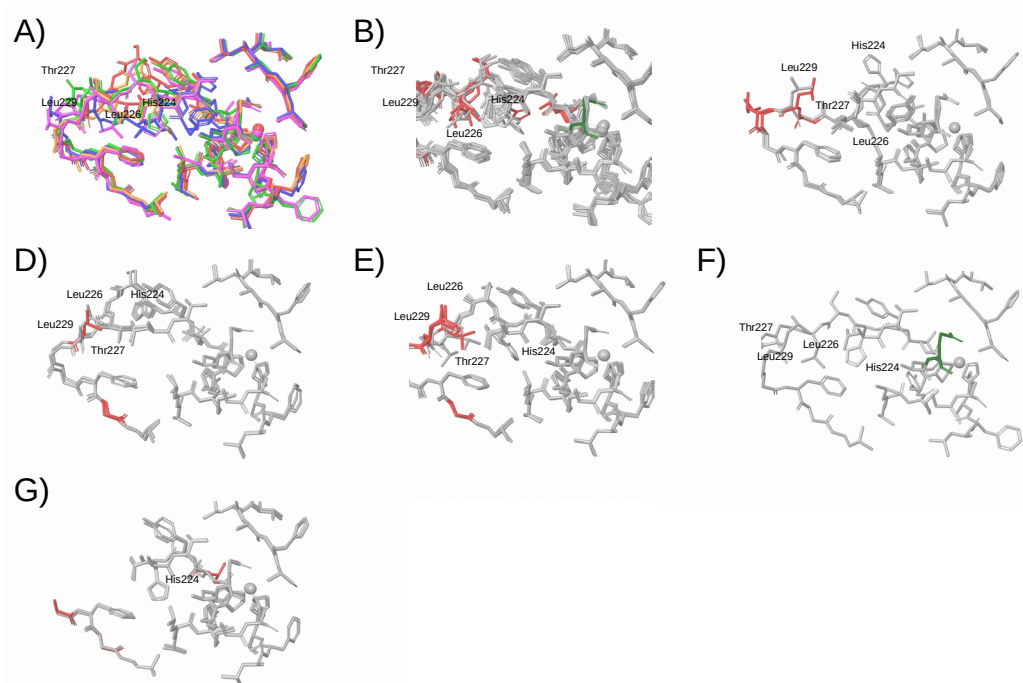


Figure S7. Ω -loop conformations of MMP-3. Panels B-F show the superposition of the subunits with available EDMs that present the conformations from A to E (see Table S14). Panel A shows a superposition of the representative structures with different Ω -loop conformations in MMP-3, the conformations from A to E in light green, blue, red, orange and pink, respectively (see Table S14). Panel G shows a superposition of the subunits that could not be classified due to the lack of coordinates for some residues of the Ω -loop in the corresponding PDB file. Residues for which the electron density has been considered insufficient to validate their conformation are in red. Non-natural mutations are shown in green. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

Manuscript 3

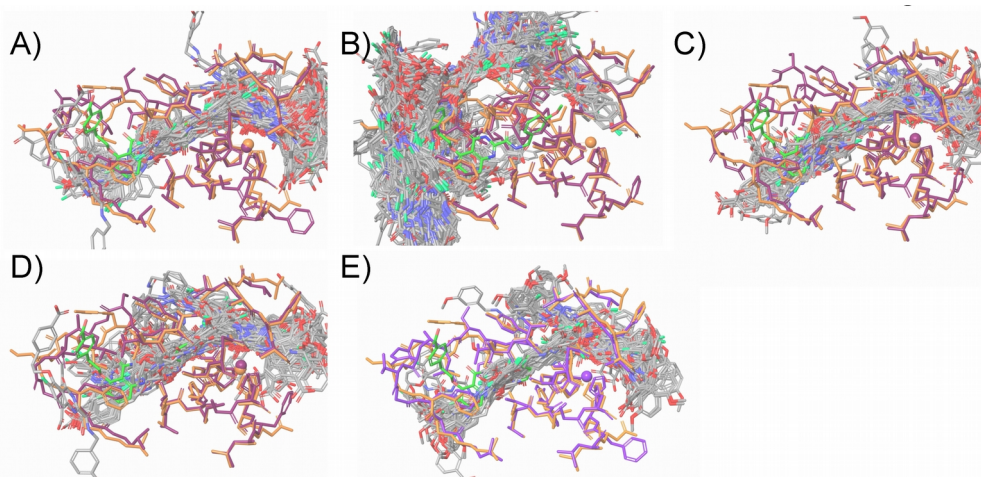


Figure S8. Docking of **21k¹⁰⁶** on MMP-3 compared with its experimental complex at MMP-13 [PDB⁶⁰ ID and subunit: 3WV1 (B)^{84,85}]. Panels from A to E show the dockings of **21k⁶⁰** on MMP-3 structures presenting the Ω -loop conformations from A to E (see Figure S7 and Table S14). The MMP-3 structures used for docking [PDB⁶⁰ IDs and subunits 1HFS (A),^{84,85} 4G9L (A),⁸⁸ 2D1O (A),¹⁰⁹ 4DPE (B),¹¹⁰ and 1C3I (B)¹⁰⁹ for panels from A to E, respectively] are shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro¹¹¹ v11.

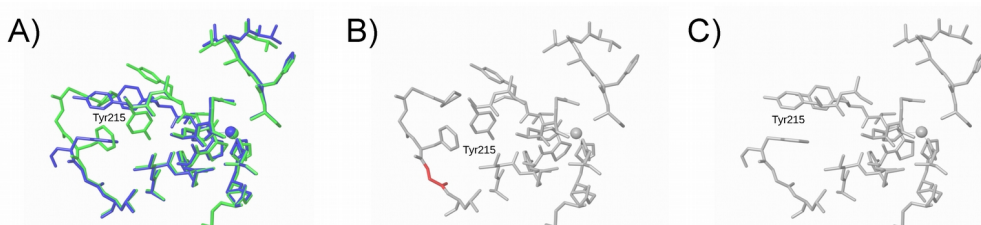


Figure S9. Ω -loop conformations of MMP-7. Panels B and C show the subunits with available EDMs that present the conformations A and B (see Table S15). Panel A shows a superposition of the A and B conformations in light green and blue, respectively. Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

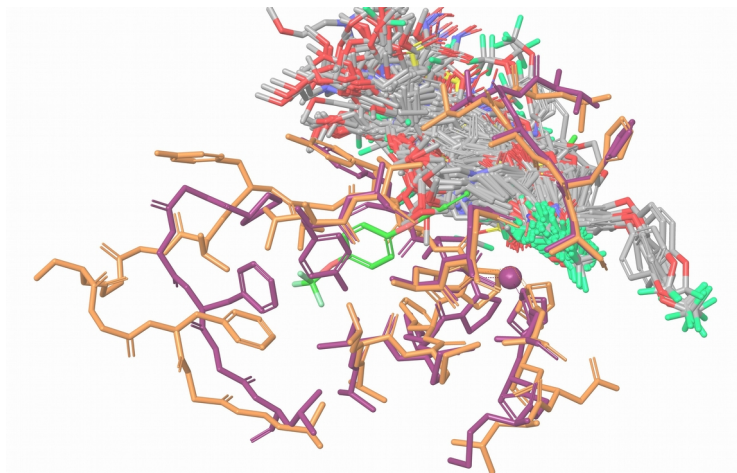


Figure S10. Docking of **19v**¹⁰⁶ on MMP-7 compared with its experimental complex at MMP-13 [PDB⁹⁶ ID and subunit: 3KRY (A)^{84,85}]. The MMP-7 structure used for docking [PDB⁹⁶ ID and subunit: 2Y6C (A)^{84,85}] is shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro⁴¹ v11.

Manuscript 3

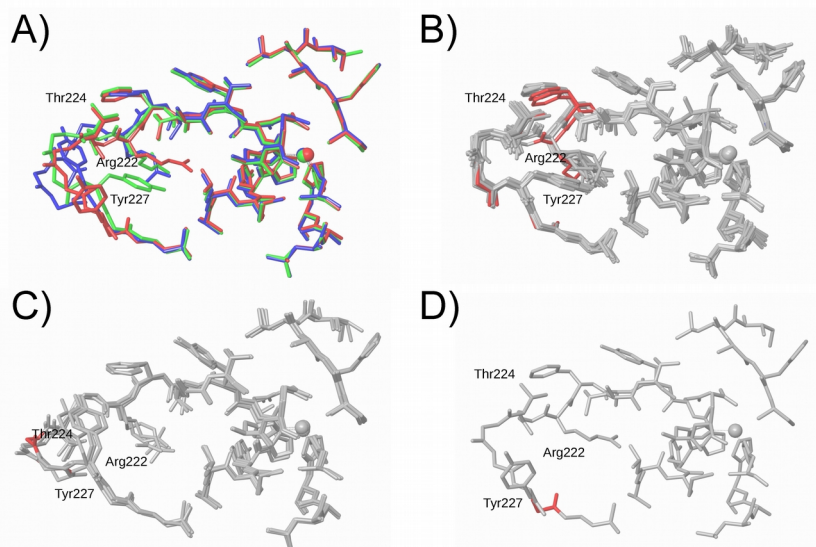


Figure S11. Ω -loop conformations of MMP-8. Panels from B to D show the subunits with available EDMs that present the conformations from A to C (see Table S16). Panel A shows a superposition of the representative structures with different Ω -loop conformations in MMP-8 (see Table S16) with conformations A, B and C in light green, blue and red respectively. Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

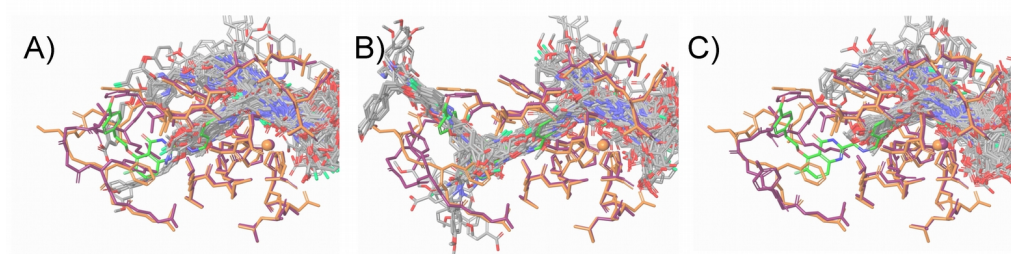


Figure S12. Docking of **21k**¹⁰⁶ on MMP-8 compared with its experimental complex at MMP-13 [PDB⁶⁰ ID and subunit: 3WV1 (B)^{84,85}]. Panels from A to C show the dockings of **21k**⁶⁰ on MMP-8 structures presenting the Ω -loop conformations from A to C (see Table S16). The MMP-8 structures used for docking [PDB⁶⁰ IDs and subunits 1173 (A),^{84,85} 3DNG (A)⁸⁹ and 5H8X (A)⁹³ for panels A, B and C, respectively] are shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro¹¹² v11.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

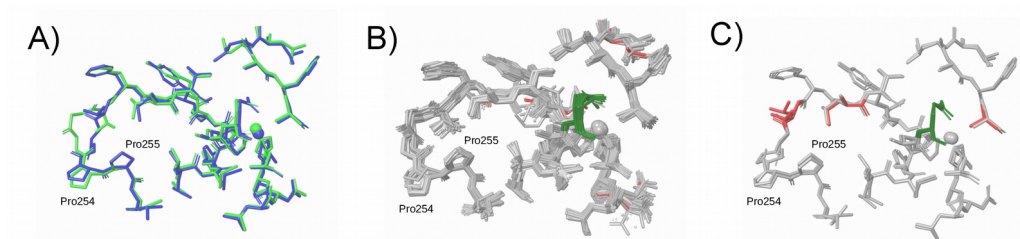


Figure S13. Ω -loop conformations of MMP-9. Panels B and C show the subunits with available EDMs that present the conformations A and B (see Table S17). Panel A shows a superposition of the representative structures with different Ω -loop conformations in MMP-9 (see Table S17) with the A and B conformations in light green and blue, respectively. Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

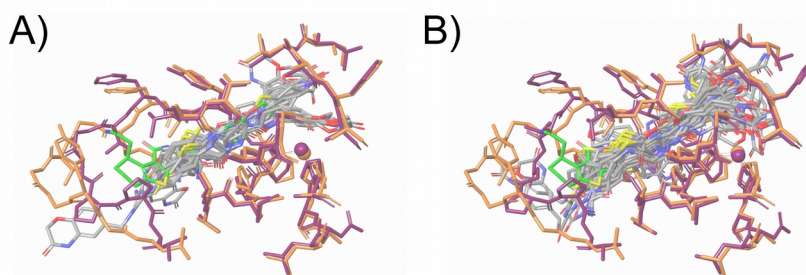


Figure S14. Docking of 2^{106} on MMP-9 compared with its experimental complex at MMP-8 [PDB⁹³ ID and subunit: 3DNG (A)^{84,85}]. Panels A and B show the dockings of 2^{93} on MMP-9 structures presenting the Ω -loop conformations A and B (see Table S17). The MMP-9 structures used for docking [PDB⁹³ IDs and subunits 4WZV (A)^{84,85} and 4H2E (A)⁹⁹, respectively] are shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. All panels are in the same orientation to facilitate comparison. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro¹¹³ v11.

Manuscript 3

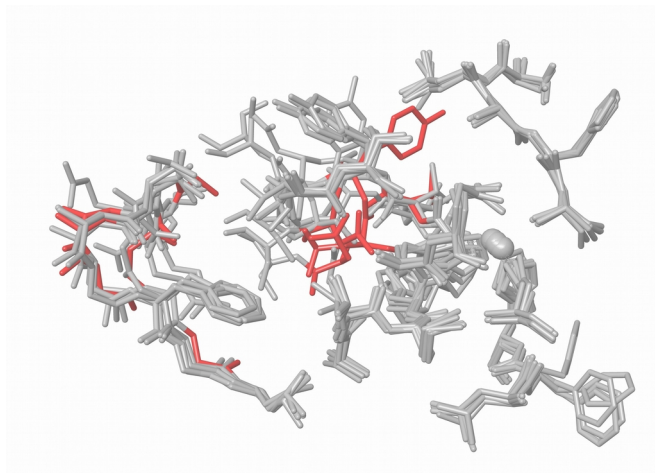


Figure S15. Superposition of the subunits of MMP-10 that could not be classified due to missing residues (see Table S18). Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. This figure was obtained with Maestro¹⁰⁶ v11.

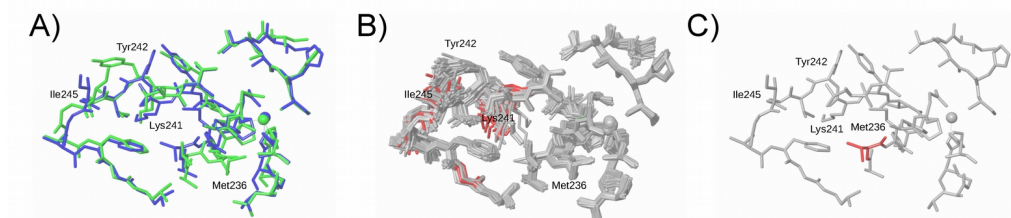


Figure S16. Ω -loop conformations of MMP-12. Panels B and C show the subunits with available EDMs that present the conformations A and B (see Table S19). Panel A shows a superposition of the representative structures with different Ω -loop conformations in MMP-12 (see Table S19) with conformations A and B in light green and blue, respectively. Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. Non-natural mutations are in green. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

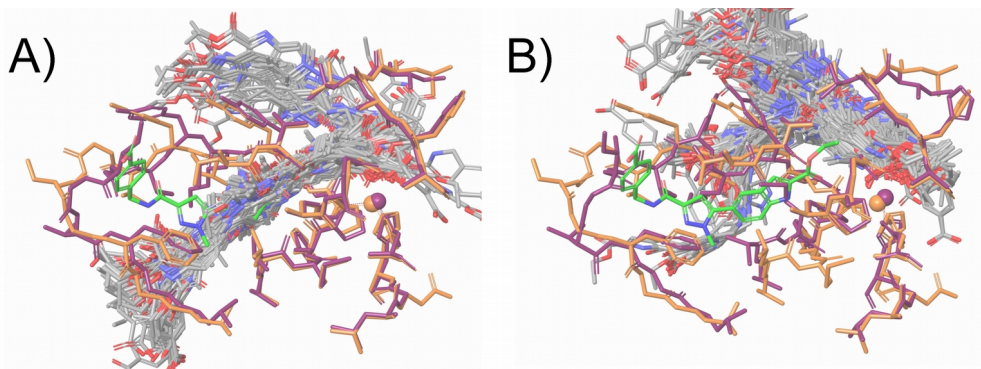


Figure S17. Docking of **15**¹⁰⁶ on MMP-12 compared with its experimental complex at MMP-13 [PDB⁶⁷ ID and subunit: 5BPA (B)^{84,85}]. Panels A and B show the dockings of **15**⁶⁷ on MMP-12 structures presenting the Ω -loop conformations A and B (see Table S19). The MMP-12 structures used for docking [PDB⁶⁷ IDs and subunits 1JIZ (A)^{84,85} and 3BA0 (A),¹¹⁴ respectively] are shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. All panels are in the same orientation to facilitate comparison. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro¹¹⁵ v11.

