



Protease inhibitors as therapeutic agents

Laura Mendieta Martínez

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

PhD programme: Biotechnology
Programa de doctorado: Biotecnología

“Protease inhibitors as therapeutic agents”

Laura Mendieta Martínez

Thesis directors:

Dr. Ernest Giralt Lledó
Prof. Universitat de Barcelona

Dra. Teresa Tarragó Clúa
iProteos

INDEX

| | |
|---|-----|
| Introduction | 1 |
| 1. Proteases | 3 |
| 2. Proteases as drug targets | 9 |
| 3. Protease inhibitors as drugs | 12 |
| 4. Case studies | 21 |
| Objectives | 43 |
| Results | 47 |
| Chapter 1: DPP IV labeling methodology and the interaction with its inhibitors | 49 |
| Chapter 1 context | 53 |
| 1.1. DPP IV recombinant expression in insect cells | 55 |
| 1.2. DPP IV characterization | 68 |
| 1.3. DPP IV dynamism | 76 |
| Chapter 1 overview | 95 |
| Chapter 2: Discovery of DPP IV inhibitors from plant extrac. | 97 |
| Chapter 2 context | 101 |
| 2.1. Selection, extraction and test | 103 |
| 2.2. <i>Coutarea latiflora</i> | 114 |
| Chapter 2 overview | 125 |
| Chapter 3: Discovery of POP inhibitors by HTS | 127 |
| Chapter 3 context | 131 |
| 3.1. Set-up fluorescence polarization experiment | 135 |
| 3.2. High throughput screening | 142 |
| 3.3. Hit validation | 151 |
| 3.4. Kinetic studies of lead candidates | 162 |
| Chapter 3 overview | 167 |

| | |
|--|---------|
| Chapter 4: Characterization of peptidyl aryl vinyl sulfones as cathepsins L and B inhibitors..... | 169 |
| Chapter 4 context | 173 |
| 4.1. Library screening..... | 175 |
| 4.2. Elucidation of the inhibition mechanism..... | 178 |
| 4.3. Docking..... | 181 |
| Chapter 4 overview..... | 185 |
| Discussion..... | 187 |
| Conclusions..... | 203 |
| Materials and methods..... | 209 |
| Chapter 1..... | 211 |
| Chapter 2..... | 232 |
| Chapter 3..... | 236 |
| Chapter 4..... | 239 |
| Resumen..... | 243 |
| Bibliography..... | 273 |

INTRODUCTION

The following introduction is divided into four main sections:

- 1: Proteases
- 2: Proteases as drug targets
- 3: Protease inhibitors as drugs
- 4: Case studies: DPP IV, POP and Cathepsins L and B

1. Proteases

Proteases are enzymes that release amino acids, peptides and proteins from larger peptides or proteins.^[1] In humans, there are 933 known and putative proteases (February 2014), and the number increases every year.^[2] First study was in 1836, by Schwann, who described pepsin.^[3] Since then, intense scientific research has been devoted to the protease field, obtaining highly valuable information not only of their function and mechanism of action, but also of their involvement in a wide range of biological pathways. The following sections are intended to obtain a global picture of proteases.

1.1. Classification

Proteolytic enzymes are widely distributed in the human body. They can be located in the extracellular space (approximately the 50%), or in the cellular cytosol, while a relatively small population is intramembranous.^[4] Given the variety of proteases, they are classified upon several parameters.

Depending on the substrate they cleave, proteases can be subdivided into exo-proteases (cleave from the N or C termini) or endo-proteases (cleave an internal peptide bond, far away from any termini).

Based on the catalytic mechanism, proteases are classified into seven subgroups.

Finally, depending on the evolutionary origin, these enzymes are classified into clans and families.

1.1.1. Catalytic mechanism

Depending on the catalytic mechanism, proteases are divided into two main groups: peptidases and lyases.

By definition, peptidases are hydrolases acting on peptide bonds. Members belonging to peptidase group fall into the EC 3.4. Given the nature of the nucleophile involved in the hydrolysis, peptidases are divided into 6 subgroups. The nucleophile can be a residue that forms an acyl intermediate during the reaction (covalent catalysis). These are the amino-terminal nucleophile hydrolases, and are divided into the subgroups of serine, cysteine and threonine proteases. On the contrary, the nucleophile can be an activated water molecule. Depending on the activation of the molecule, proteases are divided into aspartic, metallo and glutamic proteases.^[2]

Until recently, the terms “protease” and “peptidase” have been used as synonyms, but the recent discovery of peptide lyases has imposed the distinction of the concepts. This group was discovered in 2011 by Rawlings *et al.*, when studying the self-cleaving precursor of the Tsh autotransporter from *E.coli*.^[1] The enzyme contains an asparagine residue that forms a five-membered succinimide ring with its own carbonyl carbon. After the correspondent induction, the peptide bond is cleaved. The enzyme was classified as a

lyase, since these enzymes cleave C-C, C-O, C-N and other bonds by means different than hydrolysis or oxidation. Members belonging to lyase group fall into the EC 4. As proteases, there is only one subgroup to date, the asparagine peptide lyases, which has not been included yet in the IUBMB recommendations.^[2]

Thus, there are seven groups of proteases based on the catalytic mechanism, and an eighth one for the unknown mechanism' proteases. In humans, there are representatives of all of them except for glutamic peptidases and asparagine peptide lyases.

1.1.2. Evolutionary origin

Proteases that are homologous in terms of similar amino acid sequence or structure are enclosed into families. All members of a family cleave the peptide bond by the same catalytic mechanism. Families are identified by a letter (catalytic type) and a number. Some families are subdivided, depending on the ancient divergence of the proteases.^[2]

A clan is the group of one or more protease families that share a single evolutionary origin. Similar protein structure or common sequence motifs around the active site are parameters used to enclose families in a clan, but not the catalytic mechanism. Clans are identified by two letters, being the first one related to the mechanism of action. Letter "P" designs those clans where mixed mechanisms are present. As well as for families, clans can be subdivided into subclans.^[2]

1.2. Protease substrates

For a peptide or protein to be considered as an endogenous protease substrate, it has to fulfil three requisites. First, it must contain a recognition sequence for a given protease, and second, the protease must cleave the substrate candidate in an *in vitro* assay. However, the peptide or protein would not be considered as a substrate until it would be found simultaneously and in the same cellular compartment or extracellular location with the active protease^[5]. It is not only important the presence of the proteolytic enzyme, but also its ability to cleave the substrate.

Nature has been economic, and one protease cleaves more than one substrate. The subset of a protease's physiological substrate is defined as the protease degradome.

1.2.1. Identification of protease substrates

Identification of the endogenous substrates of a protease is not an easy task. Time-defined events, low concentrations or fast processing of a proteolytic product may difficult its identification in an *in vitro* assay.

Traditionally, the used methodology has been the bottom-up technique: given a substrate, study the subset of proteases in order to determine which enzyme is the responsible for its processing. The majority of proteases' substrates have been identified by this method.

For example, substrates of angiotensin-converting enzyme (ACE) were discovered by the bottom-up procedure.^[5, 6]

Recently, *in silico* bioinformatics tools have helped in the substrate validation. The substrate specificities of proteases are determined, and specific peptide sequences are obtained. Then, these sequences compared to the whole genome in order to find the containing peptide or protein. However, the calculated affinities for the small artificial substrates may substantially differ to the peptide or protein substrates. For that reason, *in silico* tools are complemented with a second method for substrate validation. One remarkable example of the application of bioinformatics tools is the identification of caspase substrates.^[5, 7]

Following the same strategy, the yeast-two-hybrid method allowed the identification of endogenous substrates based on peptide affinities.^[5] The main example would be the identification of the matrix metalloproteinase-2 substrate (the monocyte chemoattractant protein 3 -MCP3).^[8]

Genetic approaches, such as knockin, knockout models or RNA interference (RNAi), may present some disadvantages, such as knockout embryonic lethality and upregulation of compensatory protease functions.^[5] Nevertheless, they have been proved as a powerful technique in the protease degradome field. β -site APP-cleaving enzyme (BACE) was validated as a target for Alzheimer's disease, thanks to the confirmation, by genetic tools, that BACE processed the β -amyloid peptide in mouse.^[9]

Finally, proteomic tools, such as mass spectrometry (MS) have enabled the identification of protease substrates that were not possible with the previous methodologies. Proteomic methods are expected to be extremely powerful in substrate determination in the next few years.^[3]

Regarding the two last techniques, genetic approaches and proteomic tools, additional precautions should be taken. In these techniques, there is no evidence that the protease of study is responsible for the direct proteolysis of the substrate, but that it is somehow involved in its processing. For that reason, a peptide or protein is only considered a protease substrate when the cleavage is proved *in vitro*. But instead of being a disadvantage, genetic and proteomic tools have favoured the understanding of the protease mechanism of action.

1.3. Functions

Depending on the goal of the processing, proteases have two main functions: protein degradation, which affects partially or totally unfolded proteins, and protein processing which consists on cleavage of properly folded proteins.

Regarding protein degradation, proteases are kept in organelles, such as lysosomes or endosomes, or are integrated in the proteasome system.^[5] The compartmentalization

avoids the uncontrolled protein degradation, given the fact that these proteases are non-selective.

Examples of protein degradation include protein recycling and the antigen processing for their presentation in MHC I (started by the action of the proteasome) and MHC II molecules (endosomal and lysosomal proteases).^[4]

On the contrary, proteases in charge of protein processing are highly selective. In first term, the protease activates, inactivates or alters the function of a protein. However, the substrate or the product of a proteolytic event may form part of a biological pathway.^[5] Thus, final function of proteases is defined by the roles of their substrates or products.