Manuscript 3

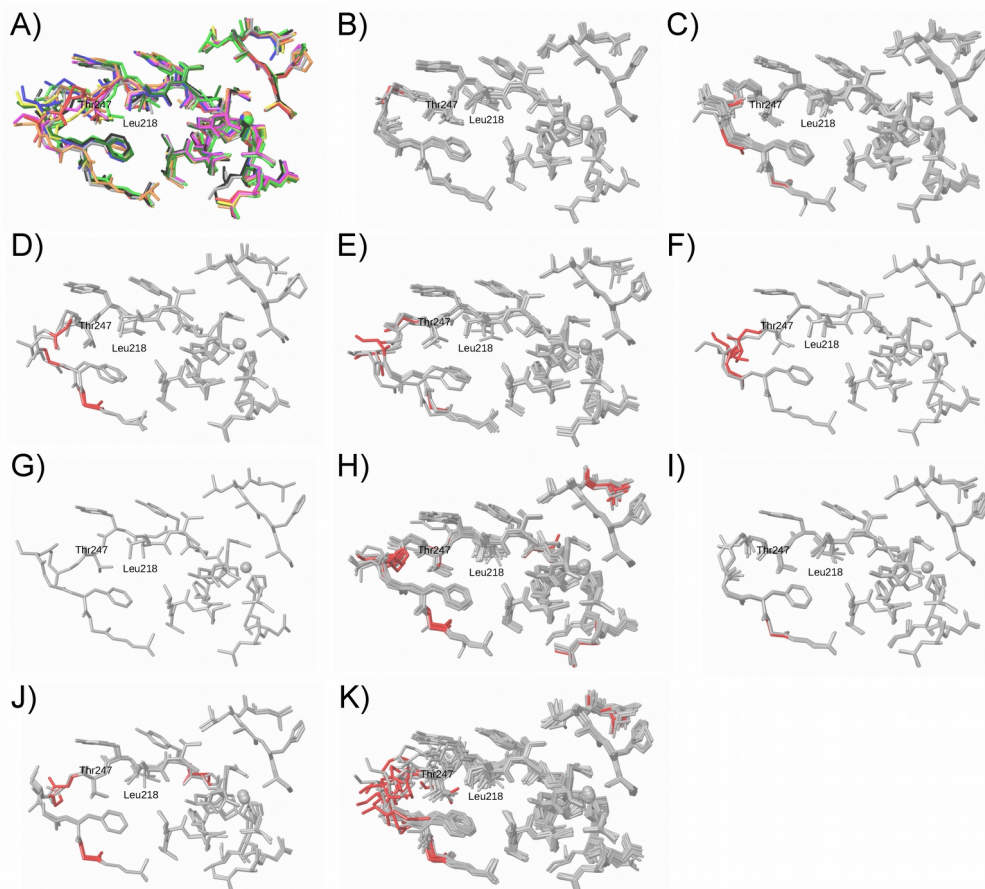


Figure S18. Ω -loop conformations of MMP-13. Panels from B to J show the subunits with available EDMs that present the conformations from A to I (see Table S20), in light green, blue, red, orange, pink, yellow, dark green, grey and black, respectively. Panel A shows a superposition of the representative structures with different Ω -loop conformations in MMP-13 (see Table S20). Panel K shows a superposition of the subunits that could not be classified due to missing residues. Residues for which the electron density has been considered insufficient to validate their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. All panels are in the same orientation to facilitate comparison. This figure was obtained with Maestro¹⁰⁶ v11.

Understanding the variability of the S1' pocket to improve MMPi selectivity profiles

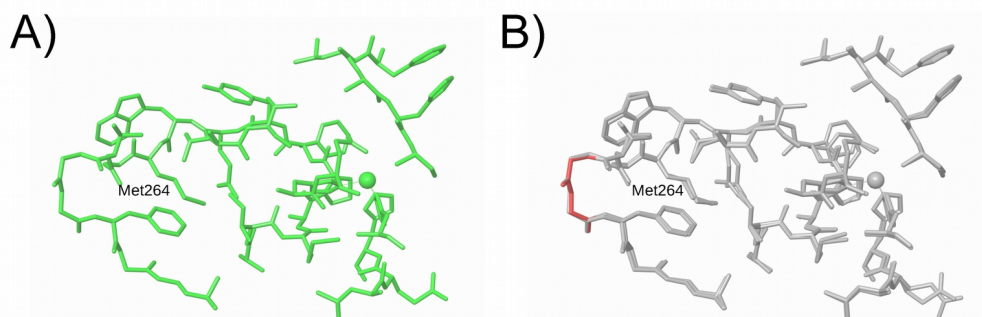


Figure S19. Ω -loop conformations of MMP-14. Panel A shows a representative structure of the Ω -loop conformation of MMP-14 (see Table S21). Panel B shows the superposition of the subunits with available EDMs that present this conformation (see Table S21). Residues for which the electron density has been considered insufficient to determine their conformation are shown in red. For clarity reasons, in all panels the side-chains facing out of the S1' pocket were not represented. This figure was obtained with Maestro¹⁰⁶ v11.

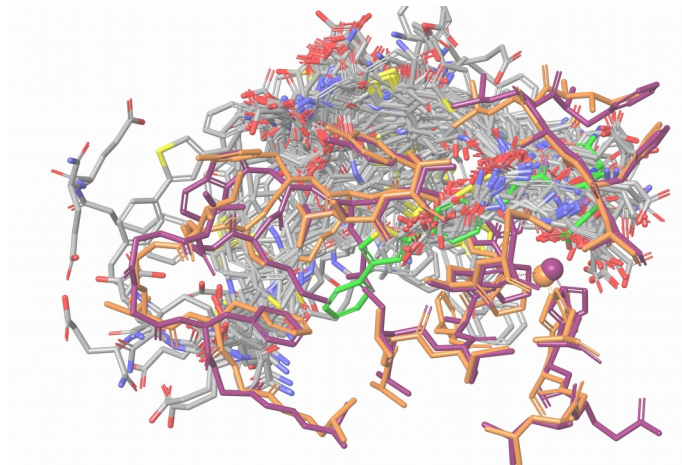


Figure S20. Docking of 3¹⁰⁶ on MMP-14 compared with its experimental complex at MMP-12 [PDB⁹¹ ID and subunit: 3TS4 (A)^{84,85}]. The MMP-14 structure used for docking [PDB⁹¹ ID and subunit: 3MA2 (A)^{84,85}] is shown in purple, the docked poses are shown in gray and the ligand and MMP structure from the experimental complex are shown in green and orange, respectively. For clarity reasons, only the docked poses in which the ring system of the ligand occupies the S1' pocket have been represented. Docking was performed with Glide¹⁰⁸ SP. This figure was obtained with Maestro¹¹⁶ v11.

References

- (1) Ala-aho, R.; Kähäri, V.-M. Collagenases in Cancer. *Biochimie* **2005**, *87* (3–4), 273–286.
- (2) Shih, C.-L. M.; Ajuwon, K. M. Inhibition of MMP-13 Prevents Diet-Induced Obesity in Mice and Suppresses Adipogenesis in 3T3-L1 Preadipocytes. *Mol. Biol. Rep.* **2015**, *42* (7), 1225–1232.
- (3) Li, H.; Wang, D.; Yuan, Y.; Min, J. New Insights on the MMP-13 Regulatory Network in the Pathogenesis of Early Osteoarthritis. *Arthritis Res. Ther.* **2017**, *19* (1), 248.
- (4) Hu, J.; Van den Steen, P. E.; Sang, Q.-X. a; Opdenakker, G. Matrix Metalloproteinase Inhibitors as Therapy for Inflammatory and Vascular Diseases. *Nat. Rev. Drug Discov.* **2007**, *6* (6), 480–498.
- (5) Rath, T.; Roderfeld, M.; Graf, J.; Wagner, S.; Vehr, A.-K.; Dietrich, C.; Geier, A.; Roeb, E. Enhanced Expression of MMP-7 and MMP-13 in Inflammatory Bowel Disease: A Precancerous Potential? *Inflamm. Bowel Dis.* **2006**, *12* (11), 1025–1035.
- (6) Solovyeva, N. I.; Timoshenko, O. S.; Gureeva, T. A.; Kugaevskaya, E. V. Matrix Metalloproteinases and Their Endogenous Regulators in Squamous Cervical Carcinoma (Review of the Own Data). *Biomeditsinskaya Khimiya* **2015**, *61* (6), 694–704.
- (7) Rabkin, S. W. The Role Matrix Metalloproteinases in the Production of Aortic Aneurysm. *Prog. Mol. Biol. Transl. Sci.* **2017**, *147*, 239–265.
- (8) Arakaki, P. A.; Marques, M. R.; Santos, M. C. L. G. MMP-1 Polymorphism and Its Relationship to Pathological Processes. *J. Biosci.* **2009**, *34* (2), 313–320.
- (9) Radosinska, J.; Barancik, M.; Vrbjar, N. Heart Failure and Role of Circulating MMP-2 and MMP-9. *Panminerva Med.* **2017**, *59* (3), 241–253.
- (10) Overall, C. M.; López-Otín, C. Strategies for MMP Inhibition in Cancer: Innovations for the Post-Trial Era. *Nat. Rev. Cancer* **2002**, *2* (9), 657–672.
- (11) Kurzepa, J.; M, A.; Czechowska, G.; Kurzepa, J.; Celiński, K.; Kazmierak, W.; Slstrokoma, M. Role of MMP-2 and MMP-9 and Their Natural Inhibitors in Liver Fibrosis, Chronic Pancreatitis and Non-Specific Inflammatory Bowel Diseases. *Hepatobiliary Pancreat. Dis. Int.* **2014**, *13* (6), 570–579.
- (12) Siloși, I.; Boldeanu, M. V.; Mogoantă, S. Ș.; Ghiluși, M.; Cojocaru, M.; Biciușcă, V.; Cojocaru, I. M.; Avrămescu, C. S.; Gheonea, D. I.; Siloși, C. A.; et al. Matrix Metalloproteinases (MMP-3 and MMP-9) Implication in the Pathogenesis of Inflammatory Bowel Disease (IBD). *Rom. J. Morphol. Embryol.* **2014**, *55* (4), 1317–1324.
- (13) Romero, A. M.; Mastromatteo-Alberga, P.; Escalona, L.; Correnti, M. [MMP-3 and MMP-8 Levels in Patients with Chronic Periodontitis before and after Nonsurgical Periodontal Therapy]. *Invest. Clin.* **2013**, *54* (2), 138–148.
- (14) Beton, O.; Arslan, S.; Acar, B.; Ozbilum, N.; Berkan, O. Association between MMP-3 and MMP-9 Polymorphisms and Coronary Artery Disease. *Biomed. reports* **2016**, *5* (6), 709–714.
- (15) Huang, H. Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances. *Sensors (Basel)*. **2018**, *18* (10), 3249.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

- (16) Gharib, S. A.; Manicone, A. M.; Parks, W. C. Matrix Metalloproteinases in Emphysema. *Matrix Biol.* **2018**, *73*, 34–51.
- (17) Rodriguez, J. A.; Orbe, J.; Martinez de Lizarrondo, S.; Calvayrac, O.; Rodriguez, C.; Martinez-Gonzalez, J.; Paramo, J. A. Metalloproteinases and Atherothrombosis: MMP-10 Mediates Vascular Remodeling Promoted by Inflammatory Stimuli. *Front. Biosci.* **2008**, *13*, 2916–2921.
- (18) Chelluboina, B.; Nalamolu, K. R.; Klopfenstein, J. D.; Pinson, D. M.; Wang, D. Z.; Vemuganti, R.; Veeravalli, K. K. MMP-12, a Promising Therapeutic Target for Neurological Diseases. *Mol. Neurobiol.* **2018**, *55* (2), 1405–1409.
- (19) Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. BRENDA in 2017: New Perspectives and New Tools in BRENDA. *Nucleic Acids Res.* **2017**, *45* (D1), D380–D388.
- (20) Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New Approach for Understanding Genome Variations in KEGG. *Nucleic Acids Res.*
- (21) Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. EXPASY: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* **2012**, *40* (W1), W597–W603.
- (22) Rawlings, N. D.; Barrett, A. J.; Thomas, P. D.; Huang, X.; Bateman, A.; Finn, R. D. The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison with Peptidases in the PANTHER Database. *Nucleic Acids Res.* **2018**, *46* (D1), D624–D632.
- (23) Li, Y.; Voorhees, J. J.; Fisher, G. J. Identification of Dihydrogambogic Acid as a Matrix Metalloproteinase 1 Inhibitor by High-Throughput Screening. *Clin. Cosmet. Investig. Dermatol.* **2017**, *Volume 10*, 499–502.
- (24) Yuan, H.; Lu, W.; Wang, L.; Shan, L.; Li, H.; Huang, J.; Sun, Q.; Zhang, W. Synthesis of Derivatives of Methyl Rosmarinate and Their Inhibitory Activities against Matrix Metalloproteinase-1 (MMP-1). *Eur. J. Med. Chem.* **2013**, *62*, 148–157.
- (25) Wang, Z.-C.; Shen, F.-Q.; Yang, M.-R.; You, L.-X.; Chen, L.-Z.; Zhu, H.-L.; Lu, Y.-D.; Kong, F.-L.; Wang, M.-H. Dihydropyrazothiazole Derivatives as Potential MMP-2/MMP-8 Inhibitors for Cancer Therapy. *Bioorg. Med. Chem. Lett.* **2018**.
- (26) Shamsara, J. Identification of Non-Zinc Binding Inhibitors of MMP-2 Through Virtual Screening and Subsequent Rescoring. *Drug Res. (Stuttg)*. **2018**, *68* (09), 529–535.
- (27) Romanchikova, N.; Trapencieris, P.; Zemītis, J.; Turks, M. A Novel Matrix Metalloproteinase-2 Inhibitor Triazolylmethyl Aziridine Reduces Melanoma Cell Invasion, Angiogenesis and Targets ERK1/2 Phosphorylation. *J. Enzyme Inhib. Med. Chem.* **2014**, *29* (6), 765–772.
- (28) Gooyit, M.; Song, W.; Mahasenan, K. V.; Lichtenwalter, K.; Suckow, M. A.; Schroeder, V. A.; Wolter, W. R.; Mobashery, S.; Chang, M. O-Phenyl Carbamate and Phenyl Urea Thiiranes as Selective Matrix Metalloproteinase-2 Inhibitors That Cross the Blood-Brain Barrier. *J. Med. Chem.* **2013**, *56* (20), 8139–8150.
- (29) Fabre, B.; Filipiak, K.; Zapico, J. M.; Díaz, N.; Carbajo, R. J.; Schott, A. K.; Martínez-Alcázar, M. P.; Suárez, D.; Pineda-Lucena, A.; Ramos, A.; et al. Progress towards Water-Soluble Triazole-Based Selective MMP-2 Inhibitors. *Org. Biomol. Chem.* **2013**, *11* (38), 6623.

Manuscript 3

- (30) Di Pizio, A.; Laghezza, A.; Tortorella, P.; Agamennone, M. Probing the S1' Site for the Identification of Non-Zinc-Binding MMP-2 Inhibitors. *ChemMedChem* **2013**, *8* (9), 1421–1482.
- (31) Higashi, S.; Hirose, T.; Takeuchi, T.; Miyazaki, K. Molecular Design of a Highly Selective and Strong Protein Inhibitor against Matrix Metalloproteinase-2 (MMP-2). *J. Biol. Chem.* **2013**, *288* (13), 9066–9076.
- (32) Zapico, J. M.; Serra, P.; García-Sanmartín, J.; Filipiak, K.; Carbajo, R. J.; Schott, A. K.; Pineda-Lucena, A.; Martínez, A.; Martín-Santamaría, S.; de Pascual-Teresa, B.; et al. Potent “Clicked” MMP2 Inhibitors: Synthesis, Molecular Modeling and Biological Exploration. *Org. Biomol. Chem.* **2011**, *9* (12), 4587.
- (33) Bencsik, P.; Kupai, K.; Görbe, A.; Kenyeres, É.; Varga, Z. V.; Pálóczi, J.; Gáspár, R.; Kovács, L.; Weber, L.; Takács, F.; et al. Development of Matrix Metalloproteinase-2 Inhibitors for Cardioprotection. *Front. Pharmacol.* **2018**, *9*, 296.
- (34) Adhikari, N.; Halder, A. K.; Mallick, S.; Saha, A.; Saha, K. D.; Jha, T. 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. *Bioorg. Med. Chem.* **2016**, *24* (18), 4291–4309.
- (35) He, W.; Jiang, J.; Yu, Z.-Q.; Zhou, J.-H. Novel 5-Hydroxy, 5-Substituted Benzenesulfonamide Pyrimidine-2,4,6-Triones Attenuate Lipopolysaccharide-Induced Acute Lung Injury via Inhibition of the Gelatinases, MMP-2 and MMP-9. *Drug Dev. Res.* **2016**, *77* (5), 251–257.
- (36) Song, J.; Peng, P.; Chang, J.; Liu, M. M.; Yu, J. M.; Zhou, L.; Sun, X. Selective Non-Zinc Binding MMP-2 Inhibitors: Novel Benzamide Iloprost Analogs with Anti-Tumor Metastasis. *Bioorganic Med. Chem. Lett.* **2016**, *26* (9), 2174–2178.
- (37) Wang, P.-F.; Qiu, H.-Y.; Baloch, S. K.; Gong, H.-B.; Wang, Z.-C.; Zhu, H.-L. Synthesis, Biological Evaluation, and Docking of Dihydropyrazole Sulfonamide Containing 2-Hydroxyphenyl Moiety: A Series of Novel MMP-2 Inhibitors. *Chem. Biol. Drug Des.* **2015**, *86* (6), 1405–1410.
- (38) Yan, X. Q.; Wang, Z. C.; Li, Z.; Wang, P. F.; Qiu, H. Y.; Chen, L. W.; Lu, X. Y.; Lv, P. C.; Zhu, H. L. Sulfonamide Derivatives Containing Dihydropyrazole Moieties Selectively and Potently Inhibit MMP-2/MMP-9: Design, Synthesis, Inhibitory Activity and 3D-QSAR Analysis. *Bioorganic Med. Chem. Lett.* **2015**, *25* (20), 4664–4671.
- (39) Rane, R. A.; Naphade, S. S.; Bangalore, P. K.; Palkar, M. B.; Patel, H. M.; Shaikh, M. S.; Alwan, W. S.; Karpoornath, R. Synthesis of Novel Hybrids Inspired from Bromopyrrole Alkaloids Inhibiting MMP-2 and -12 as Antineoplastic Agents. *Chem. Biol. Drug Des.* **2015**, *86* (2), 210–222.
- (40) Nanjan, P.; Nambiar, J.; Nair, B. G.; Banerji, A. Synthesis and Discovery of (1-3,11-3)-Biacacetin as a Novel Non-Zinc Binding Inhibitor of MMP-2 and MMP-9. *Bioorganic Med. Chem.* **2015**, *23* (13), 3781–3787.
- (41) Edman, K.; Furber, M.; Hemsley, P.; Johansson, C.; Paireudeau, G.; Petersen, J.; Stocks, M.; Tervo, A.; Ward, A.; Wells, E.; et al. The Discovery of MMP7 Inhibitors Exploiting a Novel Selectivity Trigger. *ChemMedChem* **2011**, *6* (5), 769–773.
- (42) Hou, J.; Zou, Q.; Wang, Y.; Gao, Q.; Yao, W.; Yao, Q.; Zhang, J. Screening for the Selective Inhibitors of MMP-9 from Natural Products Based on Pharmacophore Modeling and Molecular

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

- Docking in Combination with Bioassay Experiment, Hybrid QM/MM Calculation, and MD Simulation. *J. Biomol. Struct. Dyn.* **2018**, 1–50.
- (43) Shirian, J.; Arkadash, V.; Cohen, I.; Sapir, T.; Radisky, E. S.; Papo, N.; Shifman, J. M. Converting a Broad Matrix Metalloproteinase Family Inhibitor into a Specific Inhibitor of MMP-9 and MMP-14. *FEBS Lett.* **2018**, 592 (7), 1122–1134.
- (44) Gao, Q.; Wang, Y.; Hou, J.; Yao, Q.; Zhang, J. Multiple Receptor-Ligand Based Pharmacophore Modeling and Molecular Docking to Screen the Selective Inhibitors of Matrix Metalloproteinase-9 from Natural Products. *J. Comput. Aided. Mol. Des.* **2017**, 31 (7), 625–641.
- (45) Zheng, X.-Z.; Zhou, J.-L.; Ye, J.; Guo, P.-P.; Lin, C.-S. Cardioprotective Effect of Novel Sulphonamides-1,3,5-Triazine Conjugates against Ischaemic-Reperfusion Injury via Selective Inhibition of MMP-9. *Chem. Biol. Drug Des.* **2016**, 88 (5), 756–765.
- (46) Kalva, S.; Azhagiya Singam, E. R.; Rajapandian, V.; Saleena, L. M.; Subramanian, V. Discovery of Potent Inhibitor for Matrix Metalloproteinase-9 by Pharmacophore Based Modeling and Dynamics Simulation Studies. *J. Mol. Graph. Model.* **2014**, 49, 25–37.
- (47) Mori, M.; De Lorenzo, E.; Torre, E.; Fragai, M.; Nativi, C.; Luchinat, C.; Arcangeli, A. A Highly Soluble Matrix Metalloproteinase-9 Inhibitor for Potential Treatment of Dry Eye Syndrome. *Basic Clin. Pharmacol. Toxicol.* **2012**, 111 (5), 289–295.
- (48) Wang, J.; O'Sullivan, S.; Harmon, S.; Keaveny, R.; Radomski, M. W.; Medina, C.; Gilmer, J. F. Design of Barbiturate-Nitrate Hybrids That Inhibit MMP-9 Activity and Secretion. *J. Med. Chem.* **2012**, 55 (5), 2154–2162.
- (49) Senn, N.; Ott, M.; Lanz, J.; Riedl, R. Targeted Polypharmacology: Discovery of a Highly Potent Non-Hydroxamate Dual Matrix Metalloproteinase (MMP)-10/13 Inhibitor. *J. Med. Chem.* **2017**, 60 (23), acs.jmedchem.7b01001.
- (50) Nuti, E.; Cuffaro, D.; Bernardini, E.; Camodeca, C.; Panelli, L.; Chaves, S.; Ciccone, L.; Tepshi, L.; Vera, L.; Orlandini, E.; et al. Development of Thioaryl-Based Matrix Metalloproteinase-12 Inhibitors with Alternative Zinc-Binding Groups: Synthesis, Potentiometric, NMR, and Crystallographic Studies. *J. Med. Chem.* **2018**, 61 (10), 4421–4435.
- (51) Butsch, V.; Börgel, F.; Galla, F.; Schwegmann, K.; Hermann, S.; Schäfers, M.; Riemann, B.; Wünsch, B.; Wagner, S. Design, (Radio)Synthesis, and in Vitro and in Vivo Evaluation of Highly Selective and Potent Matrix Metalloproteinase 12 (MMP-12) Inhibitors as Radiotracers for Positron Emission Tomography. *J. Med. Chem.* **2018**, 61 (9), 4115–4134.
- (52) Nuti, E.; Cuffaro, D.; D'Andrea, F.; Rosalia, L.; Tepshi, L.; Fabbri, M.; Carbotti, G.; Ferrini, S.; Santamaria, S.; Camodeca, C.; et al. Sugar-Based Arylsulfonamide Carboxylates as Selective and Water-Soluble Matrix Metalloproteinase-12 Inhibitors. *ChemMedChem* **2016**, 11 (15), 1626–1637.
- (53) Aerts, J.; Vandenbroucke, R. E.; Dera, R.; Balusu, S.; Van Wonterghem, E.; Moons, L.; Libert, C.; Dehaen, W.; Arckens, L. Synthesis and Validation of a Hydroxypyronone-Based, Potent, and Specific Matrix Metalloproteinase-12 Inhibitor with Anti-Inflammatory Activity In Vitro and In Vivo. *Mediators Inflamm.* **2015**, 2015, 510679.
- (54) Santamaria, S.; Nuti, E.; Cercignani, G.; Marinelli, L.; La Pietra, V.; Novellino, E.; Rossello, A. N-O-Isopropyl Sulfonamido-Based Hydroxamates: Kinetic Characterisation of a Series of MMP-12/MMP-13 Dual Target Inhibitors. *Biochem. Pharmacol.* **2012**, 84 (6), 813–820.

Manuscript 3

- (55) Wu, Y.; Li, J.; Wu, J.; Morgan, P.; Xu, X.; Rancati, F.; Vallese, S.; Raveglia, L.; Hotchandani, R.; Fuller, N.; et al. Discovery of Potent and Selective Matrix Metalloprotease 12 Inhibitors for the Potential Treatment of Chronic Obstructive Pulmonary Disease (COPD). *Bioorg. Med. Chem. Lett.* **2012**, *22* (1), 138–143.
- (56) Badland, M.; Compère, D.; Courté, K.; Dublanchet, A.-C.; Blais, S.; Manage, A.; Peron, G.; Wrigglesworth, R. Thiophene and Bioisostere Derivatives as New MMP12 Inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *21* (1), 528–530.
- (57) Ando, N.; Terashima, S. Synthesis and Matrix Metalloproteinase-12 Inhibitory Activity of Ageladine A Analogs. *Chem. Pharm. Bull. (Tokyo)*. **2011**, *59* (5), 579–596.
- (58) Devel, L.; Garcia, S.; Czarny, B.; Beau, F.; Lajeunesse, E.; Vera, L.; Georgiadis, D.; Stura, E.; Dive, V. Insights from Selective Non-Phosphinic Inhibitors of MMP-12 Tailored to Fit with an S1' Loop Canonical Conformation. *J. Biol. Chem.* **2010**, *285* (46), 35900–35909.
- (59) Ramezani, M.; Shamsara, J. Virtual Screening on MMP-13 Led to Discovering New Inhibitors Including a Non-Zinc Binding and a Micro Molar One: A Successful Example of Receptor Selection According to Cross-Docking Results for a Flexible Enzyme. *Comb. Chem. High Throughput Screen.* **2017**, *20* (8), 719–725.
- (60) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Discovery of Novel, Highly Potent, and Selective Quinazoline-2- Carboxamide-Based Matrix Metalloproteinase (MMP)-13 Inhibitors without a Zinc Binding Group Using a Structure-Based Design Approach. *J. Med. Chem.* **2014**, *57* (21), 8886–8902.
- (61) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Thieno[2,3-d]Pyrimidine-2-Carboxamides Bearing a Carboxybenzene Group at 5-Position: Highly Potent, Selective, and Orally Available MMP-13 Inhibitors Interacting with the S1'' Binding Site. *Bioorganic Med. Chem.* **2014**, *22* (19), 5487–5505.
- (62) Fischer, T.; Riedl, R. Strategic Targeting of Multiple Water-Mediated Interactions: A Concise and Rational Structure-Based Design Approach to Potent and Selective MMP-13 Inhibitors. *ChemMedChem* **2013**, *8* (9), 1457–1461.
- (63) De Savi, C.; Waterson, D.; Pape, A.; Lamont, S.; Hadley, E.; Mills, M.; Page, K. M.; Bowyer, J.; Maciewicz, R. A. Hydantoin Based Inhibitors of MMP13 - Discovery of AZD6605. *Bioorganic Med. Chem. Lett.* **2013**, *23* (16), 4705–4712.
- (64) Gege, C.; Bao, B.; Bluhm, H.; Boer, J.; Gallagher, B. M.; Korniski, B.; Powers, T. S.; Steeneck, C.; Taveras, A. G.; Baragi, V. M. Discovery and Evaluation of a Non-Zn Chelating, Selective Matrix Metalloproteinase 13 (MMP-13) Inhibitor for Potential Intra-Articular Treatment of Osteoarthritis. *J. Med. Chem.* **2012**, *55* (2), 709–716.
- (65) De Savi, C.; Morley, A. D.; Nash, I.; Karoutchi, G.; Page, K.; Ting, A.; Gerhardt, S. Lead Optimisation of Selective Non-Zinc Binding Inhibitors of MMP13. Part 2. *Bioorganic Med. Chem. Lett.* **2012**, *22* (1), 271–277.
- (66) La Pietra, V.; Marinelli, L.; Cosconati, S.; Di Leva, F. S.; Nuti, E.; Santamaria, S.; Pugliesi, I.; Morelli, M.; Casalini, F.; Rossello, A.; et al. Identification of Novel Molecular Scaffolds for the Design of MMP-13 Inhibitors: A First Round of Lead Optimization. *Eur. J. Med. Chem.* **2012**, *47* (1), 143–152.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

- (67) Taylor, S. J.; Abeywardane, A.; Liang, S.; Muegge, I.; Padyana, A. K.; Xiong, Z.; Hill-Drzewi, M.; Farmer, B.; Li, X.; Collins, B.; et al. Fragment-Based Discovery of Indole Inhibitors of Matrix Metalloproteinase-13. *J. Med. Chem.* **2011**, *54* (23), 8174–8187.
- (68) Roth, J.; Minond, D.; Darout, E.; Liu, Q.; Lauer, J.; Hodder, P.; Fields, G. B.; Roush, W. R. Identification of Novel, Exosite-Binding Matrix Metalloproteinase-13 Inhibitor Scaffolds. *Bioorg. Med. Chem. Lett.* **2011**, *21* (23), 7180–7184.
- (69) Tommasi, R. A.; Weiler, S.; Mcquire, L. W.; Rogel, O.; Chambers, M.; Clark, K.; Doughty, J.; Fang, J.; Ganu, V.; Grob, J.; et al. Potent and Selective 2-Naphthylsulfonamide Substituted Hydroxamic Acid Inhibitors of Matrix Metalloproteinase-13. *Bioorganic Med. Chem. Lett.* **2011**, *21* (21), 6440–6445.
- (70) Choi, J. Y.; Fuerst, R.; Knapinska, A. M.; Taylor, A. B.; Smith, L.; Cao, X.; Hart, P. J.; Fields, G. B.; Roush, W. R. Structure-Based Design and Synthesis of Potent and Selective Matrix Metalloproteinase 13 Inhibitors. *J. Med. Chem.* **2017**, *60* (13), 5816–5825.
- (71) Savi, C. De; Morley, A. D.; Ting, A.; Nash, I.; Karabelas, K.; Wood, C. M.; James, M.; Norris, S. J.; Karoutchi, G.; Rankine, N.; et al. Selective Non Zinc Binding Inhibitors of MMP13. *Bioorganic Med. Chem. Lett.* **2011**, *21* (14), 4215–4219.
- (72) Vicini, P.; Crascì, L.; Incerti, M.; Ronsisvalle, S.; Cardile, V.; Panico, A. M. Benzisothiazolyliminothiazolidin-4-Ones with Chondroprotective Properties: Searching for Potent and Selective Inhibitors of MMP-13. *ChemMedChem* **2011**, *6* (7), 1199–1202.
- (73) Gao, D. A.; Xiong, Z.; Heim-Riether, A.; Amodeo, L.; August, E. M.; Cao, X.; Ciccarelli, L.; Collins, B. K.; Harrington, K.; Haverty, K.; et al. SAR Studies of Non-Zinc-Chelating MMP-13 Inhibitors: Improving Selectivity and Metabolic Stability. *Bioorganic Med. Chem. Lett.* **2010**, *20* (17), 5039–5043.
- (74) Piecha, D.; Weik, J.; Kheil, H.; Becher, G.; Timmermann, A.; Jaworski, A.; Burger, M.; Hofmann, M. W. Novel Selective MMP-13 Inhibitors Reduce Collagen Degradation in Bovine Articular and Human Osteoarthritis Cartilage Explants. *Inflamm. Res.* **2010**, *59* (5), 379–389.
- (75) Schnute, M. E.; O'Brien, P. M.; Nahra, J.; Morris, M.; Howard Roark, W.; Hanau, C. E.; Ruminski, P. G.; Scholten, J. A.; Fletcher, T. R.; Hamper, B. C.; et al. Discovery of (Pyridin-4-Yl)-2H-Tetrazole as a Novel Scaffold to Identify Highly Selective Matrix Metalloproteinase-13 Inhibitors for the Treatment of Osteoarthritis. *Bioorg. Med. Chem. Lett.* **2010**, *20* (2), 576–580.
- (76) Nara, H.; Kaieda, A.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; et al. Discovery of Novel, Highly Potent, and Selective Matrix Metalloproteinase (MMP)-13 Inhibitors with a 1,2,4-Triazol-3-Yl Moiety as a Zinc Binding Group Using a Structure-Based Design Approach. *J. Med. Chem.* **2017**, *60* (2), 608–626.
- (77) Hugenberg, V.; Wagner, S.; Kopka, K.; Schäfers, M.; Schuit, R. C.; Windhorst, A. D.; Hermann, S. Radiolabeled Selective Matrix Metalloproteinase 13 (MMP-13) Inhibitors: (Radio)Syntheses and in Vitro and First in Vivo Evaluation. *J. Med. Chem.* **2017**, *60* (1), 307–321.
- (78) Nara, H.; Sato, K.; Kaieda, A.; Oki, H.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; Uchikawa, O.; Kori, M. Design, Synthesis, and Biological Activity of Novel, Potent, and Highly Selective Fused Pyrimidine-2-Carboxamide-4-One-Based Matrix Metalloproteinase (MMP)-13 Zinc-Binding Inhibitors. *Bioorganic Med. Chem.* **2016**, *24* (23), 6149–6165.