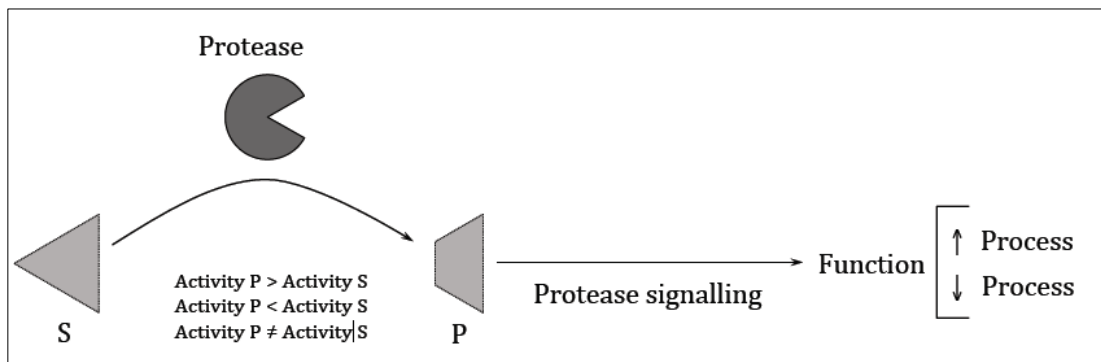


Figure i.1.: Function of a protease. The direct function is to alter the activity of a given substrate. However, the final action depends upon the participation of the substrate or product in a biological pathway.

Proteolytic enzymes are then able to regulate specific biological pathways, activating or inhibiting the route upon the appropriate stimuli. The ability of proteases for the regulation of biological processes has been named as protease signalling.^[5] As in other non-proteolytic pathways, the signal is amplified by the consecutive steps, which can be regulated. In contrast, the protease signalling is irreversible, and thus, once started products cannot revert to substrates.

It has to be mentioned that antigen presentation, carried out by protein degradation, can be classified as well as protease signalling event. However, in this text, the term protease signalling will refer only to protein processing.

1.4. Mechanism of protease signalling

For proteases is important to define what they do (their ultimate function/s) but it is crucial to elucidate how they perform their activity. This means to describe the protease signalling pathway.

A single proteolytic event may be implicated in the pathway: the cleavage of a protein substrate promotes a specific effect. For example, the protease ADAM10 (ADAM stands for a disintegrin and metalloproteinase) activates the receptor HER2 (receptor tyrosine-

protein kinase erB-2). After this event, the receptor dimerizes, provoking an autophosphorylation that initiates several signalling pathways, such as the MAPK route (mitogen-activated protein kinases – cell proliferation) among others.

Other proteases participate in consecutive reactions where the substrate is a non-protease protein.^[5] One example would be the subsequent processing of APP (amyloid precursor protein) by α - β - and γ -secretases.^[10]

The third mechanism of signal transduction is the cascade. These pathways consist on consecutive proteolytic events where the substrates are proteases.^[5] A zymogen (inactive protease) is activated by a second protease. Once active, it can activate the zymogen of a third protease, and so on. In some cases, not all the substrates are proteases, and it is difficult to decide if the term cascade is appropriate. For that reason, the more complex pathways are called protease networks.^[5] The best studied cascade is the blood-coagulation process, which was first described by MacFarlane in 1964,^[11] where consecutive serine proteases activate zymogens. The last protease is thrombin that cleaves fibrinogen into fibrin, allowing its aggregation into the blood clot. ^[3]

Apart from defining the players of a protease signalling event, full understanding of the mechanism involves the description of the relationship among all proteins. Despite proteolytic events are independent, proteases may perform their activity either alone, or in complex with other proteins (proteases or not). Protein-Protein Interactions (PPI) are crucial in some processes, such as the proteasome-protein degradation.

1.5. Regulation

Proteases are involved in key biological processes. From immune response to blood-coagulation, there are examples of protease participation in almost every physiological event. Because of the importance of the pathways where proteases take part, and the vast coverage in a multitude of routes, proteolytic enzymes are tightly regulated at different levels.

At the lowest level, transcription of protease genes is regulated by nuclear receptors. This regulation allows the spatial and temporal expression.

Respect to the location, ubiquitous proteases are found in every cell type (such as calpain-1), whereas tissue-specialized proteolytic enzymes present a restrictive expression pattern. These proteases have specific roles that are carried out only in the cell type where they are expressed. For example, cytosolic carboxypeptidase 4 is a metalloprotease expressed exclusively in the corneal endothelium. ^[12]

Regarding to the temporal expression, transcription can be constitutive or proceed after certain stimuli. Constitutively expressed proteases are the house-keeping enzymes, and proteasome is the most characteristic example. On the contrary, the expression of Lon protease, a mitochondrial serine protease involved in degradation of oxidized proteins, is induced under conditions of acute stress.^[13]

At the protein level, proteases may be active or be expressed as zymogens. Thus, subsequent activation, either autocatalytic or performed by other proteases, is required. Besides, there are proteases that still need the binding of specific cofactors in order to be fully active, some require post-translational modifications and other are only active in complex formation, such as the proteasome.

A particular case of regulation at the protein level is the presence of exosites (secondary binding sites). These areas are located apart from the active site of the protease and play important roles in substrate recognition.^[14] The exosites are key in the substrate affinity and specificity for some enzymes. It has been suggested that the interactions between exosites and substrate are important for the determination of the affinity^[14], like a recognition-docking space for the substrate. Then, after this event, binding of the substrate with the active site would promote the catalytic event. Thus, exosite interactions influence the catalysis rates.^[3]

The coagulation blood cascade is a perfect example of protease-containing exosites. Extrinsic Xase complex,^[15] Factor XIa,^[14] prothrombinase^[16] and thrombin^[17] are enzymes which exosite existence has been demonstrated.

Finally, there is a last step of regulation that avoids the proteolysis: inhibition.^[5] For most proteases, it exists a protein or peptide inhibitor capable of stopping the proteolytic activity. Therefore, this last step of regulation allows a fast switch-off of the signalling pathway, if the enzyme is involved in any. Enzymes and inhibitors (as well as the substrates) have co-evolved to accomplish a successful signalling route. Since nature is economical, these inhibitors affect several proteases at the same time, typically of the same family, and thus have low specificity.^[4]

There are two types of endogenous inhibitors that differ in the concentration and the kinetics of the inhibition.^[5] The emergency inhibitors are located in cellular compartments that avoid the contact with the protease. When required, are liberated to inhibit the enzyme in a fast manner. On the contrary, regulatory inhibitors are co-localized with the protease and perform a fine-tuning of the proteolytic event.

An important aspect of inhibitors is the definition of their binding pocket. The great majority of them occupy the active site or an area close to it, thus avoiding the binding of the substrate. However, other inhibitors bind to exosites or allosteric hotspots. While the terms "exosite" and "allosteric site" have been used indistinctly, exosites are required for activity or for substrate recognition, while allosteric areas are not involved in the proteolytic event. Nonetheless, binding of ligands to allosteric sites avoids (or enhances) the proper function of the enzyme.

The typical example of endogenous protease inhibitors is serpins. The serpin family is formed by 48 members,^[12] and are mainly serine protease inhibitors.

As mentioned previously, protease signalling involves irreversible reactions, in the sense that the product cannot reverse into substrate (except for cathepsin S under change of pH). But the fact that the reactions are irreversible does not imply that once triggered the pathway all steps were performed, like a domino effect. The several regulation points, at different stages, grant the possibility of stopping a particular event if required.

Thus, regulation assures the correct performance of the reaction or network, and allows fast changes upon stimuli, shifting the balance.

2. Proteases as drug targets

Proteases participate in a multitude of important biological processes and their action is tightly regulated. Failure in their activity or misregulation leads to aberrant signalling pathways and therefore, proteolytic enzymes are involved in a large number of diseases. For that reason, proteases are important drug targets in a wide range of diseases, such as Alzheimer's disease, cancer, cardiovascular disorders and HIV infection among others. The following sections are intended to describe the selection of a protease as a drug target.

2.1. Protease target identification

Given a disease, selection of the appropriate protease target may lead to difficulties. A thorough analysis of the pathology will provide information about the altered processes that are undergoing.

Inappropriate proteolysis plays an important role in several diseases, such as cancer and neurodegenerative diseases.

The proteolytic activity can be diminished or increased. Both can be result of genetic aberrations or altered regulations (insufficient activation or excessive inhibition for diminished proteolysis and excessive activation of insufficient inhibition for the increased enzymatic activity). For example, expression of the cysteine proteases cathepsins L and B is increased in tumor cells.

In other pathologies, complex signalling routes are altered, resulting in an undesirable effect. The protease may participate in the unbalanced pathway and tuning of its activity restores the proper running of the pathway. It can also happen that the enzyme and the cause of the disease are not related, but modification of the protease activity results in a beneficial effect. An example for this last scenario is the inhibition of ACE for the hypertension treatment. ACE cleaves angiotensin I into angiotensin II, which is a vasoconstrictor. By avoiding vessel constriction, blood pressure is lowered. Thus, ACE inhibitors restore normal blood circulation, while ACE itself is not altered in the disease.

Finally, infectious agents such as parasites, virus or bacteria, have their own proteases, which can be targeted to avoid the proliferation of these agents. The HIV-1 protease has been largely investigated as a drug target for autoimmune deficiency syndrome (AIDS).^[5]

2.2. Protease target validation

A validated target is an entity critically involved in a disease and whose modulation offers a therapeutic effect.^[18] In the case of proteins, the target should be expressed in cells linked to the disease pathology. Specifically, overexpression is a key marker in validated targets. Furthermore, a beneficial effect has to be observed in animal models of the disease after treatment with drugs against the target or in knock-out transgenic animals.

Target validation is then the term used for the recognition of the crucial role of the protein in the disease.

Historically, drug discovery was only pursued against validated targets. However, validation is time-consuming (requires typically more than one year, resulting in an extremely large step) and does not assure a successful drug discovery process.

This last comment is connected to the fact that validated targets are not necessarily druggable. For example, a large number of protease-validated targets perform their activity through protein-protein interactions (PPI). But there is no marketed inhibitor able to disrupt PPI, because its discovery is a challenge that requires a huge economic investment. Hopefully, future screening techniques and the design of novel molecules with unimaginable properties, would transform the PPI inhibitor drug discovery from challenge to a routinely process.

Besides, it is believed that the druggable validated targets have already been covered.

Thus, the drug discovery process for the current targets is extremely difficult, causing high attrition rates of drugs at discovery or developmental stages. This fact has arisen discrepancies about the usefulness of target validation and is driving some pharmaceutical companies to move to alternative strategies, suggesting that target validation is not a prerequisite for drug discovery. However, the majority of pharmaceutical industries still work on validated targets.

2.3. Study of the target

A given protease has to be fully characterized before selecting it as a drug target. It implies to have a full knowledge of the substrates, mechanism of protease signalling (if any) and its regulation. Insights on these aspects help in target selection and in the drug design as well.