Manuscript 3

- (79) Kothapalli, R.; Sivaraman Siveen, K.; Tan, T. Z.; Paul Thiery, J.; Kumar, A. P.; Sethi, G.; Swaminathan, K. Functional Characterization of Selective Exosite-Binding Inhibitors of Matrix Metalloproteinase-13 (MMP-13) - Experimental Validation in Human Breast and Colon Cancer. *Biosci. Biotechnol. Biochem.* **2016**, *80* (11), 1–10.
- (80) Fischer, T.; Riedl, R. Molecular Recognition of the Catalytic Zinc(II) Ion in MMP-13: Structure-Based Evolution of an Allosteric Inhibitor to Dual Binding Mode Inhibitors with Improved Lipophilic Ligand Efficiencies. *Int. J. Mol. Sci.* **2016**, *17* (3), 314.
- (81) Ruminski, P. G.; Massa, M.; Strohbach, J.; Hanau, C. E.; Schmidt, M.; Scholten, J. A.; Fletcher, T. R.; Hamper, B. C.; Carroll, J. N.; Shieh, H. S.; et al. Discovery of N-(4-Fluoro-3-Methoxybenzyl)-6-(2-(((2S,5R)-5-(Hydroxymethyl)-1,4-Dioxan-2-Yl)Methyl)-2H-Tetrazol-5-Yl)-2-Methylpyrimidine-4-Carboxamide. A Highly Selective and Orally Bioavailable Matrix Metalloproteinase-13 Inhibitor for the Potential Treat. *J. Med. Chem.* **2016**, *59* (1), 313–327.
- (82) Spicer, T. P.; Jiang, J.; Taylor, A. B.; Choi, J. Y.; Hart, P. J.; Roush, W. R.; Fields, G. B.; Hodder, P. S.; Minond, D. Characterization of Selective Exosite-Binding Inhibitors of Matrix Metalloproteinase 13 That Prevent Articular Cartilage Degradation in Vitro. *J. Med. Chem.* **2014**, *57* (22), 9598–9611.
- (83) Arkadash, V.; Yosef, G.; Shirian, J.; Cohen, I.; Horev, Y.; Grossman, M.; Sagi, I.; Radisky, E. S.; Shifman, J. M.; Papo, N. Development of High Affinity and High Specificity Inhibitors of Matrix Metalloproteinase 14 through Computational Design and Directed Evolution. *J. Biol. Chem.* **2017**, *292* (8), 3481–3495.
- (84) RCSB PDB <http://www.rcsb.org>.
- (85) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (86) Friesner, R. a; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. a; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
- (87) Hashimoto, H.; Takeuchi, T.; Komatsu, K.; Miyazaki, K.; Sato, M.; Higashi, S. Structural Basis for Matrix Metalloproteinase-2 (MMP-2)-Selective Inhibitory Action of β -Amyloid Precursor Protein-Derived Inhibitor. *J. Biol. Chem.* **2011**, *286* (38), 33236–33243.
- (88) Esser, C. K.; Bugianesi, R. L.; Caldwell, C. G.; Chapman, K. T.; Durette, P. L.; Girotra, N. N.; Kopka, I. E.; Lanza, T. J.; Levorse, D. A.; MacCoss, M.; et al. Inhibition of Stromelysin-1 (MMP-3) by P1'-Biphenylethyl Carboxyalkyl Dipeptides. *J. Med. Chem.* **1997**, *40* (6), 1026–1040.
- (89) Gavuzzo, E.; Pochetti, G.; Mazza, F.; Gallina, C.; Gorini, B.; D'Alessio, S.; Pieper, M.; Tschesche, H.; Tucker, P. A. Two Crystal Structures of Human Neutrophil Collagenase, One Complexed with a Primed- and the Other with an Unprimed-Side Inhibitor: Implications for Drug Design. *J. Med. Chem.* **2000**, *43* (18), 3377–3385.
- (90) Morales, R.; Perrier, S.; Florent, J. M.; Beltra, J.; Dufour, S.; De Mendez, I.; Manceau, P.; Tertre, A.; Moreau, F.; Compere, D.; et al. Crystal Structures of Novel Non-Peptidic, Non-Zinc Chelating Inhibitors Bound to MMP-12. *J. Mol. Biol.* **2004**, *341* (4), 1063–1076.
- (91) Devel, L.; Beau, F.; Amoura, M.; Vera, L.; Cassar-Lajeunesse, E.; Garcia, S.; Czarny, B.; Stura, E. A.; Dive, V. Simple Pseudo-Dipeptides with a P2' Glutamate: A Novel Inhibitor Family of Matrix Metalloproteases and Other Metzincins. *J. Biol. Chem.* **2012**, *287* (32), 26647–26656.

Understanding the variability of the S1' pocket to improve MMP1 selectivity profiles

- (92) Holmes, I. P. P.; Gaines, S.; Watson, S. P. P.; Lorthioir, O.; Walker, A.; Baddeley, S. J. J.; Herbert, S.; Egan, D.; Convery, M. A. M. A. M. A. A.; Singh, O. M. P. M. P.; et al. The Identification of β -Hydroxy Carboxylic Acids as Selective MMP-12 Inhibitors. *Bioorganic Med. Chem. Lett.* **2009**, *19* (19), 5760–5763.
- (93) Pochetti, G.; Montanari, R.; Gege, C.; Chevrier, C.; Taveras, A. G.; Mazza, F. Extra Binding Region Induced by Non-Zinc Chelating Inhibitors into the S1' Subsite of Matrix Metalloproteinase 8 (MMP-8). *J. Med. Chem.* **2009**, *52* (4), 1040–1049.
- (94) Heim-Riether, A.; Taylor, S. J.; Liang, S.; Gao, D. A.; Xiong, Z.; Michael August, E.; Collins, B. K.; Farmer, B. T.; Haverty, K.; Hill-Drzewi, M.; et al. Improving Potency and Selectivity of a New Class of Non-Zn-Chelating MMP-13 Inhibitors. *Bioorganic Med. Chem. Lett.* **2009**, *19* (18), 5321–5324.
- (95) Shieh, H. S.; Tomasselli, A. G.; Mathis, K. J.; Schnute, M. E.; Woodard, S. S.; Caspers, N.; Williams, J. M.; Kiefer, J. R.; Munie, G.; Wittwer, A.; et al. Structure Analysis Reveals the Flexibility of the ADAMTS-5 Active Site. *Protein Sci.* **2011**, *20* (4), 735–744.
- (96) Becker, D. P.; Barta, T. E.; Bedell, L. J.; Boehm, T. L.; Bond, B. R.; Carroll, J.; Carron, C. P.; Decrescenzo, G. A.; Easton, A. M.; Freskos, J. N.; et al. Orally Active MMP-1 Sparing α -Tetrahydropyranil and α -PiperidinyI Sulfone Matrix Metalloproteinase (MMP) Inhibitors with Efficacy in Cancer, Arthritis, and Cardiovascular Disease. *J. Med. Chem.* **2010**, *53* (18), 6653–6680.
- (97) Monovich, L. G.; Tommasi, R. A.; Fujimoto, R. A.; Blancuzzi, V.; Clark, K.; Cornell, W. D.; Doti, R.; Doughty, J.; Fang, J.; Farley, D.; et al. Discovery of Potent, Selective, and Orally Active Carboxylic Acid Based Inhibitors of Matrix Metalloproteinase-13. *J. Med. Chem.* **2009**, *52* (11), 3523–3538.
- (98) Tochowicz, A.; Maskos, K.; Huber, R.; Oltenfreiter, R.; Dive, V.; Yiotakis, A.; Zanda, M.; Bode, W.; Goettig, P. Crystal Structures of MMP-9 Complexes with Five Inhibitors: Contribution of the Flexible Arg424 Side-Chain to Selectivity. *J. Mol. Biol.* **2007**, *371* (4), 989–1006.
- (99) Nuti, E.; Cantelmo, A. R.; Gallo, C.; Bruno, A.; Bassani, B.; Camodeca, C.; Tuccinardi, T.; Vera, L.; Orlandini, E.; Nencetti, S.; et al. N-O-Isopropyl Sulfonamido-Based Hydroxamates as Matrix Metalloproteinase Inhibitors: Hit Selection and in Vivo Antiangiogenic Activity. *J. Med. Chem.* **2015**, *58* (18), 7224–7240.
- (100) Camodeca, C.; Nuti, E.; Tepshi, L.; Boero, S.; Tuccinardi, T.; Stura, E. A.; Poggi, A.; Zocchi, M. R.; Rossello, A. Discovery of a New Selective Inhibitor of A Disintegrin and Metalloprotease 10 (ADAM-10) Able to Reduce the Shedding of NKG2D Ligands in Hodgkin's Lymphoma Cell Models. *Eur. J. Med. Chem.* **2016**, *111*, 193–201.
- (101) Mannino, C.; Nievo, M.; Machetti, F.; Papakyriakou, A.; Calderone, V.; Fragai, M.; Guarna, A. Synthesis of Bicyclic Molecular Scaffolds (BTAA): An Investigation towards New Selective MMP-12 Inhibitors. *Bioorganic Med. Chem.* **2006**, *14* (22), 7392–7403.
- (102) Campestre, C.; Agamennone, M.; Tortorella, P.; Preziuso, S.; Biasone, A.; Gavuzzo, E.; Pochetti, G.; Mazza, F.; Hiller, O.; Tschesche, H.; et al. N-Hydroxyurea as Zinc Binding Group in Matrix Metalloproteinase Inhibition: Mode of Binding in a Complex with MMP-8. *Bioorganic Med. Chem. Lett.* **2006**, *16* (1), 20–24.
- (103) Johnson, A. R.; Pavlovsky, A. G.; Ortwine, D. F.; Prior, F.; Man, C. F.; Bornemeier, D. A.; Banotai, C. A.; Mueller, W. T.; McConnell, P.; Yan, C.; et al. Discovery and Characterization of a Novel Inhibitor

Manuscript 3

- of Matrix Metalloprotease-13 That Reduces Cartilage Damage in Vivo without Joint Fibroplasia Side Effects. *J. Biol. Chem.* **2007**, *282* (38), 27781–27791.
- (104) Marvin 16.10.10.0, 2016, ChemAxon <http://www.chemaxon.com>.
- (105) Reaxys <https://www.reaxys.com/>.
- (106) Schrödinger Release 2018-1: Maestro, Schrödinger, LLC, New York, NY, 2018.
- (107) Iyer, S.; Wei, S.; Brew, K.; Acharya, K. R. 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.* **2007**, *282* (1), 364–371.
- (108) Schrödinger Release 2018-1: Glide, Schrödinger, LLC, New York, NY, 2018.
- (109) Belviso, B. D.; Caliandro, R.; Siliqi, D.; Calderone, V.; Arnesano, F.; Natile, G. Structure of Matrix Metalloproteinase-3 with a Platinum-Based Inhibitor. *Chem. Commun.* **2013**, *49* (48), 5492.
- (110) Kohno, T.; Hochigai, H.; Yamashita, E.; Tsukihara, T.; Kanaoka, M. Crystal Structures of the Catalytic Domain of Human Stromelysin-1 (MMP-3) and Collagenase-3 (MMP-13) with a Hydroxamic Acid Inhibitor SM-25453. *Biochem. Biophys. Res. Commun.* **2006**, *344* (1), 315–322.
- (111) Steele, D. L.; El-Kabbani, O.; Dunten, P.; Windsor, L. J.; Kammlott, R. U.; Crowther, R. L.; Michoud, C.; Engler, J. A.; Birktoft, J. J. Expression, Characterization and Structure Determination of an Active Site Mutant (Glu202-Gln) of Mini-Stromelysin-1. *Protein Eng.* **2000**, *13* (6), 397–405.
- (112) Tauro, M.; Laghezza, A.; Loiodice, F.; Piemontese, L.; Caradonna, A.; Capelli, D.; Montanari, R.; Pochetti, G.; Di Pizio, A.; Agamennone, M.; et al. Catechol-Based Matrix Metalloproteinase Inhibitors with Additional Antioxidative Activity. *J. Enzyme Inhib. Med. Chem.* **2016**, *31* (October 2017), 25–37.
- (113) Antoni, C.; Vera, L.; Devel, L.; Catalani, M. P.; Czarny, B.; Cassar-Lajeunesse, E.; Nuti, E.; Rossello, A.; Dive, V.; Stura, E. A. Crystallization of Bi-Functional Ligand Protein Complexes. *J. Struct. Biol.* **2013**, *182* (3), 246–254.
- (114) Nar, H.; Werle, K.; Bauer, M. M.; Dollinger, H.; Jung, B. Crystal Structure of Human Macrophage Elastase (MMP-12) in Complex with a Hydroxamic Acid Inhibitor. *J. Mol. Biol.* **2001**, *312* (4), 743–751.
- (115) Bertini, I.; Calderone, V.; Fragai, M.; Jaiswal, R.; Luchinat, C.; Melikian, M.; Mylonas, E.; Svergun, D. I. Evidence of Reciprocal Reorientation of the Catalytic and Hemopexin-like Domains of Full-Length MMP-12. *J. Am. Chem. Soc.* **2008**, *130* (22), 7011–7021.
- (116) Grossman, M.; Tworowski, D.; Dym, O.; Lee, M.-H.; Levy, Y.; Murphy, G.; Sagi, I. The Intrinsic Protein Flexibility of Endogenous Protease Inhibitor TIMP-1 Controls Its Binding Interface and Affects Its Function. *Biochemistry* **2010**, *49* (29), 6184–6192.

Manuscript 4

Identification of selective MMP-13 inhibitors by virtual screening

Aleix Gimeno^[a], Elisa Nuti^[b], María José Ojeda-Montes^[a], Sarah Tomás-Hernández^[a],
Adrià Cereto-Massagué^[a], Raúl Beltrán-Debón^[a], Miquel Mulero^[a], Armando Rossello^[b],
Gerard Pujadas^{[a],[c].*}, Santiago Garcia-Vallvé^{[a],[c]}

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]Dipartimento di Farmacia, Università di Pisa, Via Bonanno 6, 56126 Pisa, Italy

^[c]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Abstract

Osteoarthritis is a pathology of the joint characterized by pain, swelling, ankylosis and limited mobility. Matrix metalloproteinase 13 plays a central role in osteoarthritis as its over-expression in osteoarthritis patients induces an excessive breakdown of collagen that results in an imbalance between collagen synthesis and degradation in the joint, leading to the progressive degradation of articular collagen. Therefore MMP-13 has been proposed as a key therapeutical target for osteoarthritis. Here we have developed a virtual screening workflow aimed at identifying selective MMP-13 inhibitors by targeting the deep S1' pocket of MMP-13, which is not present in other MMPs. The virtual screening workflow consisted of a molecular weight filter, a shape similarity analysis and a protein-ligand docking step. Compounds were selected based on the data reported in previous structure-activity relationship studies of MMP-13 inhibitors. As a result, three MMP-13 inhibitors with IC_{50} values of 91 μ M, 105 μ M and 15 μ M were obtained, one of which displayed at least 4-fold selectivity against MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

1. Introduction

Osteoarthritis (OA) is the most common form of arthritis,^{1,2} affecting half of the aged population (> 65 years).³ It is characterized by the progressive degradation of articular collagen and can ultimately result in the prosthetic replacement of joints as they become completely dysfunctional.³

Matrix metalloproteinase 13 (MMP-13) plays a central role in the pathology, as it is the main responsible for the cleavage of type II collagen in patients with OA.^{4,5} Several signaling pathways result in the modulation of MMP-13 activity in OA, both directly by regulatory factors such as LRP1,⁶ leptin⁷ and microRNAs that directly target MMP-13 (e.g. miR-9,⁸ miR-146a,⁹⁻¹¹ miR-127-5p,¹² miR-27b,¹³ miR-320¹⁴ and miR-136¹⁵); and at the epigenetic, transcriptional and post-transcriptional levels by transcription factors (e.g. LEF1,¹⁶ NF- κ B,¹⁷ ELF3,¹⁷ HIF2 α ¹⁷ and RUNX-2¹⁸) and non-coding microRNAs (e.g. miR-27a,¹⁹ mirR-140,¹⁹ miR-488,²⁰ miR-24,²¹ miR-148a,²² miR-222,²³ miR-22,²⁴ miR-181b,²⁵ miR-33a,²⁶ miR-145,²⁷ miR-483²⁸). As a result, MMP-13 is significantly over-expressed in the joints and articular cartilage in patients with OA and therefore has been proposed as a key therapeutical target for the treatment of OA.³

MMP-13 belongs to the MMP family, which consists of a series of enzymes responsible for the degradation of different extracellular matrix (ECM) components.²⁹ In addition to tissue remodelling, MMPs are involved in the cleavage of many non-matrix targets, such as cell surface receptors, cytokines, chemokines, cell-cell adhesion molecules, clotting factors and other proteinases.³⁰ Moreover, some MMPs have been identified to play protective roles and thus they are considered antitargets, as is the case of MMP-3 and MMP-8, which have been attributed anti-tumoral properties.³¹ Actually, many broad spectrum MMP inhibitors have failed clinical trials as patients developed musculoskeletal syndrome (MSS), possibly resulting from the alteration of the physiological functions of different members of the MMP family.^{32,33} Therefore, selectivity is currently considered a priority in the development of MMP inhibitors.

Manuscript 4

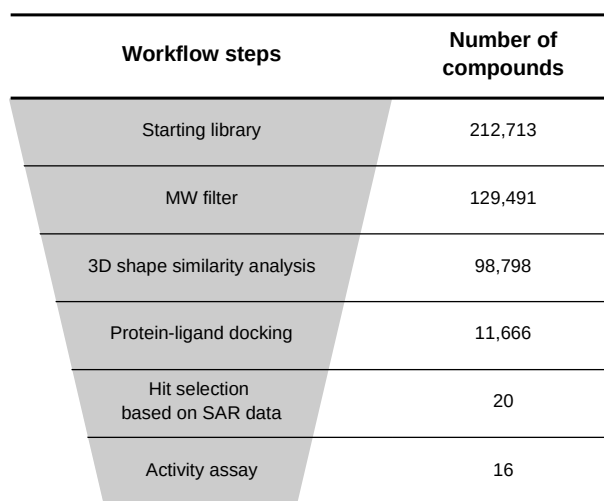
The main challenge in the development of selective MMP inhibitors arises from the fact that the catalytic sites of the enzymes in the MMP family present a high degree of homology. In this regard, the characteristics of the ligand binding site of MMP-13 are slightly different from other MMPs, thus providing an edge on the identification of selective inhibitors for this enzyme. More specifically, an adjacent region to the catalytic site, known as the S1' pocket, is different in MMP-13 as the loop that delimits the pocket (Ω -loop) is longer and shows more flexibility in MMP-13 than in other MMPs.³⁴ This allows for an opportunity to identify inhibitors with a different binding mode not possible in the binding sites of other enzymes of the MMP family. Several inhibitors use this difference in the MMP-13 binding site to achieve selectivity for this enzyme, as it is shown by their X-ray structures.³⁵⁻³⁸ The aim of the present work is to use this crystallographic data to design a virtual screening (VS) methodology able to identify MMP-13 inhibitors that can adopt a similar binding mode, and therefore achieve selectivity towards MMP-13.