In signalling pathways, early stage inhibitions afford a more effective blocking of the final effect/s. If the proteolytic enzyme is at the end of the cascade, the inactivator concentration must be higher due to the signal amplification effect. Thus, it may compromise the safety of the treatment.

Another important aspect in signalling routes that is of crucial study before drug discovery is the existence of related proteases. Proteases participating in a same route may have similar active sites, and thus exogenous inhibitors might alter not only the target, but also the other proteolytic enzymes. Knowing beforehand which proteases can be affected is critical for the finding of selective inhibitors.

Regarding regulation, the presence of exosites and allosteric sites may represent an advantage in the selection of proteases as drug targets. These sites imply extra protein surfaces that can be object of ligand binding and that offer, normally, better selectivity in front of other proteolytic enzymes.

In general, inhibition of targets is easier than activation, and for that reason, drug discovery in protease field normally refers to inhibitors. However, activation of some enzymes can be achieved by other means than activation. For example, the activity of a proteolytic enzyme can be increased by inhibition of the endogenous inhibitors.^[5] Then, analysis of the pros and cons of a given protease as a drug target helps in the selection; in the way that target identification and study are iterative processes.

3. Protease inhibitors as drugs

The implication of proteases in a multitude of diseases makes them valuable drug targets. In fact, around 5-10% of the pharmaceutical targets are proteolytic enzymes^[3] and there are several successful protease inhibitors in the market. Inactivators of thrombin and factor Xa, designed for the treatment of coagulopathies, have reached US\$3 billion sales, and ACE inhibitors, against hypertension, US\$6 billion.^[3] In 2013 the DPP IV inhibitor Januvia, for type 2 diabetes treatment, raised the sales generating US\$3 billion in total since it was launched in 2006. Thus, proteases inhibitors market is very attractive for pharmaceutical companies and is an area of growing investment.

The following sections are intended for the description of the drug discovery process as well as the definition of the ideal protease inhibitor. Finally, their use is represented by some successful examples.

3.1. Drug discovery strategies: obtaining a hit

The objective of inhibitors is binding, in a specific and potent manner, to their target, ablating the activity. In the case of proteases, inactivators block the entrance or processing of the substrate.

As well as for other proteins, drug discovery in proteolytic enzymes can be pursued by a vast group of strategies. For some of them, there is a requirement of knowledge of the protease's three-dimensional structure, while for other strategies, it is not mandatory. However, although not being essential, previous information about the binding domain to be targeted is valuable for drug discovery and refinement.

3.1.1. Structure-based drug discovery

Classic inhibitors are substrate peptide derivatives, and so, their development does not require structure elucidation of the protein. However, the target three-dimensional structure can be used in the design of non-peptide derivative compounds. As a result, more potent and stable compounds are obtained. Inhibitors of renin and DPP IV have been obtained following this strategy. Combinatorial chemistry is a useful tool in order to explore several substituents preferences in the protein subsite.

With the advent of new technologies, virtual screening (VS) has appeared as an *in silico* alternative. It pursues the discovery of new ligands with the use of computer-based methods and previous information of protein structures. There are some successful cases in the bibliography of the VS application in the protease inhibitor search. For example, inactivators of the serine protease of dengue virus^[19, 20], or the aspartic acid β -secretase^[21] among many others. Furthermore, VS is a powerful technique in order to bring drugs to market,^[22] but it has to be combined with other strategies in order to confirm the hits, due to its intrinsic weaknesses. As an *in silico* technique, it deals with docking and scoring

problems. The calculation of the binding affinities, which depends on solvation energies, is not still accurate despite the efforts in the field.^[23, 24] Besides, proteins are flexible entities and suffer modifications after the inhibitor binding. Unfortunately, VS fails in the calculation of the conformational changes.^[23, 25]

3.1.2. Structure-independent drug discovery

The classic drug discovery strategy of protease inhibitors is based on previous information of substrate/s. First-generation inactivators consists on substrate derivatives. Therefore, they are peptide molecules with low metabolic stability. Then, peptidomimetics gain in biophysical properties, such as stability. However, optimal conditions are difficult to achieve for some targets. In these scenarios, structural knowledge may be necessary in order to guide the selection of molecule substituents.

High-throughput Screening (HTS) is a technique that allows testing of thousands of compounds against a target in a fast way due to automated platforms.^[26] HTS is applicable to a variety of targets, with the only requirement of a screening technique for the identification of the hits, such as Fluorescence Polarization.

HTS was envisaged in the 90's as a promising tool for drug discovery, but results were not as expected. Among the causes for failures, compound libraries quality was detected as the principal one. Libraries, despite being large, do not cover the entire chemical space, which is estimated to be around 1×10^{40} to 1×10^{128} molecules.^[3] Besides, properties of the compounds are, in general, not optimized, and then molecules have low bioavailability. Thus, small-focused libraries were proposed as a more efficient tool for the application of HTS in drug discovery.^[5] Improved libraries are now being used and fruits of these changes will be gathered in the coming years. As an example, inhibitors of methionine aminopeptidase 1b from *Plasmodium falciparum* have been successfully obtained by HTS.^[27] This technique also offers the possibility of in-cell testing of compounds. In this regard, instead of emphasizing the search of potent compounds, other properties, as toxicity and cell permeability are analysed as well. As a result, hits are more likely to be selective drugs with good biophysical properties and minimum side effects.^[28-30]

Fragment-based screening oriented for the finding of small molecules that interact with the target. Due to the reduced molecular weight, interactions are weak, usually in the milimolar or high micromolar range. Fragments are identified with techniques as nuclear magnetic resonance (NMR), mass spectrometry (MS), surface-plasmon resonance (SPR) or x-ray crystallography, which also allow the identification of the binding area of small molecules (except SPR). The technique relies in the identification of at least two binding sites in the target, close in space. Fragments binding to different hot-pockets are then chemically linked in order to obtain a more potent ligand, with affinities that can reach the nanomolar range.^[3]

For some scientists, fragment-based screening is a structure-based drug discovery method, while for others it is not, but in fact, it can be both. Depending on the availability, or not, of the three-dimensional structure of the protein, the identification technique is chosen. While X-ray crystallography implies previous structure elucidation, NMR and MS can be used in both scenarios. Thus, fragment-based drug screening is a versatile tool for

drug discovery and there are several examples in the bibliography. Abbot Laboratories used this technique in combination with NMR for the finding of metalloproteinase-3 (MMP3) inhibitors. A potent compound, ABT-518 was found to inhibit the enzyme with an IC50 value of 12 nM.^[31] It has reached phase I but no further development is reported, thus actual state is unknown.^[32]

Similarly to virtual screening, there is computational-based strategy that can be applied without target structural knowledge. It consists on the screening of molecules against similarity versus known ligands or a pharmacophore.^[33] As an example, inhibitors of the hepatitis C virus protease have been obtained through this approach.^[34, 35]

Finally, even if the target is unknown, drug discovery is possible by phenotype screening. Testing of compounds directly into animal models or *in vitro* cell cultures has been applied since the beginning of medicinal chemistry and still nowadays is a source for new drugs.

Cell-based assays have gained popularity in the discovery of new chemical entities.^[36] This strategy consists on testing small molecules in disease-related cultures, such as tumor cells. If a beneficial effect is detected, a set of biochemical and computational methods are used for target identification and study of its mechanism of action.

Phenotype screenings require longer time periods than target-based approaches, because target identification has to be accomplished during the process. Sometimes, target determination is impossible, mainly because of specificity derived issues. However, their application is common in some fields, as medicinal extracts tests. Inhibitors of hepatitis C virus protease^[37] and matrix metalloproteases 9 and 12^[38] have been found by indirect drug discovery, though specificity of these molecules has not been reported.

3.2. In the quest for the perfect drug: obtaining a lead

An inhibitor obtained by the classic approach must be not only active, but also it has to fulfil requirements such as selectivity and low toxicity. The balance between potency, selectivity and bioavailability is difficult to achieve, and sometimes only it is possible by sacrificing the efficacy. Thus, there is no rule to follow in the designing of an ideal drug, but some guidelines may be helpful.

3.2.1. Inhibitor design

Regarding to the inhibitor size, small molecules are preferred by the medicinal chemists due to their biophysical properties. Large inhibitors may have difficulties in entering the cells, but are suitable for extracellular or secreted proteases. However, they are restricted by the administration route.

First approach of large inhibitors was based on analogs of endogenous inhibitors, such as the use of serpins as trypsin inhibitors for the treatment of cystic fibrosis.^[39] Protein and peptide derived molecules are also described in the literature.^[3] For example, hirudin, a thrombin inhibitor, and its derivatives were designed as anticoagulants.^[5] However, biological agents, in general, may cause immunogenicity. Other large molecules used in

drug discovery are monoclonal antibodies, which are in preclinical and clinical development.^[5] Antibodies offer high specificity, therefore are useful in targeting proteases without modifying the activity of the closely related enzymes. Nevertheless, monoclonal antibodies are expensive and still have the main hurdle of large molecules (administration route is restricted).

Respect to the binding kinetics, reversible inhibitors represents better choice than irreversible molecules. One disadvantage of the latter is that covalently modified proteins may cause antigenicity. However, the main drawback is specificity. A reversible inhibitor might affect other proteases, but only partially, depending on the affinity respect to the off-targets. On the contrary, an irreversible inhibitor will block the activity of all the proteases that it binds to, in a permanent manner, owing to important side effects.

Thus, irreversible inhibitors are only recommended if the selectivity is high enough, are fast-cleared and for their restrictive usage in acute situations and life-threatening diseases.^[5] Apart from its clinical use, irreversible inhibitors are vastly applied in basic research given its ability to label targets and the reliability of its synthesis, which can be easily achieved by attaching a warhead that recognises the active-site residue.

It has to be pointed out that not all covalent inhibitors are irreversible, since some bonds are kinetically reversible.^[3]

With regard to the catalytic mechanism, inhibitors can be competitive or not. Mainly all inhibitors in the market are competitive because traditionally they were preferred. Besides, there is an ensemble of techniques for the finding of this type of inactivators, which makes its discovery easier. However, competitive inhibitors might present selectivity issues. Specificity is crucial in all types of drug treatment. As mentioned before, in proteases the finding of a selective inhibitor may be more difficult than in other targets given the similarity of the active site of related proteolytic enzymes. Cross-reactivity occurs in a small percentage for all inhibitors, but the lower the percentage, the better. Moreover, depending on the roles of the off-targets, consequences will not have the same impact. Increase of the selectivity of a molecule can be achieved by structure-activity studies (SAR). Structural differences in the binding site between the target and similar proteases are profitable for the design of specific inhibitors, which will yield fewer side effects.

Given the selectivity disadvantages of competitive molecules, there is an actual tendency for the use of non-competitive inhibitors. These molecules bind to allosteric sites or to exosites, thus avoiding the entrance or the processing of the substrate in the active site.^[5] While active sites of related proteases are similar, the non-competitive areas present substantial differences. For that reason, selectivity of non-competitive inhibitors is generally higher. Besides, since these inhibitors are not competing with the substrate, the concentration that yields inhibition is lower. Thus, non-competitive inhibitors are potentially more selective and less toxic.

3.2.2. Inhibitor physiological properties

Good bioavailability of an inhibitor implies to have acceptable ADMET properties (absorption, distribution, metabolism, excretion and toxicity). In order to assess them, a series of *in vitro* assays are available.

Oral administration is the preferred route for a drug treatment. However, some drugs have stomach-stability issues or are unable to cross the intestinal wall. Caco-2 cell assays are a reliable test used for the determination of potential intestine permeation. Besides, solubility and stability assays (in different medias, depending on the selected administration route) are recommended.

Inhibitors, despite being active and selective, may fail if they do not reach their target in the organism.

In the case of intracellular targets, inhibitors must have cell-penetrating properties. While some molecules are already able to enter into cells, some require modifications. Cell-penetrating peptides (CPP) are effective tools for intracellular delivery.^[40] CPP are versatile molecules that can be chemically modified and tailored: their resistance to proteases is increased by the use of D-amino acids, and specific sequences and/or substituents are used depending on the selected internalization process.

The D-amino acid derivative of sweet arrow peptide (SAP), a proline-rich peptide, has been described as a potent CPP with protease resistance.^[41, 42] Besides, further addition of hydrophobic substituents, improved its cell penetration efficiency, making SAP derivatives a promising tool for the internalization of drug cargoes.^[43] Nevertheless, there is no protease inhibitor coupled to CPP in development stage. This may be because they are expensive to produce and are not suitable for oral administration.

Cellular membranes are not the unique barriers in the organism. In order to protect the brain from potential damaging entities, the central nervous system is safeguarded by the blood-brain barrier (BBB). It is an extremely restrictive barrier where crossing of molecules is tightly regulated. This physiological constraint is vital for the proper running of the brain, but it also avoids the entrance of medicinal drugs. It has been estimated that only the 2% of potential neurotherapeutics are able to cross the BBB.^[44] Therefore, a great percentage of drugs require shuttles to cross it. Diketopiperazines (DKP) are organic molecules that result from a side reaction in peptide synthesis, when proline is at the first or second position from the resin (C terminal). Despite being unwanted products, DKP are described as tools for the transport of cargoes across the BBB through passive diffusion.^[45] Peptides have also been exploited as BBB shuttles. A successful strategy is the receptor-mediated transport. It consists on the use of peptides that selectively target receptors of the BBB endothelial cells, such as the transferrin receptor. These shuttles have been proved to carry even large cargoes, as gold nanoparticles, through the BBB.^[46] An alternative to shuttles is the use of prodrugs that once in the central nervous system are processed to afford the active drug. An example is baicalein, the aglycone moiety of baicalin, which potentially crosses the BBB by passive diffusion.^[47]

Regarding drug metabolism, it is carried out by the family of enzymes under the name of cytochrome P450, named as CYP, and located principally in the liver. Testing the processing of the inhibitor by these enzymes affords information with regard to the

products of its metabolism, and possible toxicity. Besides, not all the CYP have the same affinity for the exogenous molecules, and thus, the specific CYP that metabolizes the drug can be defined. Analysing the CYP and comparing it with CYP of other molecules, drug-drug interactions can be established.

Excretion of drug and its metabolites is usually through the kidneys. Inappropriate excretion may lead to toxicity issues.

With respect to toxicity, hepatotoxicity of drugs and/or metabolites is assayed in cell-based tests.^[48] Alternately, toxicity can be analysed by measuring cell viability *in vitro*, for example by an MTT assay.

Finally, after *in vitro* tests, inhibitors are assayed in animal models, where potency, toxicity and bioavailability are carefully analysed. It has to be pointed out that these models do not resemble humans in a substantial number of biological processes, and thus, drugs that have depicted acceptable efficacy and physiological properties in animal models may not show the same result in humans. It depends on the translational degree of the animal model with respect to humans: for some diseases models resemble the human pathology in a higher degree than for others. In those maladies where the model is not appropriate, lack of reproducibility is one point where many drugs are discarded.

3.3. Protease inhibitors as drugs (examples)

In order to highlight the importance of proteolytic enzymes inactivators in the pharmaceutical market, two examples are selected.

3.3.1. ACE inhibitors

The angiotensin-converting enzyme (ACE, peptidyl dipeptidase A) is a membrane-bound zinc metalloprotease of 150-180 KDa (somatic form).

It cleaves the C-terminal dipeptide of substrates when the last residue is not a glutamic or an aspartic and the penultimate position is not occupied by a proline. One of the ACE substrates is angiotensin I, which is processed to angiotensin II. This molecule is responsible, among other functions, for the vasoconstriction and increase of blood pressure. Therefore, ACE is a validated target in cardiovascular conditions such as hypertension, heart failure and heart attack.^[5]

ACE crystallographic structure was not solved until 2003^[49], then first ACE inhibitors, in the 80's, were developed without knowledge about the target architecture. Thus, molecule design was based on the carboxypeptidase A structure, which was believed to be structurally related to ACE. Inhibitors were small, contained a metal-chelating group and lacked of selectivity, producing side effects such as angioedema.^[5]

Years later, crystallographic elucidation revealed that ACE structure presented significant differences to carboxypeptidase A, thus explaining the failure in selectivity of the first-

generation inhibitors. ACE is formed by two catalytic domains that display different affinities for substrates, and therefore, for inhibitors. The C-terminal domain is involved in the processing of angiotensin I, while the N-terminal domain is in charge of haemoregulation processes. Based on this finding, a second generation of ACE inhibitors is being investigated.

Currently there are 37 ACE inhibitors in the market, being captopril the first one in 1980. However the number will increase in the coming years by the discovery of new structure-designed ACE inhibitors.

3.3.2. HIV protease inhibitors

The aspartic protease of HIV-1 is a homodimeric enzyme of 10.8 KDa. It is in charge of the processing of viral gag and gag-pol polyprotein precursors. Thus, its inhibition results in the formation of immature and non-infectious virions. For that reason HIV-1 protease is a validated target for the treatment of AIDS.

The crystallographic structure elucidation supposed a breakthrough in the drug discovery process, and allowed a fast reach to the market of the inhibitors. The protein was discovered in 1985^[50] and 10 years were enough for inhibitors to reach the market.^[32, 51]

Nowadays, there are 9 HIV-1 protease inhibitors in the market.^[32] Pharmaceutical companies and brand names are in brackets.

In 1995, Saquinavir (Roche; Fortovase) appeared in the market as the first HIV-1 protease inhibitor.

In 1996, Merck launched Indinavir under the commercial name of Crixivan. It is currently in phase II for the therapy of Kaposi's sarcoma by the Istituto Superiore di Sanita.

Same year, Ritonavir appeared (Abbot, Norvir); and in 1997, Nelfinavir (Agouron-Pfizer/Roche; Viracept) was released. This last drug was withdrawn from the European market in 2007 due to an impurity (ethyl methansulphonate). However, the license was re-instated the same year. Besides its application in the AIDS treatment, its use is being studied in other conditions such as severe acute respiratory syndrome (SARS), systemic lupus erythematosus or pancreatic cancer among many.

In 2000, Abbot launched Lopinavir (Aluviran) and three years later Atazanavir (Novartis, Reyataz) reached the market for AIDS treatment. The Brigham and Women's hospital is conducting phase II experiments for its use in diabetic patients.

Also in 2003 another HIV-1 protease inhibitor was launched. Fosamprenavir (GlaxoSmithKline; Lexiva, Telzir) contains two reactive moieties, a sulphone and a phosphonate, that were not described previously. Fosamprenavir is the prodrug of Amprenavir. Curiously, amprenavir was an approved drug (Kissei; Agenerase, Prozei), but in 2005 the pharmaceutical company decided to withdraw it from the market of Japan, and later, in 2010, from Europe.

The sulfone moiety was also used in subsequent molecules tipranavir and darunavir.

The first one was launched in 2005 under the name of Aptivus (Pfizer and Boehringer Ingelheim). It is a non-peptidic-derived from a screening programme.

The second one, darunavir, appeared one year later (Janssen; Prezista).