2. Results and Discussion

Co-crystallized MMP-13 inhibitors that bind to the S1' pocket show a similar binding mode.³⁵⁻³⁸ All of them present two common characteristics: **a)** they contain two aromatic rings or ring systems at both ends of the molecule (with the exception of the co-crystallized inhibitor in the structure with PDB code 3KEK,³⁶ which instead presents a cyclohexyl group at one of its ends); and **b)** they adopt a characteristic *U* shape, establishing hydrogen bond interactions between the core of the molecule and Thr245, Thr247, or both residues. We use here these characteristics to design a VS workflow capable of identifying compounds that bind in a similar manner to the S1' pocket of MMP-13. The VS workflow, which was applied to an initial library of 212,713 compounds obtained from Specs,³⁹ consisted of 4 steps that are described in this section (see Figure 1). First, the compounds were filtered by MW; then, a shape-based alignment was performed to keep only the compounds that could adopt a conformation similar to that of the co-crystallized ligands; next, protein-ligand docking was performed

Identification of selective MMP-13 inhibitors by virtual screening

on MMP-13; and, finally, compounds were selected based on the interactions with the binding site of MMP-13 that improved MMP-13 inhibitor activity in previously reported structure-activity relationship (SAR) studies.



Workflow steps	Number of compounds
Starting library	212,713
MW filter	129,491
3D shape similarity analysis	98,798
Protein-ligand docking	11,666
Hit selection based on SAR data	20
Activity assay	16

Figure 1. Diagram of the virtual screening workflow that indicates the different filters used and the number of compounds that overcome each one of them.

2.1. Molecular weight filter

The compounds obtained from the Specs database were filtered by molecular weight (MW) in order to discard compounds too small to fulfill the posterior protein-ligand docking constraints and compounds too large compared to the reference ligands used for the subsequent shape-based alignment, therefore reducing the computational time of the subsequent steps. The 300 – 700 Da range was selected as filter, taking in to account that the compounds used as reference in the shape-based alignment step have a MW between 392 and 491 Da. As a result of this first step, 83,222 compounds were filtered out (see Figure 1).

2.2. Shape-based similarity

Next, a maximum of 10 conformations were generated for each compound that survived the MW filter. These conformations were compared to those of selective

Manuscript 4

MMP-13 inhibitors co-crystallized with MMP-13 that bind to the S1' pocket and do not contain a zinc-binding group. Only the conformations similar to those of co-crystallized ligands were kept in order to reduce the computational cost of the VS, as compounds unable to adopt a similar shape than that of the co-crystallized ligands would likely not fit in the S1' cavity during the protein-ligand docking step. As a result of this step, 30,693 compounds were filtered out (see Figure 1).

2.3. Protein-ligand docking

The ligands resulting from the shape similarity filter were docked onto MMP-13, using the crystal structure with PDB^{40,41} code 3WV1,³⁷ as it contains an inhibitor that binds to the S1' pocket. In order to discard ligands unable to adopt a similar binding mode in the S1' pocket to that of previously known selective MMP-13 inhibitors, two positional constraints were defined (one closer to the zinc binding region and another one deep in the S1' pocket) to be fulfilled by aromatic atoms. Moreover, it was required that the ligand performed a hydrogen bond interaction with Thr245 or Thr247, as all the co-crystallized inhibitors performed at least one of these interactions with the core of the molecule.

2.4. Analysis of SAR data

In order to select the compounds that performed the appropriate interactions in the S1' pocket of MMP-13, several SAR studies were analyzed to obtain information regarding which interactions are important to achieve high activity towards MMP-13. Eventually, several conclusions were reached on how to increase the bioactivity of inhibitors by targeting different areas of the S1' pocket. The main features that a potent MMP-13 inhibitor should have are:

1. Have a negatively charged ring substituent that can establish a salt bridge interaction with Lys140 in the S1' pocket

In a SAR study reported by Nara *et al.*,³⁷ the modification of compound **21h** by the addition of a carboxylic acid group in the *para* position of the ring that occupies the S1'

Identification of selective MMP-13 inhibitors by virtual screening

pocket resulted in compound **21k**, which presented a 10-fold increase of bioactivity for MMP-13. As compound **21k** is able to establish a salt bridge interaction with Lys140, this could be the reason for the notable increase in potency. In another study reported by Nara *et al.*,⁴² the authors incorporated a carboxylic acid group to compound **26a**, generating compound **26c**, which displayed >25-fold increase in MMP-13 bioactivity compared to the original compound. In similar structures with different linkers to the S1' pocket, they also incorporated this modification to compounds **32a**, **29a**, **35a** and **41** to generate the compounds **32c**, **29c**, **35c**, and **38**, respectively. In all these cases, the incorporation of the carboxylic acid group resulted in an increase of bioactivity compared to the original compound (5-fold, 8-fold, 3-fold and 2-fold, respectively). Similarly, in a study reported by Taylor *et al.*,³⁸ compounds presenting a carboxylic acid in this region (*i.e.* **15** and **16**) also showed a higher bioactivity (up to 150-fold) compared to compounds with a different substituent (*i.e.* **13**, **14** and **17**). This data indicates that a negatively charged substituent (preferably a carboxylic acid) that can establish a salt bridge interaction with Lys140 is highly favored in this position.

2. Make a π - π interaction with Tyr246 and Phe252

In a SAR study reported by Nara *et al.*,³⁷ the compound **9m** bearing a phenyl moiety displayed slightly higher bioactivity (1.4-fold) than the compound **21f** bearing a cyclohexyl moiety, indicating that the formation of a π - π interaction with Tyr246 and Phe252 results in an increase of bioactivity for MMP-13. In the same study, the modification of compound **21h** bearing a phenyl moiety to generate compound **21i** bearing a *p*-fluorophenyl moiety resulted in a 3-fold decrease of bioactivity. This case could be explained by the weakening of the π - π interaction in **21i** respect to **21h**, as the fluor atom in *para* position of the benzyl ring acts as an electron withdrawing group, thus weakening the negative charge of the π system and therefore weakening the π - π interactions of the compound with Tyr246 and Phe252. This data indicates that a π - π interaction with Tyr246 and Phe252 slightly favors compound bioactivity towards MMP-13.

3. Have a hydrophobic moiety occupying the S1'' pocket

In a SAR study reported by Nara *et al.*,³⁷ compound **37** was modified by incorporating a substituent that extended to the S1'' pocket. In most cases (*i.e.*, compounds **26b**, **26c**, **21a**, **9m**, **21b**, **21c**, **21d**, **21e**, **26d**, **34**, **36** and **21f**) this resulted in an increase of the inhibitory activity (5-fold, 23-fold, 39-fold, 17-fold, 23-fold, 8-fold, 3-fold, 1.2-fold, 4-fold, 7-fold, 75-fold and 12-fold, respectively). Similarly, in a study reported by Taylor *et al.*,³⁸ compound **11** was modified to further extend to the S1'' pocket. In this case, most of the resulting compounds (*i.e.*, **13**, **14**, **15**, **16** and **17**) showed a considerable increase in bioactivity (*i.e.*, 28-fold, 21-fold, 2500-fold, 833-fold and 17-fold, respectively). This data indicates that designing compounds that occupy the S1'' pocket of MMP-13 can have a very favorable effect on inhibitor bioactivity.

4. Have an appropriate linker to join the S1' pocket with the S1'' pocket

Nara *et al.*⁴² reported several compound derivatives with different linkers extending from the core of the compound designed in order to reach the S1'' pocket. Compounds **32a**, **29a**, **35a** and **41**, which contained amide linkers (*i.e.* $(CH_2)CONH$, $(CH_2)NHCO$, $CONH(CH_2)$ and $NHCO(CH_2)$, respectively) showed higher bioactivity values towards MMP-13 than compounds **26a**, **26d**, **26f**, which contained oxygen (*i.e.* $(CH_2)O(CH_2)$), sulphur (*i.e.* $(CH_2)S(CH_2)$) and *NMe* (*i.e.* $(CH_2)NMe(CH_2)$) linkers, respectively. Among the different amide linkers, the one which had the greatest contribution to inhibitor activity was the linker $(CH_2)NHCO$, present in compounds **29a** and **29c**, as these compounds showed higher bioactivity values compared to the respective compounds with other amide linkers (*i.e.* **32a**, **35a** and **41**; and **32c**, **35c** and **38**). This data shows that amide linkers can be accommodated well in the region between the S1' and the S1'' pockets, thus opening a window to further increase MMP-13 inhibitor bioactivity.

5. Have a hydrogen or halogen bond acceptor towards Met253 N_i

In a SAR study reported by Nara *et al.*,³⁷ substituents were incorporated to interact with the MMP-13 residue Met253. Substituting a hydrogen atom (compound **37**) for a

Identification of selective MMP-13 inhibitors by virtual screening

fluorine atom (compound **9e**) in that position led only to a slight increase in bioactivity (1.1-fold), possibly due to the formation of a weak halogen bond with Met253. Nevertheless, introducing methoxy (compound **9b**), ethoxy (compound **17**) and *SMe* (compound **19a**) groups in this position resulted in greater bioactivity gains (3-fold and 5-fold, respectively). These groups may enhance bioactivity by accepting the hydrogen atom of the backbone nitrogen of Met253. The introduction of hydrophobic substituents (e.g. methyl group in compound **9f** and trifluoromethyl group in compound **9g**) and bulkier substituents (e.g. *OCF₃* in compound **9h**, *OBn* in compound **9i**, *SO₂Me* in compound **19b**) in this position led to losses in bioactivity, presumably due to the hydrophilic character and size of this region of the pocket. Therefore, this data shows that a small hydrophilic substituent able to establish a hydrogen bond with Met253 by accepting its main chain hydrogen atom is beneficial for MMP-13 inhibitory activity. On the other hand, introducing a hydrophobic group towards this region should be avoided as it results in a decrease of bioactivity.

6. Make a hydrophobic interaction with Pro255

In a SAR study reported by Schnute *et al.*,³⁶ a methyl group was introduced to several compounds in the region close to Pro255. Compounds that presented the methyl group in this position (i.e. compounds **16b**, **17b**, **18b**, **19b**, **20b** and **23b**) had higher bioactivity values than their respective non-methylated analogues (i.e. compounds **16a**, **17a**, **18a**, **19a**, **20a** and **23a**). The methyl substituent on the pyridine ring of the compounds which significantly improved MMP-13 inhibition (12-fold, 3-fold, 12-fold, 9-fold, 12-fold and 7-fold, respectively) was found to fill the hydrophobic region of the S1' pocket, buttressing against Pro255. Therefore, hydrophobic interactions with Pro255 may contribute to a tighter binding by MMP-13 inhibitors and the increase of inhibitor bioactivity. In another study, Nara *et al.*,⁴² introduced modifications in the same region, obtaining that a thiophene was preferred to a furan ring in that position (e.g. compound **7d** displayed a 13-fold increase in bioactivity respect to compound **7e**). They also obtained that a *NH*-containing ring was preferred to a *N(CH₃)*-containing ring in that position (e.g. compound **7j** displayed a 79-fold increase in bioactivity respect to compound **7i**). In both of these cases, the varying atoms face the residue Pro255 and

Manuscript 4

thus it can be assumed that the hydrophobic character of the substituents in that position confers these compounds a higher MMP-13 inhibitory activity, probably by establishing hydrophobic interactions with Pro255.

7. Have a hydrogen bond acceptor towards the side chain of Thr247

In a SAR study reported by Nara *et al.*,³⁷ substituents were introduced in the region facing Thr247. The modification of compound **37** for compound **9i**, which presented a methyl group towards this area, resulted in a 2-fold decrease in bioactivity, possibly due to unfavorable contacts between this substituent and the side chain of Thr247. In contrast, the addition of atoms able to accept the hydroxyl group hydrogen in the side chain of Thr247 resulted in the generation of compounds with greater potency. For instance, compound **9m**, which contained an O-based linker, displayed a 6-fold improvement in potency compared to compound **21d**, which instead contained a NH-based linker. Compound **9m** also displayed 14-fold and 4-fold potency increases compared to compounds **21e** and **26d**, which contained more hydrophobic S-based and CH₂-based linkers, respectively. This data shows that forming a hydrogen bond with Thr247 by accepting the hydrogen atom of its side-chain thiol group is the preferred way of increasing inhibitor bioactivity by targeting this residue. Introducing a hydrophobic group towards this region should be avoided as it results in a decrease of bioactivity.

8. Establish hydrophobic contacts in the S1' pocket

In a study by Nara *et al.*,⁴² adding methyl substituents in different positions of the ring system that occupied the S1' pocket resulted in higher bioactivities (e.g. compounds **7b**, **7c** and **7d** respectively showed 22-fold, 8-fold and 20-fold increased bioactivity respect to compound **7a**). In the same study, modifications of the ring system size and polarity also resulted in changes of bioactivity. More concretely, the substitution of a phenyl-based ring system in compound **44** for more polar pyridine-based ring systems in compounds **7m** and **7l** decreased inhibitor bioactivity by 17-fold and 267-fold, respectively. This data shows how maximizing the hydrophobic contacts in the S1' pocket is an important factor for MMP-13 inhibitor potency.

9. Hydrogen bond acceptor towards Thr245 N_i

In a study by Nara *et al.*,⁴² compounds with several linkers between the core of the molecule and the part of the molecule closer to the zinc atom were reported. The amide linker in compound **46** displayed >385-fold higher potency than the secondary amine linker in compound **13** and the reverse amide linker in compound **10**. While the replacement of the amide nitrogen in compound **46** with a methylene group in compound **23** resulted only in a minor drop of potency (8-fold), the reduction of this carbonyl group to a secondary hydroxyl group in compound **21** resulted in a drastic drop in potency (>45-fold). This can be explained by the ability of the carbonyl group present in compounds **46** and **23** to establish a hydrogen bond with Thr245.

10. Have a hydrogen bond donor towards Ala238 C_i

In the study by Nara *et al.*,⁴² the 8-fold higher potency displayed by compound **46**, with an amide linker, compared to compound **23**, with a carbonyl linker, may be attributed to the ability of the amide linker to establish a hydrogen bond interaction with Ala238. Therefore, establishing a hydrogen bond with Ala238 should contribute to a slight increase of inhibitory potency.

11. Have an appropriate ring substituent in the region close to the zinc-binding group

The substituents of the ring close to the zinc region also have been shown to have an effect on bioactivity. Taylor *et al.*³⁸ showed that the introduction of a $COO(CH_2)CH_3$ group in *para* position resulted in increases in bioactivity compared to the compound with no ring substituents (>26000-fold in compound **15** respect to compound **19**) as well as compounds with other substituents (25000-fold and 250-fold in compound **15** respect to compounds **18** and **20**, respectively, >13-fold in compound **1** respect to compounds **2** and **3** and 7-fold in compound **4** respect to compound **5**). The introduction of a $(CH_2)O(CH_2)CH_3$ group (compound **20**) in *para* position also resulted in a >104-fold increase in bioactivity compared to the compound with no ring substituents

Manuscript 4

(compound **19**). In a study by Schnute *et. al.*,³⁶ compounds that beared a *p*-methoxy substituent in the ring located in the same position showed higher activity than compounds that beared other substituents. For instance, compound **19a** had 2-fold and 6-fold higher bioactivity than compounds **16a** and **21a**, which contained a fluorine and a chlorine atom in that position, respectively. The same occurred with compound **19b**, which had 5-fold and 1.2-fold higher bioactivity than compounds **15b** and **16b** that contained a hydrogen and a fluorine atom in that position, respectively. In that study, modifications were also introduced in the *meta* position, the methoxy group resulting to confer the highest potency as well. Compound **20a** displayed 36-fold, 9-fold and 8-fold higher bioactivity compared to compounds **17a**, **22a** and **23a**, with fluoro, cholo and trifluoromethyl groups, respectively, in that position. Compound **20b** displayed 159-fold, 124-fold and 14-fold higher bioactivity compared to compounds **17b**, **15b** and **23b**, with a fluoro substituent, a hydrogen atom and a trifluoromethyl substituent, respectively, in that position. Compound **29g** displayed 196-fold, 138-fold, 42-fold, 9-fold and 9-fold higher bioactivity compared to compounds **29a**, **29d**, **29e**, **29f** and **29h**, which in that position presented respectively a hydrogen atom, a fluoro group, a cyano group, a hydroxyl group and a $OCH(CH_3)_2$ group. Compound **29k** displayed 25-fold, 15-fold, 5-fold and 2-fold higher bioactivity compared to compounds **29i**, **29b**, **29j** and **29m**, with a fluoro group, no substituent, a hydroxyl group and a trifluoromethyl group, respectively, in that position.

2.5. Hit selection

After docking, the top 100 docked poses for each compound were selected based on their docking scores. The interactions they performed with the protein were carefully inspected in order to select the compounds that accomplished the criteria obtained through the analysis of SAR data. Ideally in this step we would like to obtain a compound that accomplished all the 11 criteria. However, this was not the case, as the compound that accomplished more criteria was compound **1**, with a total of 8. Therefore, this compound and compounds that fulfilled most of the criteria were selected, obtaining 20 compounds for *in vitro* tests (see Figure 2). The structures of these 20 compounds were compared, using their molecular fingerprints, to those of

Identification of selective MMP-13 inhibitors by virtual screening

previously reported MMP-13 inhibitors in the Reaxys⁴³ database. Except for compound **14**, which showed a Tanimoto value of 0.71 with a described MMP-13 inhibitor, the Tanimoto similarity values of the selected compounds with any of the previously reported MMP-13 inhibitors analyzed were below 0.5. Upon visual inspection, compound **14** was not considered structurally similar to the MMP-13 inhibitor obtained from Reaxys⁴³ (see Figure S1). Moreover, hit compounds that were structurally similar to previously selected hit compounds were discarded to ensure that the final 20 compounds selected for *in vitro* tests were structurally different from each other (see Figure 3). Therefore, the hit compounds obtained by this virtual screening methodology not only proved to be different from previously reported MMP-13 inhibitors, but also structurally diverse.