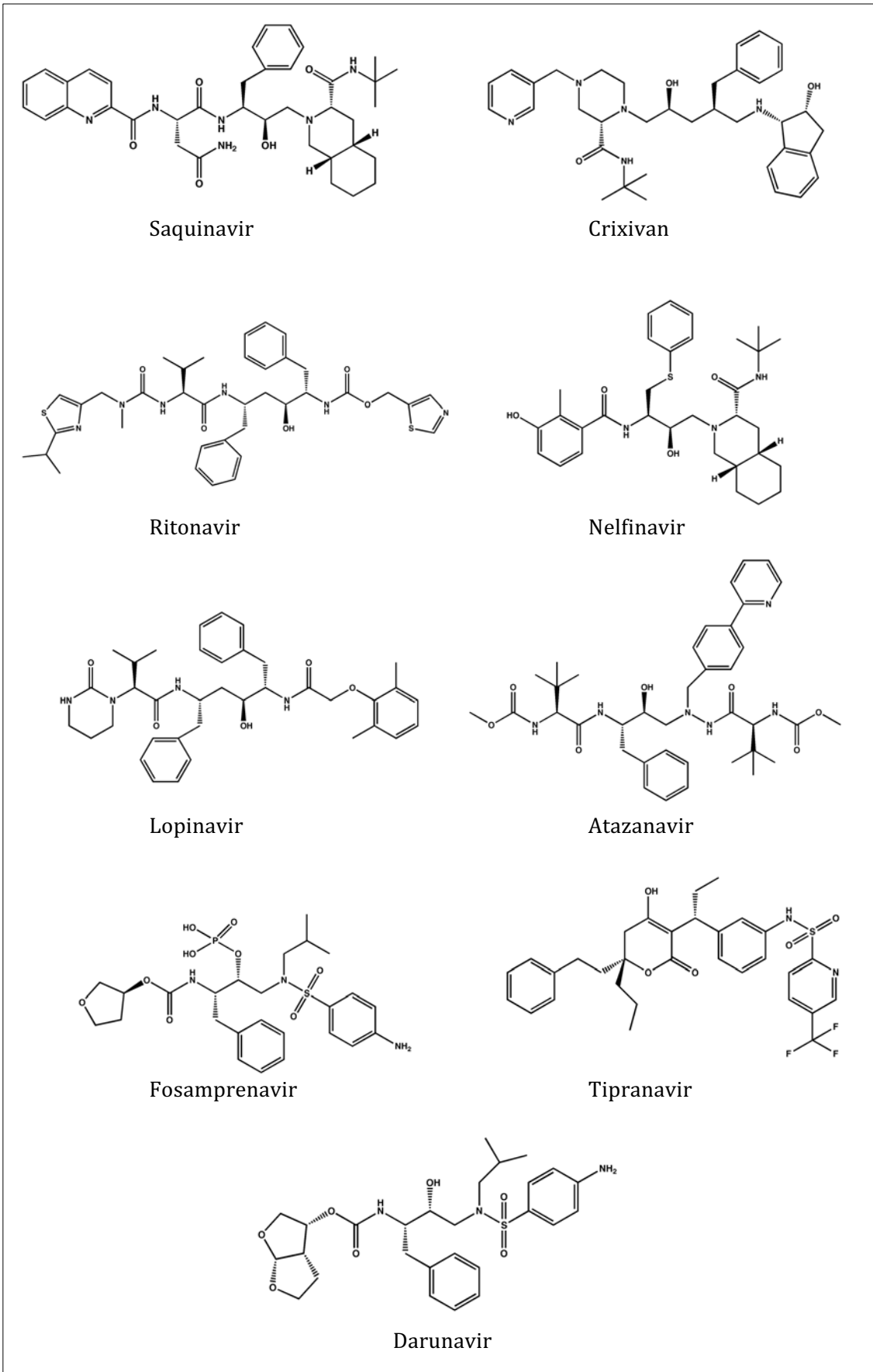


Figure i.2.: HIV-1 protease marketed inhibitors

3.4. Protease inhibitors for target validation

Apart from the use of protease inhibitors in the treatment of certain diseases, they can be used for target validation. However, since the work covered by the thesis is not related to this field, this application is briefly pointed out.

Despite the great advance in science during the 20th century, the complete understanding of the biological processes is far away. For some proteases too little is known about its physiological substrates, mechanism or regulation. In these cases, the use of inhibitors is a powerful tool for defining the function of the enzyme. They can also help in the target validation process.^[5]

This strategy has not to be confused with the phenotype screening, where the target was unknown. In this case, the protein is defined but its role in the disease is not fully characterized. Inhibitors can thus be obtained by a multitude of strategies, not only phenotype screening.

Protease inhibitors are useful tools for target validation, and it is a technique that is being followed by pharmaceuticals and laboratories around the world. Instead of following the traditional scheme of target validation followed by drug discovery, the processes are parallelized. By this, time and costs are reduced, and less time-to-market implies higher revenues.

The only disadvantage of the application of inhibitors in target validation is selectivity of the molecules.^[4] If the specificity is low, not only the target will be modulated, but also the closely related proteins. Thus, the relationship between protein and disease may be compromised.

4. Case studies

4.1. Dipeptidyl peptidase IV (DPP IV)

Chapters one and two of results are devoted to this protease.

4.1.1. The protein

Dipeptidyl Peptidase IV (DPP IV, CD26, EC 3.4.14.5) is a serine exopeptidase that belongs to subfamily B of the family S9 (clan SC). It processes substrates containing a N-terminal X-Pro or X-Ala. DPP IV cleaves after C terminus of P1, when this amino acid is a proline or an alanine, the N terminal is free and the P1' position is not a proline.^[52]

4.1.1.1. Structure

DPP IV is a 220 KDa homodimeric glycosylated type II transmembrane protein.^[53] It is mainly extracellular (760 residues form the ectodomain), and it is anchored by a 22-amino acid transmembrane sequence, leaving a short intracellular N-terminus tail of 6 residues.^[52]

There are several crystal structures of DPP IV, either alone or complexed, since it was first elucidated by Rasmussen et al. in 2003.^[53] Each monomer contains two domains: an α/β -hydrolase domain and a β -propeller domain. The former comprises a central β -sheet, made of 8 β -strands, and surrounded by 6 α -helices, while the latter consists of 8 blades, each formed by 4 antiparallel β -strands. The propeller can be subdivided into subdomain 1, comprising blades 2-5, which are bent out of cavity, and subdomain 2, formed by blades 6-8 and 1 that are bent towards the funnel. ^[53]

The active site of the enzyme is located in a cavity between the both domains, and it is formed by the catalytic triad Ser639, Asp708 and His740. The cavity is accessible by two openings, fact that has arisen controversies around which are the routes for the entrance of the substrate and the exit of the product. One access is the funnel in the centre of the β -propeller domain and the other is a side opening between the two domains. Though it has been not totally demonstrated, evidences support the idea that entrance of substrate is by the side opening, since if it was by the β -propeller domain, peptide substrates would be forced to make a strong turn in order to be processed, which, is structurally hindered.^[53]

DPP IV dimerizes by non-covalent interactions, in a hydrophilic interface. The transmembrane region, residues 658-661, 713-736 and 746-757 contribute to dimerization. Besides, an extended arm in the propeller domain stabilizes the dimer. In is formed by the elongation of the β -strand 2 of blade 4, which creates a two-stranded β -sheet arm.^[53]

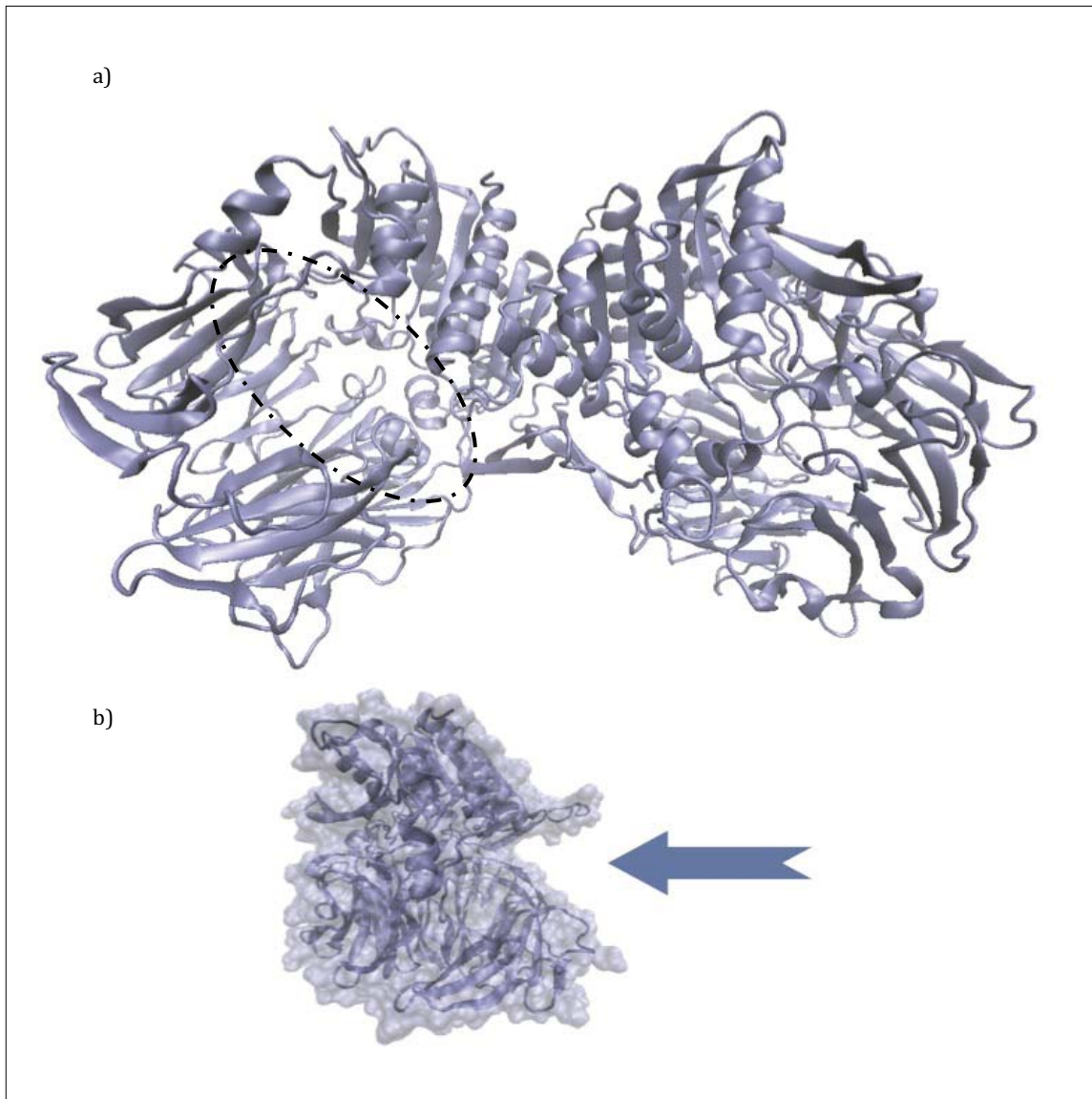


Figure i.3.: DPP IV crystallographic structure (pdb 4A5S). a) Dimeric DPP IV, displaying the α/β -hydrolase domain and the β -propeller domain. The active site is located between the two domains. The entrance of the substrate is through the opening circled in black. b) To better visualize the substrate entrance, a single monomer is shown. It has been turned, so that the entrance is placed at the right of the structure (depicted by an arrow).

4.1.1.2. Expression

DPP IV is found ubiquitously, though it is expressed in higher levels in the kidney cortex and also in small intestine brush-border membranes.^[54] Furthermore, it is expressed in intestine, liver and pancreas suggesting a digestive role in the degradation of peptides derived from nutritional proteins.^[55] In the adult brain DPP IV is found in capillaries but also in the blood-brain barrier, in circumventricular organs and in leptomeninge cells.^[55] The protease is also expressed on T and B lymphocytes as well as in natural killer (NK) cells.^[56] Regarding T-cells, the DPP IV expression levels are increased after lymphocyte activation.

4.1.1.3. Functions

As a protease, DPP IV hydrolyzes a broad spectrum of substrates (table i.1.). It has been proposed that a proline residue at the N-terminal is a natural protection for bioactive peptides, since few proteases can cleave these residues. Thus, DPP IV is an important protein in the processing of these substrates. (References in table: [52, 54, 57-66])

GLP-1, GLP-2 (Glucagon-Like Peptide 1 and 2) and GIP (Glucose Insulinotropic Peptide) are incretin molecules produced by L cells of the duodenum (GLP-1 and GLP-2) and K cells (GIP) in the intestine. These molecules travel through the blood stream to the pancreas, where they stimulate an increase in insulin production and release.

Substance P is a pancreatic neuropeptide involved in the regulation of blood flow, pancreatic neural signalling and exocrine function. [67]

SDF-1 is a proinflammatory chemokine (CXCL12)22 that attracts leukocytes to the Central Nervous System (CNS). Besides, SDF-1 is involved in tissue homeostasis and in metastasis by chemotaxis of tumoral cells.[68]

GASTROINTESTINAL HORMONES

| DPP IV substrate | Main expression | Function/Roles | Reference |
|--|--|--|-----------|
| Glucagon Glucagon Like Peptide-1 (GLP-1) | Pancreatic α -cells L cells of the ileum and colon | Induction of insulin production and secretion by the pancreatic β -cells in a glucose-dependent manner | 57 58 |
| Glucose-dependent Insulinotropic Peptide (GIP) | K cells of the duodenum | Induction of insulin production and secretion by the pancreatic β -cells in a glucose-dependent manner | 57 58 |
| Glucagon Like Peptide-2 (GLP-2) | L cells of the ileum and colon | Upregulation of intestinal glucose transport Inhibition of food intake Gastric emptying Reduction of bone resorption Promotion of neuronal proliferation Inhibition of cell apoptosis in the intestinal crypt | 57 59 |
| Pituitary adenylate cyclase-activating peptide (PACAP) | Pancreatic islet and hypothalamus | Increase of insulin and glucagon secretion in a glucose-dependent manner | 60 |
| Secretin | S cells of the jejunum and duodenum | Control of homeostasis in the pancreatic duct | 61 |
| Vasoactive Intestinal Peptide (VIP) | Nervous system Appendix Cardiac myocytes | Increase of insulin and glucagon secretion in a glucose-dependent manner Stimulation of intestinal fluid secretion Vasodilation | 54 |
| Peptide histidine methionine (PHM) | Nervous system | Stimulation of release of insulin and glucagon Vasodilation | 54 |
| Growth hormone-releasing factor (GRF) | Hypothalamus | Stimulation of GH production | 54 |

(continued in next page)

| CHEMOKINES | DPP IV substrate | Main expression | Function/Roles | Reference |
|------------|---|---|---|-----------|
| | Macrophage Inflammatory Protein-1 α (MIP-1 α /CCL3) | Monocytes T cells Mast cells Fibroblasts Lymphocytes | Antiviral defense Promotion of T _H 1 immunity Attraction of: monocytes, NK cells, T cells, basophils and dendritic cells Chemotaxis for inflammation signal | 54 |
| | Regulated on activation, normal T expressed and secreted (RANTES/CCL5) Eotaxin (CCL11) | Endothelium Monocytes Epithelium T cells | Role in allergy Attraction of: eosinophils, monocytes and T cells | 54 |
| | Macrophage-derived chemokine (MDC/CCL22) | Epithelial cells Macrophages T _H 2 cells | Chemoattraction of T _H 2 cells Chemotaxis of monocytes, dendritic cells and NK cells | 63 |
| | Gro β (CXCL2) | Monocytes Fibroblasts Endothelium | Activation of neutrophils Fibrosis Angiogenesis Attraction of: neutrophils, naive T cells and monocytes | 54 |
| | GCP-2 (CXCL6) | Macrophages Epithelial and mesenchymal cells | Chemotaxis of neutrophilic granulocytes | 54 |
| | MIG (CXCL9) | Keratinocytes | Activation of macrophages to secrete mediators of inflammation | 54 |
| | Interferon gamma-induced protein-10 (IP-10, CXCL10) | Keratinocytes T cells Monocytes Fibroblasts Endothelium | Immunostimulation Antioangiogenesis promotion Promotion of T _H 1 immunity Attraction of: resting T cells, NK cells and monocytes | 52 |
| | ITAC (CXCL11) | Keratinocytes Fibroblasts Endothelial cells | Chemotaxis of interleukin-activated T cells Induction of calcium release in activated T cells | 54 |
| | Stromal cell derived Factor-1 (SDF-1/CXCL12) | Stromal cells | B-cell development Lymphocyte homing Attraction of: naive T cells, progenitor B cells | 54 |

(continued in next page)

| PANCREATIC POLYPEPTIDE FAMILY | | | |
|---------------------------------|--|---|-----------|
| DPP IV substrate | Main expression | Function/Roles | Reference |
| Pancreatic Polypeptide (PP) | Pancreas | Regulation of pancreas secretions | 52 |
| Neuropeptide Y (NPY) | Hypothalamus | Reduction of appetite | 64 |
| Peptide YY (PYY) | L cells of the ileum and colon Pancreas | Increase of absorption in the intestine Reduction of appetite Regulation role in the postprandial pancreatic exocrine secretion | 65 |
| OTHER | | | |
| DPP IV substrate | Main expression | Function/Roles | Reference |
| Substance P | Secreted by nerves and inflammatory cells | Pancreatic neural signalling Blood flow regulation | 66 |
| β -Casomorphin | Intestine | Regulation of pancreas exocrine function | |
| Endomorphin-1 | Brain cortex | Exorphine-Analgesia | 54 |
| Endomorphin-2 | Brain cortex | Analgesia | 54 |
| Gastrin Releasing Peptide (GRP) | Brain Intrinsic neurons of the gut Parasympathetic neurons of the pancreas | Analgesia Stimulation of GLP-1, gastrin and other gastrointestinal hormones release | 54 57 |

Table i.1: DPP IV substrates.

Besides its proteolytic function, DPP IV interacts with a set of proteins. Basically, there are two types of interactions: the ones promoting the co-stimulation of T-cell activation and proliferation (table i.2.), and those involving proteins from the extracellular matrix (table i.3.). (References in table: [12, 69-79])

Adenosine Deaminase (ADA; EC 3.5.4.4) is an extracellular enzyme that hydrolyses adenosine and 2'-deoxyadenosine to inosine. DPP IV/ADA occurs at the T-cell surface and has two potential objectives. First, recruitment of ADA at the extracellular space of T lymphocytes may favor adenosine degradation, which is a T-cell proliferation inhibitor.^[80] Second, interaction of the two proteins, promotes the activation of T lymphocytes by co-stimulation of the TCR-CD3 complex.^[81] This last function is independent of their respective enzymatic activities.

Seprase or Fibroblast Activation Protein (FAP) is an endopeptidase whose main substrates are gelatinase, α -antiplasmin and collagen.^[12, 69, 70, 80] FAP is supposed to be involved tissue remodeling and enhancement of metastasis.^[80]

| DPP IV interacting protein | Interaction region | Function/Roles | Reference |
|--|--------------------|---|-----------|
| Adenosine Deaminase (ADA) | Extracellular | Hydrolysis of Adenosine Co-activator of T-cells by binding to DPP IV | 70 |
| Receptor-type tyrosine-protein phosphatase C (PTPRC - CD45) | Extracellular | Co-activation of T-cells after binding to DPP IV Tyrosine phosphatase activity | 71 |
| Cation-independent mannose-6-phosphate receptor (CIMPR) or Insulin-like growth factor-2 receptor (IGF2R) | Transmembrane | Transport of phosphorylated enzymes to the cell surface Co-activation of T-cells by binding to DPP IV | 12 |
| Caveolin-1 (CAV1) | Membrane | Induction of T-cell activation by costimulation of TCR in a CD3-dependent manner | 74 |
| Caspase recruitment domain-containing protein 11 (CARD11 or CARMA1) | Intracellular | Induction of T-cell activation by costimulation of TCR in a CD3-dependent manner Induction of T-cell proliferation and activation of NF- κ B after DPP IV binding | 73 |
| Seprase - Fibroblast Activating Protein (FAP) | Extracellular | Tissue remodeling by collagen degradation Proteolysis of ECM Promotion of migration of endothelial cells | 69 70 |
| Urokinase-type plasminogen activator (uPA) | Extracellular | Cleavage of plasminogen to form plasmin which degrades certain ECM proteins | 76 |

Table i.2.: DPP IV interacting partners involved in T-cell activation

| DPP IV interacting protein | Interaction region | Function/Roles | Reference |
|--|--------------------|---|-----------|
| Collagen | Extracellular | Connective tissue component Fibrous tissue component | 79 |
| Fibronectin III | Extracellular | Involvement in cell adhesion, motility and wound healing (interaction only seen in rat DPPIV) | 75 |
| Glypican-3 | Cell surface | Growth modulation of mesodermal tissues Regulation of tumor growth | 78 |
| Na ⁺ - H ⁺ exchanger isoform 3 | Transmembrane | pH regulation Signal transduction | 76 |

Table i.3.: DPP IV interacting partners from the extracellular matrix.

4.1.2. DPP IV and its therapeutic implication

The protease is involved in conditions such as diabetes, cancer, rheumatoid arthritis and autoimmune diseases. The role of DPP IV in them can be mediated either through PPI or by the enzymatic activity. It is in the latter scenario that DPP IV inhibitors may provide potential pharmacological drugs.

4.1.2.1. Type 2 Diabetes Mellitus

The incretin molecules GLP-1 and GIP are relevant molecules for blood glucose homeostasis due to their role in insulin excretion. After nutrient uptake, these two molecules are released to the blood flow. Then, once in the pancreas, they bind to their specific receptors and activate the expression and secretion of insulin.