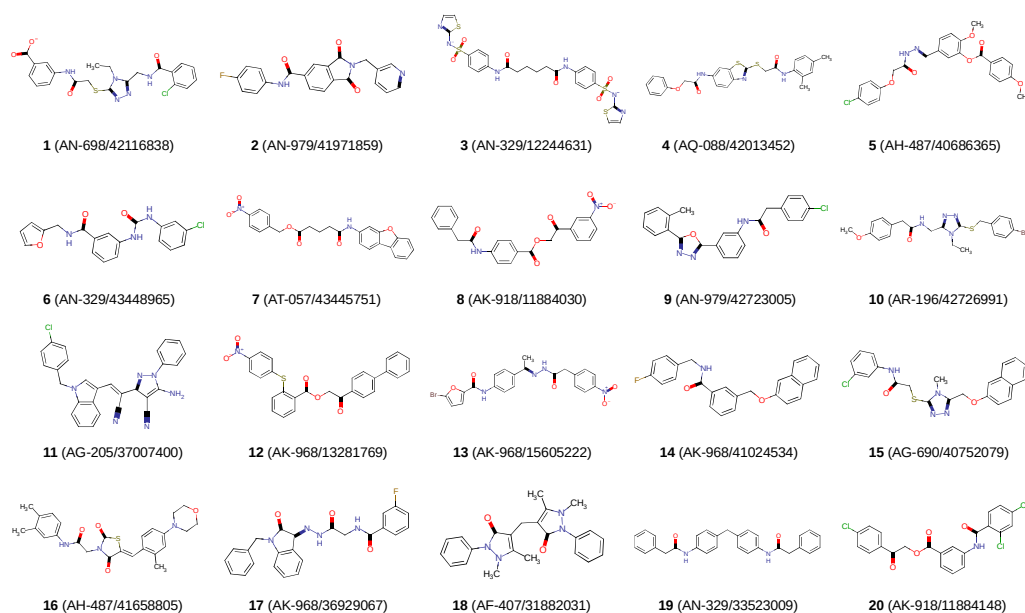


Figure 2. 2D structures of the 20 hit compounds. Each compound is identified with its Specs ID number. MarvinSketch⁴⁴ was used to draw the structures. The protonation state of each compound corresponds to the protonation state of the docked pose selected for that compound.

Manuscript 4

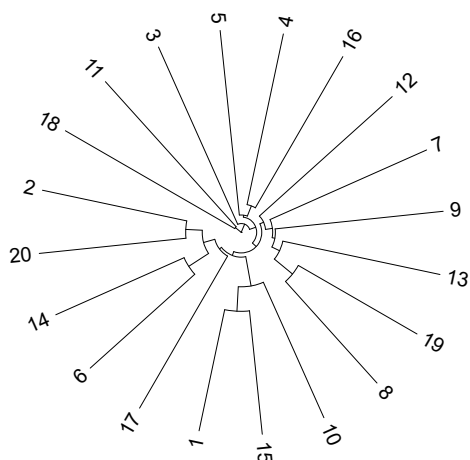


Figure 3. Dendrogram based on fingerprints showing the structural diversity of the 20 hit compounds. The fingerprint used to obtain the distance matrix was the OpenEyePath fingerprint.⁴⁵ iTOL⁴⁶ was used to draw the dendrogram.

2.6. Activity assays

After the selection of the 20 hit compounds, they were purchased from Specs³⁹ and their activity for MMP-13 was analyzed *in vitro* at a concentration of 100 μ M (see Table 1). Compounds **11**, **12** and **13** displayed the highest MMP-13 inhibitory activity and their IC₅₀ values were obtained (91 μ M, 105 μ M and 15 μ M, respectively). Next, the inhibitory activities of these three compounds towards MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14 were determined (see Table 1). Whereas compounds **11** and **12** did not show a preference for the inhibition of MMP-13 respect to other MMPs, compound **13** displayed at least 4-fold selectivity towards MMP-13 compared to the rest of the MMPs tested.

Identification of selective MMP-13 inhibitors by virtual screening

Table 1. In vitro activity of compounds **1-20** (percentage of MMP-13 inhibition at 100 μM and IC_{50} values for different MMPs).

Compound	MMP-13 inhibition at 100 μM (%)	MMP-1 IC_{50} (μM)	MMP-2 IC_{50} (μM)	MMP-8 IC_{50} (μM)	MMP-9 IC_{50} (μM)	MMP-12 IC_{50} (μM)	MMP-13 IC_{50} (μM)	MMP-14 IC_{50} (μM)
1	12.5	ND	ND	ND	ND	ND	ND	ND
2	18.1	ND	ND	ND	ND	ND	ND	ND
3	5.8	ND	ND	ND	ND	ND	ND	ND
4	22.8	ND	ND	ND	ND	ND	ND	ND
5	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a
6	2.6	ND	ND	ND	ND	ND	ND	ND
7	29.1	ND	ND	ND	ND	ND	ND	ND
8	8.8	ND	ND	ND	ND	ND	ND	ND
9	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a
10	15.2	ND	ND	ND	ND	ND	ND	ND
11	54.2	99 \pm 8	60 \pm 4.6	89 \pm 10	66 \pm 4.6	97 \pm 13	91 \pm 8.5	94 \pm 5.5
12	48.7	75 \pm 4	47 \pm 1.6	76 \pm 6.9	72 \pm 4.8	77 \pm 7.7	105 \pm 5.8	67 \pm 4.6
13	63.1	91 \pm 0.8	99 \pm 6	95 \pm 4.4	68 \pm 3	63 \pm 9	15 \pm 1.6	63 \pm 5.4
14	28.5	ND	ND	ND	ND	ND	ND	ND
15	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a
16	ND ^b	ND ^b	ND ^b	ND ^b	ND ^b	ND ^b	ND ^b	ND ^b
17	20.9	ND	ND	ND	ND	ND	ND	ND
18	1.9	ND	ND	ND	ND	ND	ND	ND
19	11.1	ND	ND	ND	ND	ND	ND	ND
20	0	ND	ND	ND	ND	ND	ND	ND

IC_{50} activity assays were run in triplicate. The final values given here are the mean \pm standard deviation of three independent experiments. ND refers to "not determined". ^a Fluorescent at 400 nm. ^b Not soluble in DMSO.

4. Experimental section

4.1. Shape-based similarity

Conformations were generated using Omega^{47,48} with default parameters and requiring a maximum of 10 conformations. The co-crystallized inhibitors used as reference in the

Manuscript 4

shape comparison corresponded to the ligands of the crystal structures with the following PDB^{40,41} codes: 2OW9,³⁵ 2OZR,³⁵ 3KEC,³⁶ 3KEJ,³⁶ 3KEK,³⁶ 3WV1,³⁷ 5BPA.³⁸ Shape similarity between the library compounds and the reference compounds was calculated with ROCS^{49,50} using the *ShapeTanimoto* coefficient, a value between 0 and 1 calculated by the following equation:

$$\text{ShapeTanimoto}_{f,g} = O_{f,g} / (I_f + I_g - O_{f,g})$$

in which the I terms are the self-volume overlaps of each molecule, while the O term is the overlap between the two functions.

Conformations with a *ShapeTanimoto* value lower than 0.5 to any of the reference compounds were discarded.

4.2. Ligand setup for docking

Before docking, ligand molecules were prepared with LigPrep⁵¹ with default parameter values except for the following options: **a)** respect chiralities from input geometry when generating stereoisomers; **b)** use Epik⁵² for ionization and tautomerization; **c)** use 7.0 as effective pH; and **d)** use 2.0 as pH tolerance for generated structures.

4.3. Protein preparation

After verifying the fitting of the coordinates of the residues in the binding site relative to their corresponding electron density map with VHELIBS, the B chain of the crystal structure with PDB^{40,41} code 3WV1³⁷ was prepared by using Maestro's Protein Preparation Wizard⁵³ through the following procedure: **a)** align to: 1ROS, chain A; **b)** remove original hydrogens; **c)** cap termini; **d)** generate ionization and tautomeric states of the ligand with Epik;⁵² **e)** assign hydrogen bonds at pH 7 with PROPKA; **f)** use force field OPLS_2005 to minimize the structure at 0.30 Å; and **g)** remove all water molecules from the structure.

4.4. Grid generation

The grid for protein-ligand docking was generated with Maestro⁵⁴ by using default parameter values except for the following settings: **a)** the grid center coordinates were (46.0, 80.0, -1.0); **b)** halogens were included as acceptors; **c)** the inner box size was (10, 10, 10); **d)** the outer box size was (30, 30, 30); **e)** hydrogen bond constraints were defined on the backbone nitrogens of the residues Thr245 and Thr247 as well as the side-chain oxygen of the residue Thr245; and **f)** two positional constraints with a radius of 2 Å were defined on the coordinates (46.3, 80.1, -7.5) and (51.2, 80.5, 4.6), respectively.

4.5. Docking

Protein-ligand docking was performed with Glide⁵⁵ by using default parameter values except for the following settings: **a)** SP precision; **b)** enhance planarity of conjugated π groups; **c)** include halogens as acceptors; **d)** write out at most 10 poses per ligand; **e)** include 50 poses per ligand in post-docking minimization; **g)** require the accomplishment of both positional constraints by aromatic atoms; and **f)** require the accomplishment of one hydrogen bond constraint.

4.6. MMP inhibition assays

Recombinant human MMP-14 catalytic domain was a kind gift of Prof. Gillian Murphy (Department of Oncology, University of Cambridge, UK). Pro-MMP-1, pro-MMP-2, pro-MMP-9, pro-MMP-8 and pro-MMP-13 were purchased from Merck Millipore. Pro-MMP-12 was purchased from Bio-Techne. p-Aminophenylmercuric acetate (APMA) was from Sigma-Aldrich. Proenzymes were activated immediately prior to use with APMA 2 mM for 1 h at 37 °C for MMP-2 and MMP-8, APMA 2 mM for 2 h at 37 °C for MMP-1, APMA 1 mM for 30 min at 37 °C for MMP-13, APMA 1 mM for 4 h at 37 °C for MMP-12 and APMA 1 mM for 1 h at 37 °C for MMP-9). For assay measurements, the purchased compound stock solutions (10 mM in DMSO) were further diluted for each MMP in the fluorimetric assay buffer (FAB: Tris 50 mM, pH = 7.5, NaCl 150 mM, CaCl₂ 10 mM, Brij 35 0.05% and DMSO 1%). Activated enzyme (final concentration 0.56 nM for MMP-2,

Manuscript 4

0.3 nM for MMP-13, 1.3 nM for MMP-9, 1.4 nM for MMP-8, 1 nM for MMP-14cd, and 2.0 nM for MMP-1, 2.3 nM for MMP-12) and inhibitor solutions were incubated in the assay buffer for 3 h at 25 °C. After the addition of 200 µM solution of the fluorogenic substrate Mca-Lys-Pro-Leu-Gly-Leu-Dap(Dnp)-Ala-Arg-NH₂ (Merck Millipore) in DMSO (final concentration 2 µM), the hydrolysis was monitored every 15 sec for 15 min recording the increase in fluorescence ($\lambda_{\text{ex}} = 325 \text{ nm}$, $\lambda_{\text{em}} = 400 \text{ nm}$) using a Molecular Devices SpectraMax Gemini XPS plate reader. The assays were performed in triplicate in a total volume of 200 µL per well in 96-well microtitre plates (Corning, black, NBS). The MMP inhibition activity was expressed in relative fluorescent units (RFU). Percent of inhibition was calculated from control reactions without the inhibitor. The inhibitory effect of the tested compounds was routinely estimated at a concentration of 100 µM towards MMP-13. Those derivatives found to be active were tested at additional concentrations and IC₅₀ was determined using at least five concentrations of the inhibitor causing an inhibition between 10% and 90%, using the formula: $V_i/V_o = 1/(1 + [I]/IC_{50})$, where V_i is the initial velocity of substrate cleavage in the presence of the inhibitor at concentration $[I]$ and V_o is the initial velocity in the absence of the inhibitor. Results were analyzed using SoftMax Pro software and Origin 6.0 software.

3. Conclusions

In order to obtain potent and selective MMP-13 inhibitors, here we have developed a virtual screening workflow aimed at identifying compounds that target the S1' pocket of MMP-13, a region in the MMP binding site that has been shown to be different for MMP-13 respect to other MMPs. For this, we have first applied MW filter to discard compounds unlikely to survive subsequent filters. Next, we have used a shape-based similarity analysis to restrict the initial library of compounds to those able to adopt the characteristic U shape adopted by co-crystallized selective MMP-13 inhibitors. Then, we have performed protein-ligand docking simulations to predict the binding modes of these compounds. Finally, we have analyzed previously reported SAR studies to identify MMP-13 inhibitor interactions with the protein that are important for activity and

we have selected the docked poses obtained in the protein-ligand docking according to these criteria. The bioactivity assays have revealed that three hit compounds are capable of inhibiting MMP-13 at the μM range, one of which displays at least 4-fold selectivity against MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14.

Aknowledgments

The authors thank Gillian Murphy (Department of Oncology, University of Cambridge, U.K.). AG contract is supported by grant 2015FI_B00655 of the Catalonia Government. We thank OpenEye Scientific Software and Cresset BioMolecular Discovery Ltd. for kindly providing us with a software bursary for using their programs.

References

- (1) Nara, H.; Kaieda, A.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; et al. Discovery of Novel, Highly Potent, and Selective Matrix Metalloproteinase (MMP)-13 Inhibitors with a 1,2,4-Triazol-3-yl Moiety as a Zinc Binding Group Using a Structure-Based Design Approach. *J. Med. Chem.* **2017**, *60* (2), 608–626.
- (2) Arabelovic, S.; McAlindon, T. E. Considerations in the Treatment of Early Osteoarthritis. *Curr. Rheumatol. Rep.* **2005**, *7* (1), 29–35.
- (3) Li, H.; Wang, D.; Yuan, Y.; Min, J. New Insights on the MMP-13 Regulatory Network in the Pathogenesis of Early Osteoarthritis. *Arthritis Res. Ther.* **2017**, *19* (1), 248.
- (4) Choi, J. Y.; Fuerst, R.; Knapinska, A. M.; Taylor, A. B.; Smith, L.; Cao, X.; Hart, P. J.; Fields, G. B.; Roush, W. R. Structure-Based Design and Synthesis of Potent and Selective Matrix Metalloproteinase 13 Inhibitors. *J. Med. Chem.* **2017**, *60* (13), 5816–5825.
- (5) Takaishi, H.; Kimura, T.; Dalal, S.; Okada, Y.; D'Armiento, J. Joint Diseases and Matrix Metalloproteinases: A Role for MMP-13. *Curr. Pharm. Biotechnol.* **2008**, *9* (1), 47–54.
- (6) Yamamoto, K.; Okano, H.; Miyagawa, W.; Visse, R.; Shitomi, Y.; Santamaria, S.; Dudhia, J.; Troeberg, L.; Strickland, D. K.; Hirohata, S.; et al. MMP-13 Is Constitutively Produced in Human Chondrocytes and Co-Endocytosed with ADAMTS-5 and TIMP-3 by the Endocytic Receptor LRP1. *Matrix Biol.* **2016**, *56*, 57–73.
- (7) Iliopoulos, D.; Malizos, K. N.; Tsezou, A. Epigenetic Regulation of Leptin Affects MMP-13 Expression in Osteoarthritic Chondrocytes: Possible Molecular Target for Osteoarthritis Therapeutic Intervention. *Ann. Rheum. Dis.* **2007**, *66* (12), 1616–1621.
- (8) Jones, S. W.; Watkins, G.; Le Good, N.; Roberts, S.; Murphy, C. L.; Brockbank, S. M. V.; Needham, M. R. C.; Read, S. J.; Newham, P. The Identification of Differentially Expressed MicroRNA in Osteoarthritic Tissue That Modulate the Production of TNF-Alpha and MMP13. *Osteoarthr. Cartil.* **2009**, *17* (4), 464–472.
- (9) Yamasaki, K.; Nakasa, T.; Miyaki, S.; Ishikawa, M.; Deie, M.; Adachi, N.; Yasunaga, Y.; Asahara, H.; Ochi, M. Expression of MicroRNA-146a in Osteoarthritis Cartilage. *Arthritis Rheum.* **2009**, *60* (4), 1035–1041.
- (10) Li, X.; Gibson, G.; Kim, J.-S.; Kroin, J.; Xu, S.; van Wijnen, A. J.; Im, H.-J. MicroRNA-146a Is Linked to Pain-Related Pathophysiology of Osteoarthritis. *Gene* **2011**, *480* (1–2), 34–41.
- (11) Boldin, M. P.; Taganov, K. D.; Rao, D. S.; Yang, L.; Zhao, J. L.; Kalwani, M.; Garcia-Flores, Y.; Luong, M.; Devrekanli, A.; Xu, J.; et al. MiR-146a Is a Significant Brake on Autoimmunity, Myeloproliferation, and Cancer in Mice. *J. Exp. Med.* **2011**, *208* (6), 1189–1201.
- (12) Park, S. J.; Cheon, E. J.; Lee, M. H.; Kim, H. A. MicroRNA-127-5p Regulates Matrix Metalloproteinase 13 Expression and Interleukin-1 β -Induced Catabolic Effects in Human Chondrocytes. *Arthritis Rheum.* **2013**, *65* (12), 3141–3152.