However, both hormones are rapidly inactivated in the blood flow by DPP IV.^[55] (The half-life of intact GLP-1 is lower than 1 min).^[82] Interestingly, the inactivated forms of these incretin molecules retain the capacity to interact with their respective receptors at the pancreatic β -cell, acting as antagonists without the correspondent insulinotropic effect (figure i.4.)

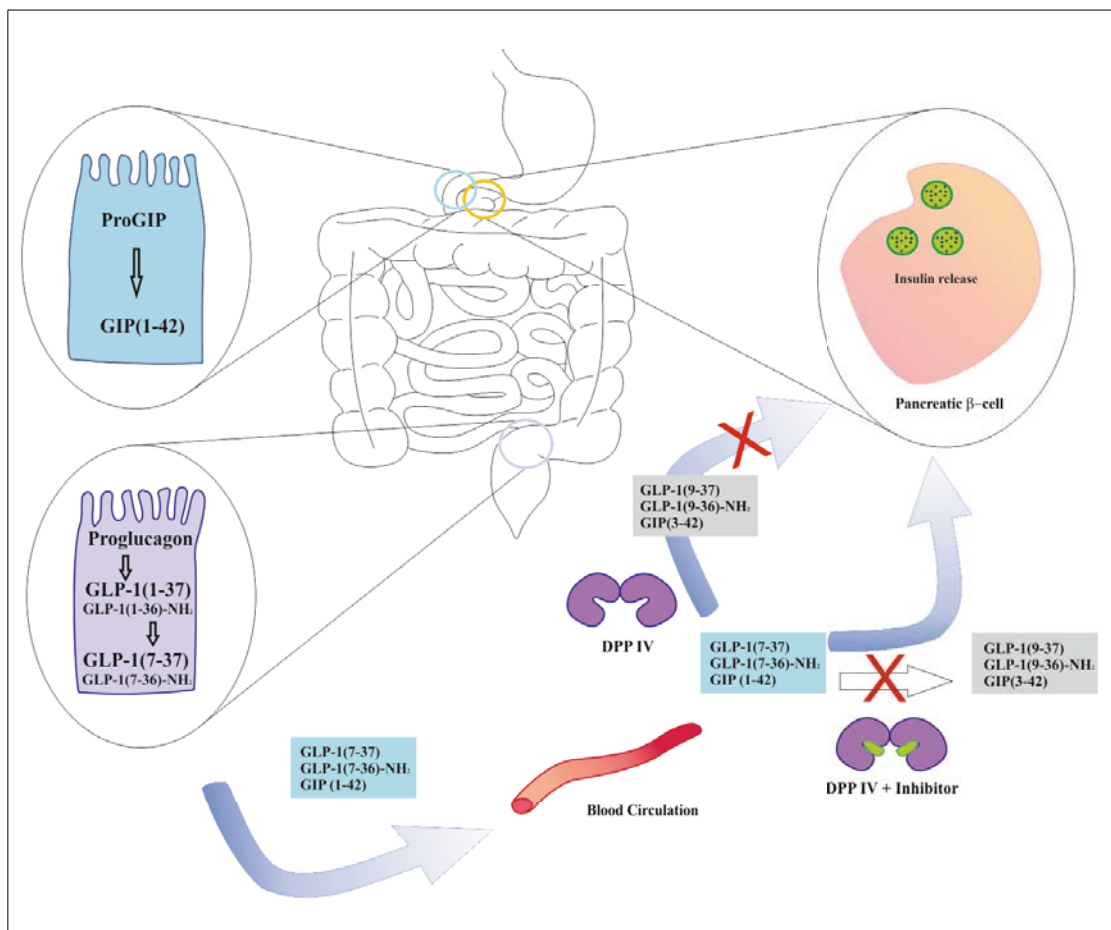


Figure i.4.: Schematic representation of the participation of DPP IV in incretin molecule inactivation.

DPP IV is a validated target for the treatment of Type 2 Diabetes Mellitus (T2DM) and its inhibitors are widely reported in the literature as effective blood glucose-lowering molecules.^[83-85]

4.1.2.2. Other pathologies

The connexion of DPP IV to cancer is controversial.^[86]

In some cancers, DPP IV is downregulated or not modified.^[54, 87] For example, in melanoma cells, the expression is lost while the tumor is in progression.^[54] The peptidase is also downregulated in ovarian carcinoma and it has been shown that transfection of DPP IV is able to revert tumor malignancy.

In other cancer types, the protease is overexpressed and has increased activity.

It is the case of most of T-cell malignancies^[54, 87] (3,42_Review) as well as in B-chronic lymphocytic leukemia cells.^[87] Thus, DPP IV inhibitors are useful for the treatment of certain T-cell cancers.

Regarding DPP IV involvement in rheumatoid arthritis, again it is a controversial issue.

On one hand, it has been observed that DPP IV inhibitors block the lymphocyte transendothelial migration, therefore reducing synovial inflammation.^[88]

On the other hand, inactivation of DPP IV reinforces the activity of chemokines. For example, SDF-1, a DPP IV substrate, promotes leukocyte chemotaxis and inflammatory response. Thus, the inhibition of the protease boosts inflammation.^[89]

Despite the disagreement among the scientific communication, several DPP IV inhibitors have been patented for rheumatoid arthritis treatment and are currently under development.

Finally, DPP IV is involved in the activation and proliferation of T lymphocytes by numerous PPI. However, the proteolytic is essential for the cleavage of cytokines and for T-cell transendothelial migration.^[90] On the basis of this observation, DPP IV inhibitors are being used for immune suppression in organ transplantation and in autoimmune diseases.^[90]

4.1.3. DPP IV inhibitors

The story of DPP IV inhibitors drug discovery is the typical example of the classic evolution of protease inactivators. Starting with substrates analogs to the small organic molecules of nowadays, there are plenty of reported DPP IV inhibitors in the bibliography, that differ in their chemical structure basis, affecting thus essential properties like potency, specificity, stability and toxicity. Here, the most relevant ones are highlighted.^[86]

In the early '80s DPP IV substrates or products analogs were studied as inhibitors. Diprotins A and B, extracted from supernatants from *Bacillus cereus* culture,^[91] presented modest IC₅₀ values of 3.2 mM and 16.7 mM respectively.

Late in the '90s the use of proline derivatives afforded more potent inhibitors. Flentke et al designed DPP IV inhibitors holding the α - amino boronic acid analog of proline in the P1 position.^[92] Though Ki values for DPP IV were in the low nanomolar range, boronic acid analogs were unstable.^[93]

Further modification of the proline position led to the discovery of pyrrolidides and thiazolidides as highly potent and stable DPP IV inhibitors. For that reason, these moieties have been extensively used in inhibitor scaffolds.^[94]

Following this discovery, DPP IV-inhibitors based on pyrrolidine and thiazolidine were described.^[93] Such a small change afforded substantial gain in potency and specificity.^[94]

In 2002 a totally new scaffold was proposed as DPP IV inhibitor, based on xanthine molecules and with IC₅₀ values in the nanomolar range.^[95, 96]

Nowadays, there are five DPP IV inhibitors in the market (table i.4.).

In 2006, Merck launched the first DPP IV inhibitor onto the market, namely sitagliptin (MK-0431, Januvia, Glactiv).^[97] This molecule is based on a beta-amino heterocyclic scaffold and is a once-daily oral treatment for T2DM.

In 2007, Novartis started commercializing vildagliptin (LAF-237, Galvus), a cyanopyrrolidine based compound. As well as Januvia, Galvus is a once-daily oral treatment for T2DM treatment.

In 2009, Bristol-Myers Squibb and AstraZeneca launched Saxagliptin (Bristol-Myers Squibb -477118, Onglyza).

In 2010 Takeda started commercializing alogliptin (SYR-322, Nesina, Vipidia)^[98] in Japan. Later, in 2013 launched it in the U.S. and in after being approved by the UE the same year, it started its commercialization in the U.K. in 2014.

Finally, Linagliptin (BI 1356), co-developed by Boehringer Ingelheim and Lilly and under active development, is available in US under the name of Tradjenta, in Europe as Trajenta and in Japan as Trazenta.

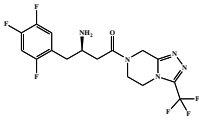
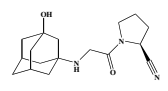
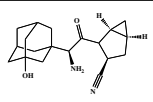
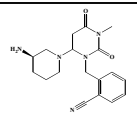
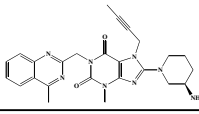
| Generic name | Formula | Countries | Launch year | Company | Commercial name |
|--------------|---|---|--|---|--|
| Sitagliptin |  | Mexico, U.S. Germany, U.K. Canada Spain Japan | 2006 2007 2008 2008 2009 2009 | Merck Merck Merck Merck Sharp & Dohme MSD KK Ono | Januvia Januvia Januvia Januvia Januvia Glactiv |
| Vildagliptin |  | Mexico, Brasil Germany, U.K. Japan | 2007 2008 2010 | Novartis Novartis Novartis | Galvus Galvus Equa |
| Saxagliptin |  | Germany, U.K., U.S. Switzerland | 2009 2009 | AstraZeneca and BMS AstraZeneca | Onglyza Onglyza |
| Alogliptin |  | Japan U.S. U.K. | 2010 2013 2014 | Takeda | Nesina Nesina Vipidia |
| Linagliptin |  | U.S. Europe Japan | 2011 2011 2011 | BI & Lilly | Tradjenta Trajenta Trazenta |

Table i.4.: Marketed DPP IV inhibitors ordered by their launch year.

4.2. Prolyl oligopeptidase (POP)

Chapter three of results is devoted to this protease.

4.2.1. The protein

Prolyl oligopeptidase (POP, EC 3.4.21.26) is a serine endopeptidase that belongs to subfamily A of the family S9 (clan SC). It processes small peptides of around 30 residues, cleaving after C terminus of proline.^[99]

4.2.1.1. Structure

POP is a 81 KDa monomeric cytosolic protein.

The porcine POP crystal structure was elucidated by Fulop et al. in 1998.^[100] As POP is structurally related to DPP IV, its structure reminds to a DPP IV monomer. POP consists of two domains as well: an α/β -hydrolase (or catalytic domain) and a β -propeller domain (or structural domain). While the DPP IV propeller domain consisted on eight blades, the one of POP is a seven-bladed β -propeller, which is more regular.^[53]

The active site of is located in a cavity between the both domains, and it is formed by the catalytic triad Ser554, Asp641 and His680. Regarding the two openings, since the propeller domain contains an extra blade, compared to DPP IV, the funnel in the centre of it is smaller, and thus, entrance of substrate or exit of product is not likely to occur through this opening.