Identification of selective MMP-13 inhibitors by virtual screening

- (13) Akhtar, N.; Rasheed, Z.; Ramamurthy, S.; Anbazhagan, A. N.; Voss, F. R.; Haqqi, T. M. MicroRNA-27b Regulates the Expression of Matrix Metalloproteinase 13 in Human Osteoarthritis Chondrocytes. *Arthritis Rheum.* **2010**, *62* (5), 1361–1371.
- (14) Meng, F.; Zhang, Z.; Chen, W.; Huang, G.; He, A.; Hou, C.; Long, Y.; Yang, Z.; Zhang, Z.; Liao, W. MicroRNA-320 Regulates Matrix Metalloproteinase-13 Expression in Chondrogenesis and Interleukin-1 β -Induced Chondrocyte Responses. *Osteoarthr. Cartil.* **2016**, *24* (5), 932–941.
- (15) Liu, Q.; Zhang, X.; Hu, X.; Dai, L.; Fu, X.; Zhang, J.; Ao, Y. Circular RNA Related to the Chondrocyte ECM Regulates MMP13 Expression by Functioning as a MiR-136 “Sponge” in Human Cartilage Degradation. *Sci. Rep.* **2016**, *6*, 22572.
- (16) Yun, K.; Im, S.-H. Transcriptional Regulation of MMP13 by Lef1 in Chondrocytes. *Biochem. Biophys. Res. Commun.* **2007**, *364* (4), 1009–1014.
- (17) Goldring, M. B.; Otero, M.; Plumb, D. A.; Dragomir, C.; Favero, M.; El Hachem, K.; Hashimoto, K.; Roach, H. I.; Olivetto, E.; Borzi, R. M.; et al. Roles of Inflammatory and Anabolic Cytokines in Cartilage Metabolism: Signals and Multiple Effectors Converge upon MMP-13 Regulation in Osteoarthritis. *Eur. Cell. Mater.* **2011**, *21*, 202–220.
- (18) Tetsunaga, T.; Nishida, K.; Furumatsu, T.; Naruse, K.; Hirohata, S.; Yoshida, A.; Saito, T.; Ozaki, T. Regulation of Mechanical Stress-Induced MMP-13 and ADAMTS-5 Expression by RUNX-2 Transcriptional Factor in SW1353 Chondrocyte-like Cells. *Osteoarthr. Cartil.* **2011**, *19* (2), 222–232.
- (19) Tardif, G.; Hum, D.; Pelletier, J.-P.; Duval, N.; Martel-Pelletier, J. Regulation of the IGFBP-5 and MMP-13 Genes by the MicroRNAs MiR-140 and MiR-27a in Human Osteoarthritic Chondrocytes. *BMC Musculoskelet. Disord.* **2009**, *10*, 148.
- (20) Song, J.; Kim, D.; Lee, C. H.; Lee, M. S.; Chun, C.-H.; Jin, E.-J. MicroRNA-488 Regulates Zinc Transporter SLC39A8/ZIP8 during Pathogenesis of Osteoarthritis. *J. Biomed. Sci.* **2013**, *20* (1), 31.
- (21) Philipot, D.; Guérit, D.; Platano, D.; Chuchana, P.; Olivetto, E.; Espinoza, F.; Dorandeu, A.; Pers, Y.-M.; Piette, J.; Borzi, R. M.; et al. P16INK4a and Its Regulator MiR-24 Link Senescence and Chondrocyte Terminal Differentiation-Associated Matrix Remodeling in Osteoarthritis. *Arthritis Res. Ther.* **2014**, *16* (1), R58.
- (22) Vonk, L. A.; Kragten, A. H. M.; Dhert, W. J. A.; Saris, D. B. F.; Creemers, L. B. Overexpression of Hsa-MiR-148a Promotes Cartilage Production and Inhibits Cartilage Degradation by Osteoarthritic Chondrocytes. *Osteoarthr. Cartil.* **2014**, *22* (1), 145–153.
- (23) Song, J.; Jin, E.-H.; Kim, D.; Kim, K. Y.; Chun, C.-H.; Jin, E.-J. MicroRNA-222 Regulates MMP-13 via Targeting HDAC-4 during Osteoarthritis Pathogenesis. *BBA Clin.* **2015**, *3*, 79–89.
- (24) Iliopoulos, D.; Malizos, K. N.; Oikonomou, P.; Tsezou, A. Integrative MicroRNA and Proteomic Approaches Identify Novel Osteoarthritis Genes and Their Collaborative Metabolic and Inflammatory Networks. *PLoS One* **2008**, *3* (11), e3740.
- (25) Song, J.; Lee, M.; Kim, D.; Han, J.; Chun, C.-H.; Jin, E.-J. MicroRNA-181b Regulates Articular Chondrocytes Differentiation and Cartilage Integrity. *Biochem. Biophys. Res. Commun.* **2013**, *431* (2), 210–214.
- (26) Kostopoulou, F.; Malizos, K. N.; Papatheanasiou, I.; Tsezou, A. MicroRNA-33a Regulates Cholesterol Synthesis and Cholesterol Efflux-Related Genes in Osteoarthritic Chondrocytes. *Arthritis Res. Ther.* **2015**, *17* (1), 42.

Manuscript 4

- (27) Martinez-Sanchez, A.; Dudek, K. A.; Murphy, C. L. Regulation of Human Chondrocyte Function through Direct Inhibition of Cartilage Master Regulator SOX9 by MicroRNA-145 (MiRNA-145). *J. Biol. Chem.* **2012**, *287* (2), 916–924.
- (28) Qi, Y.; Ma, N.; Yan, F.; Yu, Z.; Wu, G.; Qiao, Y.; Han, D.; Xiang, Y.; Li, F.; Wang, W.; et al. The Expression of Intronic MiRNAs, MiR-483 and MiR-483*, and Their Host Gene, Igf2, in Murine Osteoarthritis Cartilage. *Int. J. Biol. Macromol.* **2013**, *61*, 43–49.
- (29) Alberts, B.; Johnson, A. D.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 6th editio.; W. W. Norton & Company, 2014.
- (30) Vandenbroucke, R. E.; Libert, C. Is There New Hope for Therapeutic Matrix Metalloproteinase Inhibition? *Nat. Rev. Drug Discov.* **2014**, *13* (12), 904–927.
- (31) Alcantara, M. B.; Dass, C. R. Pigment Epithelium-Derived Factor as a Natural Matrix Metalloproteinase Inhibitor: A Comparison with Classical Matrix Metalloproteinase Inhibitors Used for Cancer Treatment. *J. Pharm. Pharmacol.* **2014**, *66* (7), 895–902.
- (32) Pulkoski-Gross, A. E. Historical Perspective of Matrix Metalloproteases. *Front. Biosci.* **2015**, *7* (JUNE 2015), 125–149.
- (33) Cathcart, J. M.; Cao, J. MMP Inhibitors: Past, Present and Future. *Front. Biosci. (Landmark Ed.)* **2015**, *20*, 1164–1178.
- (34) Fabre, B.; Ramos, A.; De Pascual-Teresa, B. Targeting Matrix Metalloproteinases: Exploring the Dynamics of the S1' Pocket in the Design of Selective, Small Molecule Inhibitors. *J. Med. Chem.* **2014**, *57* (24), 10205–10219.
- (35) Johnson, A. R.; Pavlovsky, A. G.; Ortwine, D. F.; Prior, F.; Man, C. F.; Bornemeier, D. A.; Banotai, C. A.; Mueller, W. T.; McConnell, P.; Yan, C.; et al. Discovery and Characterization of a Novel Inhibitor of Matrix Metalloprotease-13 That Reduces Cartilage Damage in Vivo without Joint Fibroplasia Side Effects. *J. Biol. Chem.* **2007**, *282* (38), 27781–27791.
- (36) Schnute, M. E.; O'Brien, P. M.; Nahra, J.; Morris, M.; Howard Roark, W.; Hanau, C. E.; Ruminski, P. G.; Scholten, J. A.; Fletcher, T. R.; Hamper, B. C.; et al. Discovery of (Pyridin-4-Yl)-2H-Tetrazole as a Novel Scaffold to Identify Highly Selective Matrix Metalloproteinase-13 Inhibitors for the Treatment of Osteoarthritis. *Bioorg. Med. Chem. Lett.* **2010**, *20* (2), 576–580.
- (37) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Discovery of Novel, Highly Potent, and Selective Quinazoline-2- Carboxamide-Based Matrix Metalloproteinase (MMP)-13 Inhibitors without a Zinc Binding Group Using a Structure-Based Design Approach. *J. Med. Chem.* **2014**, *57* (21), 8886–8902.
- (38) Taylor, S. J.; Abeywardane, A.; Liang, S.; Muegge, I.; Padyana, A. K.; Xiong, Z.; Hill-Drzewi, M.; Farmer, B.; Li, X.; Collins, B.; et al. Fragment-Based Discovery of Indole Inhibitors of Matrix Metalloproteinase-13. *J. Med. Chem.* **2011**, *54* (23), 8174–8187.
- (39) Specs <http://www.specs.net>.
- (40) RCSB PDB <http://www.rcsb.org>.
- (41) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (42) Nara, H.; Sato, K.; Naito, T.; Mototani, H.; Oki, H.; Yamamoto, Y.; Kuno, H.; Santou, T.; Kanzaki, N.; Terauchi, J.; et al. Thieno[2,3-d]Pyrimidine-2-Carboxamides Bearing a Carboxybenzene Group at 5-

Identification of selective MMP-13 inhibitors by virtual screening

Position: Highly Potent, Selective, and Orally Available MMP-13 Inhibitors Interacting with the S1'' Binding Site. *Bioorganic Med. Chem.* **2014**, 22 (19), 5487–5505.

- (43) Reaxys <https://www.reaxys.com/>.
- (44) Marvin 16.10.10.0, 2016, ChemAxon <http://www.chemaxon.com>.
- (45) OpenEye scientific software, OEChem, 2013.
- (46) Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Res.* **2016**, 44 (W1), W242-5.
- (47) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, 50 (4), 572–584.
- (48) OMEGA 3.0.1.2: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T.
- (49) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, 50 (1), 74–82.
- (50) ROCS 3.2.2.2: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.
- (51) Schrödinger Release 2016-3: LigPrep, Schrödinger, LLC, New York, NY, 2016.
- (52) Schrödinger Release 2016-3: Epik, Schrödinger, LLC, New York, NY, 2016.
- (53) Schrödinger Release 2016-3: Schrödinger Suite 2016-3 Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2016.
- (54) Schrödinger Release 2016-3: Maestro, Schrödinger, LLC, New York, NY, 2016.
- (55) Schrödinger Release 2016-3: Glide, Schrödinger, LLC, New York, NY, 2016.

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

Supplementary material for

***Identification of selective MMP-13 inhibitors by
virtual screening***

Aleix Gimeno^[a], Elisa Nuti^[b], María José Ojeda-Montes^[a], Sarah Tomás-Hernández^[a],
Adrià Cereto-Massagué^[a], Raúl Beltrán-Debón^[a], Miquel Mulero^[a], Armando Rossello^[b],
Gerard Pujadas^{[a],[c].*}, Santiago Garcia-Vallvé^{[a],[c]}

^[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

^[b]Dipartimento di Farmacia, Università di Pisa, Via Bonanno 6, 56126 Pisa, Italy

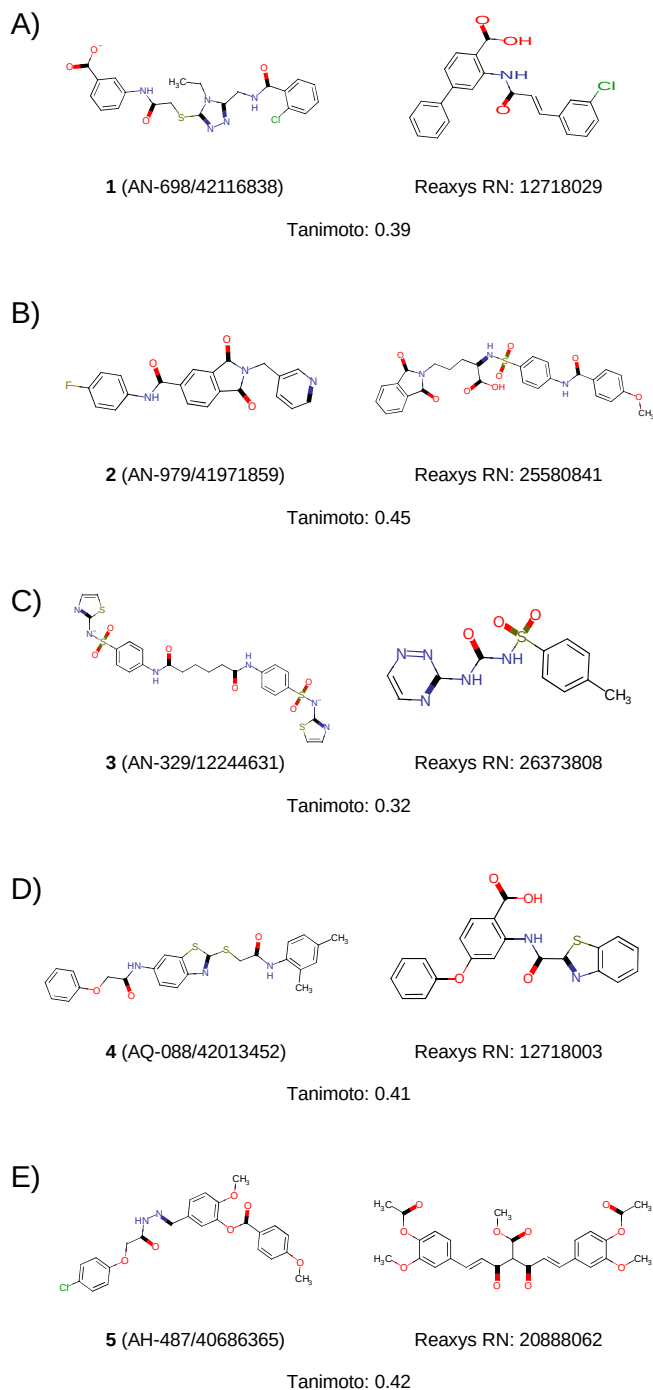
^[c]EURECAT, TECNIO, CEICS, Avinguda Universitat, 1, 43204 Reus, Catalonia, Spain

*Correspondence to: Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

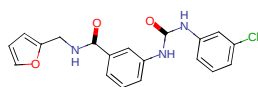
oo

Identification of selective MMP-13 inhibitors by virtual screening

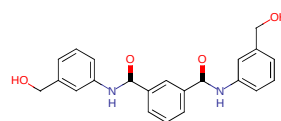


Manuscript 4

F)



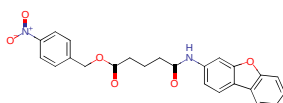
6 (AN-329/43448965)



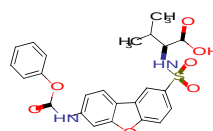
Reaxys RN: 13925759

Tanimoto: 0.49

G)



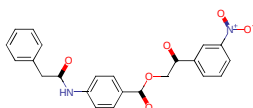
7 (AT-057/43445751)



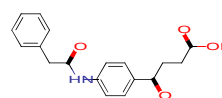
Reaxys RN: 25593517

Tanimoto: 0.42

H)



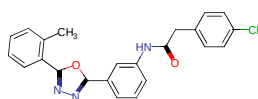
8 (AK-918/11884030)



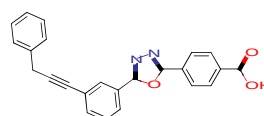
Reaxys RN: 13914187

Tanimoto: 0.54

I)



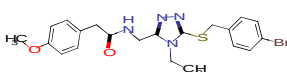
9 (AN-979/42723005)



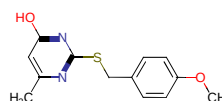
Reaxys RN: 11809753

Tanimoto: 0.49

J)



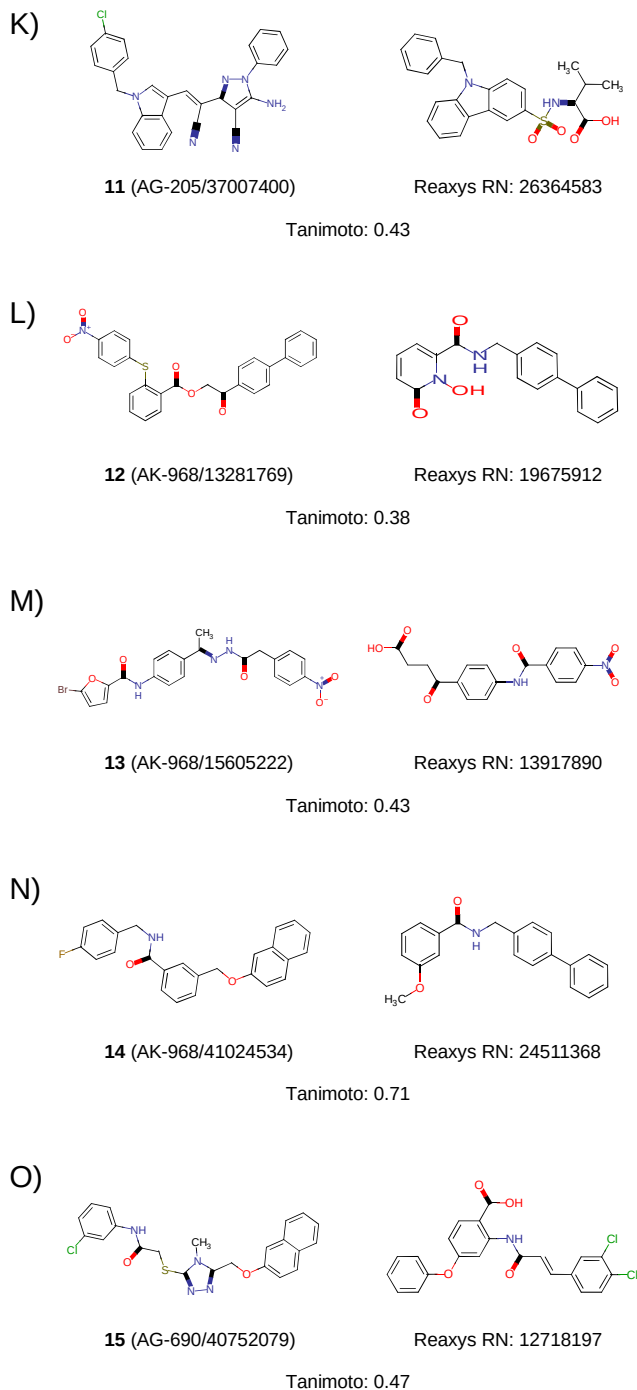
10 (AR-196/42726991)



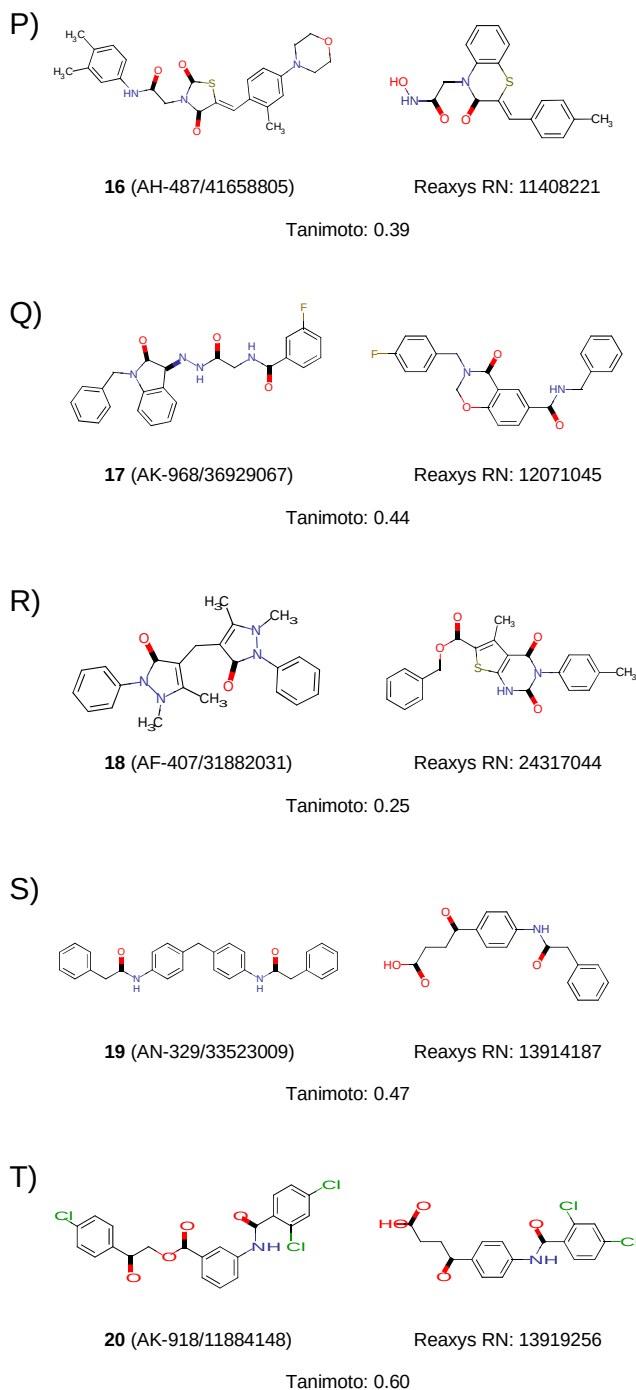
Reaxys RN: 799987

Tanimoto: 0.31

Identification of selective MMP-13 inhibitors by virtual screening



Manuscript 4



Identification of selective MMP-13 inhibitors by virtual screening

Figure S1. In each panel, a hit compound is represented in 2D, together with the most similar active compound found in the Reaxys¹ database, which is labeled with its Reaxys Registry Number. The Tanimoto similarity value resulting from the comparison of the OpenEyePath² fingerprint of both compounds is shown below. MarvinSketch³ was used to draw the structures.

References

- (1) Reaxys <https://www.reaxys.com/>.
- (2) OpenEye scientific software, OEChem, 2013.
- (3) Marvin 16.10.10.0, 2016, ChemAxon <http://www.chemaxon.com>.



UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

In the present thesis we focused on the identification of PTP1B and MMP-13 inhibitors by VS for the treatment of obesity and obesity-associated disorders.

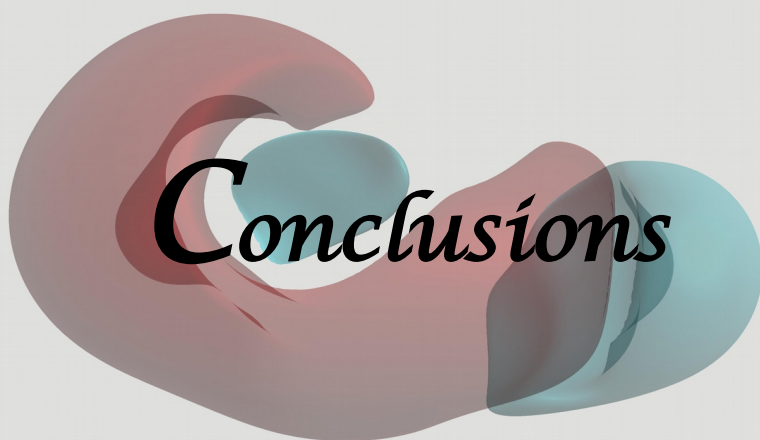
As the inhibition of PTP1B has been shown to have a positive effect on obesity and type 2 diabetes, we developed a virtual screening workflow in order to identify structurally diverse PTP1B inhibitors (**Manuscript 2**). The virtual screening consisted of a combination of ligand- and receptor-based methods. First, a molecular weight filter was applied to discard molecules that were considered either too small or too large to fit on the PTP1B binding site. Secondly, a random forest model based on fingerprints was applied to rapidly rule out the compounds least likely to be active. Next, a constrained protein-ligand docking was conducted in order to discard compounds that would not fit in the binding site and compounds that would not be capable of establishing hydrogen bonds with the P-loop residues, as this particular interaction was performed by all the ligands co-crystallized with PTP1B. Then, the electrostatic potential similarity between docked poses and co-crystallized ligands was analyzed to prioritize compounds with similar electrostatic properties as active PTP1B inhibitors. Finally, the hits obtained were clustered to select a diverse subset of 20 compounds for which PTP1B activity was tested *in vitro*. Of these 20 hit compounds, 15 were found to inhibit PTP1B and 2 of them had IC₅₀ values of 1.4 μ M and 2.1 μ M, the highest bioactivity reported by a PTP1B inhibitor in any VS.

The inhibition of MMP-13 has been postulated to have a positive effect on obesity and osteoarthritis. However, the design of MMP inhibitors that bind to the catalytic zinc ion in the binding site of MMPs has often resulted in the identification of broad-spectrum MMP inhibitors, which have been reported to produce severe side effects in clinical trials. Thus, in order to obtain relevant MMP-13 inhibitors, first we needed to determine how to achieve selectivity for MMP-13. Although selective inhibitors of different MMPs have been successfully identified by targeting the variable S1' pocket adjacent to the catalytic region of the binding site, the characteristics of the S1' pocket that determine inhibitor selectivity are often not described and, in many cases, challenging to identify. Therefore, we performed an analysis of the variability of the S1' pocket among the members of the MMP family in order to search for particular characteristics of each

Summarizing Discussion

MMP that would affect inhibitor activity respect to other MMPs (**Manuscript 3**). This analysis consisted of: **a)** a classification of crystal structures according to their S1' pocket for each MMP; **b)** protein-ligand docking simulations of previously reported selective MMP inhibitors to assess possible steric hindrances with the protein; and **c)** a comparison of the electrostatic properties and the hydrophobicity of MMP S1' pockets and previously reported MMP inhibitors. In the end, we identified concrete regions of the MMP binding site that play a crucial role in inhibitor selectivity, either due to their size, shape, electrostatic potential or hydrophobicity. In the case of MMP-13, the major reason for inhibitor selectivity was the larger size of its S1' pocket respect to other MMPs, which presents an adjacent subpocket referred to as S1'' pocket, not present in other members of the MMP family. In addition, we observed that the presence of a negatively charged group at the region of the MMP-13 inhibitor that occupied the S1' pocket in MMP-13 could prevent its interaction with MMP-8.

Once we had established that targeting the S1'' pocket of MMP-13 was a legitimate approach of achieving MMP-13 selectivity, we designed a virtual screening workflow to identify selective MMP-13 inhibitors in which compounds would aim to occupy this region of the binding site (**Manuscript 4**). This VS consisted of the following steps: First, a molecular weight filter was applied to discard compounds either too small or too large to survive subsequent VS filters. Secondly, a shape-based similarity analysis was performed to discard compounds unable to adopt the characteristic *U* shape observed in co-crystallized ligands that occupy the S1'' pocket of MMP-13. Next, a constrained protein-ligand docking was conducted in order to select the compounds that were able to fit in the MMP-13 binding site while occupying the S1'' pocket. Then, SAR studies of previously reported MMP-13 inhibitors were evaluated, thus extracting a series of protein-ligand interactions that had been shown to contribute to an increase of MMP-13 inhibitory activity. The docked poses obtained were then classified according to the fulfillment of these interactions and, finally, 20 diverse compounds were selected for experimental activity tests. Of these 20 hit compounds, 3 were found to inhibit MMP-13 (IC₅₀ values of 91 μ M, 105 μ M and 15 μ M) and one of them displayed at least 4-fold selectivity against MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14.



UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

The conclusions reached in this thesis are the following:

1. We have developed a new virtual screening procedure to identify structurally diverse PTP1B inhibitors. This VS procedure consists of four steps: (i) a molecular weight filter; (ii) a random forest model based on fingerprints; (iii) a constrained protein-ligand docking and (iv) an electrostatic potential similarity search using previously reported PTP1B inhibitors.
2. Using the above VS procedure we have identified 15 new and diverse PTP1B inhibitors. Two of these compounds, **5** and **7**, have respective IC_{50} values of 1.4 μ M and 2.1 μ M, being the most potent PTP1B inhibitors reported by any VS.
3. MMP inhibitor selectivity can be achieved by taking advantage of the variability of the S1' pocket of MMPs, as the analysis of the differences in size, shape and electrostatic potential of this pocket among the members of the MMP family has allowed us to identify particular characteristics of each MMP binding site that are relevant for inhibitor selectivity.
4. Selective MMP-13 inhibitors can be obtained by targeting its S1'' pocket. The presence of a negatively charged group at the region of the inhibitor that occupies the S1'' pocket of MMP-13 could prevent its interaction with MMP-8.
5. We have developed a new VS procedure to identify selective MMP-13 inhibitors. This VS procedure consists of four steps: (i) a molecular weight filter; (ii) a shape-based similarity search using known MMP-13 selective inhibitors; (iii) a constrained protein-ligand docking and (iv) an inspection of protein-ligand interactions relevant for MMP-13 activity obtained from previously reported SAR studies.
6. Using the above VS procedure we have identified three diverse MMP-13 inhibitors (i.e. compounds **11**, **12**, and **13**; with IC_{50} values of 91 μ M, 105 μ M and 15 μ M, respectively), one of which displayed at least 4-fold selectivity against MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 and MMP-14.

Les conclusions obtingudes en aquesta tesi són les següents:

1. Hem desenvolupat un nou procediment de cribratge virtual per identificar inhibidors de PTP1B estructuralment diversos. Aquest procediment consta de quatre passos: (i) un filtre de pes molecular; (ii) un model *random forest* basat en *fingerprints*; (iii) un *docking* proteïna-l·ligand i (iv) un anàlisi de similitud de potencial electrostàtic amb inhibidors de PTP1B prèviament descrits.
2. Utilitzant el procediment de cribratge virtual anterior hem identificat 15 inhibidors de PTP1B nous i d'estructures diverses. Dos d'aquests compostos, **5** i **7**, tenen valors d' IC_{50} d'1,4 μ M i 2,1 μ M respectivament, sent els inhibidors de PTP1B més potents obtinguts per qualsevol cribratge virtual.
3. La selectivitat dels inhibidors de MMPs es pot aconseguir aprofitant la variabilitat de la butxaca S1', ja que l'anàlisi de les diferències de mida, forma i potencial electrostàtic d'aquesta butxaca entre els membres de la família de les MMP ens ha permès identificar característiques particulars dels llocs d'unió de cada MMP que són rellevants per a la selectivitat dels inhibidors.
4. Es poden obtenir inhibidors selectius per MMP-13 utilitzant la seva butxaca S1'. La presència d'un grup carregat negativament a la regió de l'inhibidor que ocupa la butxaca S1' de MMP-13 pot evitar la seva interacció amb MMP-8.
5. Hem desenvolupat un nou procediment de cribratge virtual per identificar inhibidors selectius per MMP-13. Aquest procediment consta de quatre passos: (i) un filtre de pes molecular; (ii) un anàlisi de similitud d'estructura tridimensional utilitzant inhibidors selectius per MMP-13 coneguts; (iii) un *docking* proteïna-l·ligand i (iv) la inspecció d'interaccions proteïna-l·ligand rellevants per a l'activitat MMP-13 obtingudes a partir d'estudis *SAR*.
6. Utilitzant el procediment de cribratge virtual anterior hem identificat tres inhibidors de MMP-13 d'estructura diversa (**11**, **12** i **13**, amb valors d' IC_{50} de 91 μ M, 105 μ M i 15 μ M), un dels quals té almenys quatre vegades més activitat per MMP-13 que per MMP-1, MMP-2, MMP-8, MMP-9, MMP-12 i MMP-14.



UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

1. **Gimeno A**, Ardid-Ruiz A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Valls C, Aragonès G, Suárez M, Pujadas G, Garcia-Vallvé S. Combined Ligand- and Receptor-Based Virtual Screening Methodology to Identify Structurally Diverse Protein Tyrosine Phosphatase 1B Inhibitors. *ChemMedChem* 13, 1939–1948 (2018).
DOI: 10.1002/cmdc.201800267. PubMed ID: 30024103.
Impact factor 2017: 3.009. Categories: Medicinal Chemistry (Q2), Pharmacology & Pharmacy (Q2).
2. **Gimeno A**, Beltrán-Debón R, Mulero M, Pujadas G, Garcia-Vallvé S. Understanding the variability of the S1' pocket to improve matrix metalloproteinase inhibitor selectivity profiles.
Submitted to *Drug Discovery Today*.
3. **Gimeno A**, Ojeda-Montes MJ, Tomas-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Pujadas G, Garcia-Vallvé S. The light and dark sides of virtual screening: what is there to know?
In preparation
4. **Gimeno A**, Nuti E, Ojeda-Montes MJ, Tomas-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Rossello A, Garcia-Vallvé S, Pujadas G. Identification of selective MMP-13 inhibitors by virtual screening.
In preparation
5. Ojeda-Montes MJ, **Gimeno A**, Tomas-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Valls C, Mulero M, Pujadas G, Garcia-Vallvé S. Activity and selectivity cliffs for DPP-IV inhibitors: Lessons we can learn from SAR studies and their application to virtual screening. *Medicinal Research Reviews* 38, 1874–1915 (2018).
DOI: 10.1002/med.21499. PubMed ID: 29660786.
Impact factor 2017: 8.290. Categories: Medicinal Chemistry (Q1), Pharmacology & Pharmacy (Q1).

Publication List

6. Ojeda-Montes MJ, Casanova-Martí À, **Gimeno A**, Tomás-Hernández S, Cereto-Massagué A, Wolber G, Beltrán-Debón R, Valls C, Mulero M, Pinent M, Pujadas G, Garcia-Vallvé S. A virtual screening strategy to mine large molecular databases to find new leads with low similarity to known actives: application to find new DPP-IV inhibitors.
In preparation
7. Ojeda-Montes MJ, Ardid-Ruiz A, Tomás-Hernández S, **Gimeno A**, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Garcia-Vallvé S, Pujadas G, Valls C. Ephedrine as a lead compound for the development of new DPP-IV inhibitors. *Future Medicinal Chemistry* 9, 2129–2146 (2017).
DOI: 10.4155/fmc-2017-0080. PubMed ID: 29172693.
Impact factor 2017: 3.969. Categories: Medicinal Chemistry (Q1).
8. Tomas-Hernandez S, Garcia-Vallvé S, Pujadas G, Valls C, Ojeda-Montes M, **Gimeno A**, Cereto-Massagué A, Roca-Martinez J, Suárez M, Arola L, Blanco J, Mulero M, Beltran-Debón R. Anti-inflammatory and Proapoptotic Properties of the Natural Compound *o*-Orsellinaldehyde. *Journal of Agricultural and Food Chemistry*. 66, 10952–10963 (2018).
DOI: 10.1021/acs.jafc.8b00782. PubMed ID: 30269491.
Impact factor 2017: 3.412. Categories: Multidisciplinary Agriculture (Q1), Applied Chemistry (Q1), Food Science & Technology (Q1).
9. Bürgler S, **Gimeno A**, Parente-Ribes A, Wang D, Os A, Devereux S, Jebsen P, Bogen B, Tjønnfjord GE, Munthe LA. Chronic lymphocytic leukemia cells express CD38 in response to Th1 cell-derived IFN- γ by a T-bet-dependent mechanism. *Journal of Immunology* 194, 827–35 (2015).
DOI: 10.4049/jimmunol.1401350. PubMed ID: 25505279.
Impact factor 2015: 4.985. Categories: Immunology (Q1).

UNIVERSITAT ROVIRA I VIRGILI
IDENTIFICATION BY VIRTUAL SCREENING OF PROTEIN TYROSINE PHOSPHATASE 1B AND MATRIX
METALLOPROTEINASE 13 INHIBITORS FOR THE TREATMENT OF OBESITY AND OBESITY-ASSOCIATED DISORDERS
Aleix Gimeno Vives

oo