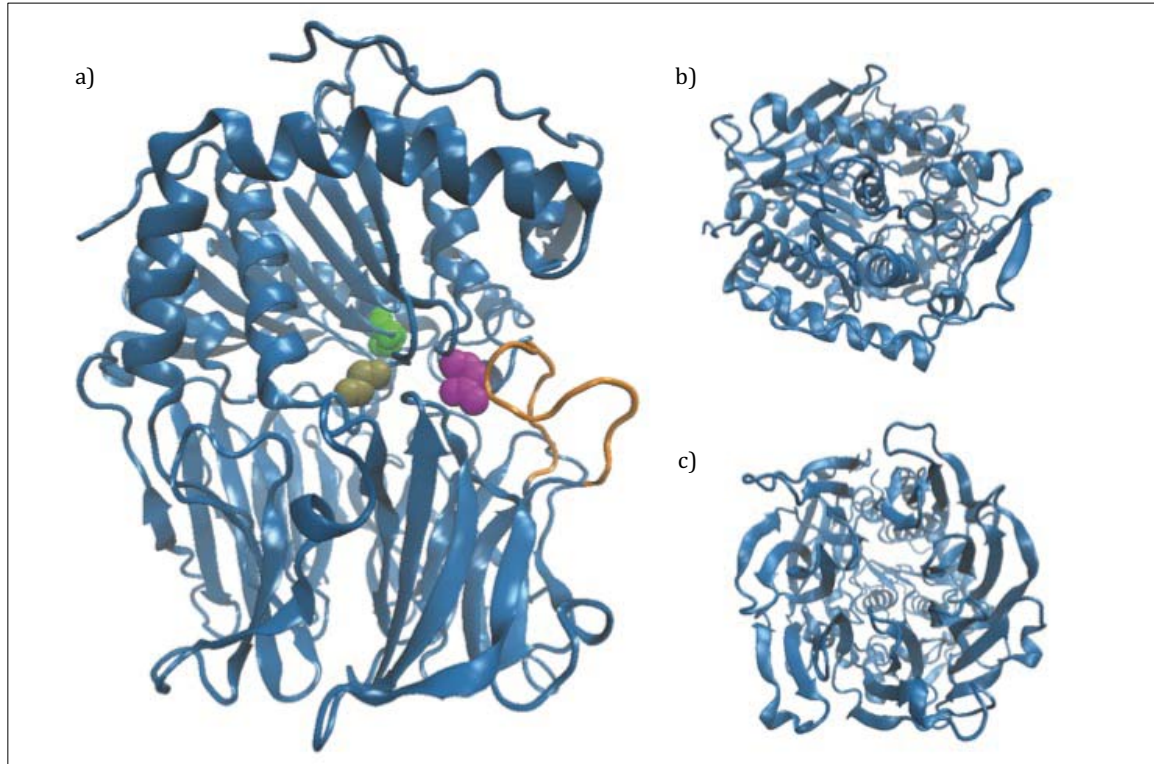


Figure i.5.: POP crystallographic structure (pdb 1H2W). a) Lateral view, loop in the opening (orange) and active site residues are highlighted: Ser (green) Asp (magenta) and His (tan); b) Top view (α/β -hydrolase domain); c) Bottom view (β -propeller domain)

POP is believed to be a highly dynamic protein.^[101] In NMR studies, equilibrium between different conformations was observed. While structural information of the conformations is not available, it was detected that POP samples two states, open and close, which are in equilibrium. Besides, the addition of POP inhibitors is able to displace it. This theory is in contraposition with the suggested induced-fit.^[102]

4.2.1.2. Expression

POP is expressed ubiquitously, including the central nervous system (CNS). It is mainly located in the cytoplasm of neuronal projections, but also it concentrates in the nucleus of young and old neurons. It is precisely, in developmental stages of neurons when POP activity is increased.^[103]

4.2.1.3. Functions

As a protease, POP hydrolyses most neuropeptides and peptide hormones, such as substance P, oxytocin or neurotensin among others. However, the cleavage was demonstrated *in vitro*, but it is not explained how an intracellular protease can degrade extracellular substrates.

Neuropeptides are molecules used by neurons to communicate between them. Among their actions, they can activate genes. Neuropeptides co-exist with neurotransmitters, which depolarize or hyperpolarize neurons.

POP interacts with other proteins apart from its proteolytic activity.

Growth Associated Protein 43 (GAP43) was identified as a POP-interacting protein.^[104] Despite performing a weak and transient interaction, they were co-localized in HeLa cells.^[105] GAP43 is a cytosolic protein, expressed in neurons, where it is involved in neuronal growth cone formation, axon guidance and synaptic plasticity.^[99]

POP also interacts with α -synuclein, an intrinsically disordered cytosolic protein that aggregates into fibrils. As observed in Parkinson's disease patients, α -synuclein fibrils accumulate into Lewy bodies. The interaction of POP/ α -synuclein was suggested as an aggregation accelerator since treatment of cells with POP inhibitors produced a decrease in α -synuclein aggregation.^[99]

4.2.2. POP and its therapeutic implication

4.2.2.1. Cognitive disorders

A role for POP in neurological disorders, such as schizophrenia^[106] or bipolar disease, was proposed after the discovery of some evidences. First, the fact that the protease was able to cleave *in vitro* certain neuropeptides. Second, that treatment of rats with POP inhibitors increased levels of such substrates and neurotransmitters (acetylcholine and dopamine) as well.^[107] And third, the detection of altered POP activity in the serum of patients suffering from mood disorders.^[108-110]

All this together suggested that POP was responsible for *in vivo* neuropeptide degradation. This processing lowered the neurotransmitters concentration affecting the neuronal communication and stimulation.

Precisely, patients suffering from cognitive disorders have lower levels of neurotransmitters, thus, POP was proposed as a therapeutic target in these conditions.

However, as mentioned before, neuropeptides are extracellular while POP is in the cytosol. Besides, a study of 2012,^[111] demonstrated that POP inhibitors had little effect on acetylcholine levels. Therefore, the hypothesis was rejected.

Later, in 1999, mutants of the amoeba *Dyctiostelium discoideum*, which did not express POP, were found to have increased levels of inositol triphosphate (IP3). The relationship between POP and IP3 was further confirmed by treatment of wild type amoebas with POP inhibitors. As a result, concentrations of IP3 were increased as well.^[112]

The connexion of IP3 with neurological disorders is explained by its signalling pathway. After interaction of the neuropeptides with their neuronal receptors, intracellular levels of IP3 are increased, which provokes an augmented release of ion calcium (Ca^{+2}) from the endoplasmic reticulum. Once in the cytosol, Ca^{+2} is involved in memory and learning.^[113] Thus, decreased levels of the ion and members of the pathway lead to mood disorders.^[114] Then, as POP inhibitors can increase IP3 concentration, they are expected to be drugs for the treatment of cognitive disorders.

However, the direct relationship of POP and IP3 was not described.

Furthermore, there is no full consensus in the literature about the relationship between Ca^{+2} levels and mood disorders.

In 2009, the POP null mutant mouse was generated.^[104] It was observed that the neurons presented alterations in growth cone dynamics that could be restored with addition of POP, regardless of its proteolytic activity. Thus, the effect was suggested to be a result of a PPI, which was then identified as POP/GAP interaction. The finding that this PPI was essential in neuronal growth explained the role of POP in cognitive disorders, yet the overall relationship of POP/GAP/IP3 is not clear.

For that reason POP is not a validated target for the treatment of cognitive disorders.

4.2.2.2. Parkinson's disease

As explained before, POP interacts with the intrinsically disordered protein α -synuclein. It has been observed that aggregation of this protein can be reverted by addition of POP inhibitors.^[115] For that reason, the serine protease is believed to play a role in Parkinson's disease, where aggregates of α -synuclein are detected. However, it is neither a validated target for this condition.

4.2.2.3. Cancer

In 2010, an increased POP activity was detected in tumors.^[116] In the same year, POP was described as an inducer of angiogenesis, and that POP inhibitors were able to reduce the process.^[117] Thus, the protease is a potential target for cancer treatment.

4.2.3. POP inhibitors

As explained in the previous section, the relationship of POP with several diseases is yet no proved, but treatment of animal models with POP inhibitors has suggested a therapeutic application for those inactivators.

But, if the mechanism by which POP is related to the conditions is unclear, the positive effects seen by treatment with inhibitors are even of more difficult explanation.

However, it is believed that the action of inhibitors is not important because they ablate the proteolytic activity, but because they modify the three-dimensional structure of POP, avoiding its interaction with protein partners.

Next, a historical description of POP inhibitors development is detailed and structures are shown in figure i.5.^[106]

In 1988, the molecule Benzyloxycarbonyl-prolyl-prolinal (ZPP) was developed by Microbial Chemistry Research Foundation and Yakult Honsha as a POP inhibitor for the treatment of neurothic pain and as an anti-HIV agent.^[32] It is a covalent reversible inhibitor that binds to the active Ser554 through a hemiacetal formation. Despite the highest phase it reached was preclinical, ZPP is the reference structure used for the development of POP inhibitors.

Modifications of P1' ZPP position led to the discovery of JTP-4819 and KYP-2047.

In 1994, Mitsubishi Tanabe Pharma and Japan Tobacco describe JTP-4819. Despite the beneficial effects observed in scopolamine treated rats,^[106] the highest phase it reached was phase II.

Years later, in 2000, the University of Eastern Finland developed KYP-2047 as a POP inhibitor for the treatment of cognition disorders. It reached the preclinical phase.

P2 modifications afforded the finding of ONO-1603 and S-17092.

In 1988, Ono started investigations on ONO-1603 for the treatment of cognition disorders and it reached phase II.

In 1996, S-17092, discovered by Servier, was rejected at phase I.

Finally, in 1990 P3 modifications gave rise to compound Z-321 for treatment of Alzheimer's disease and dementia. Developed by Zeria, it reached phase I.

To date, there is only one POP inhibitor under active development.^[32] It is (-)-Epigallocatechin gallate, a compound present in the aqueous extract of the Japanese green tea leaves. It was registered in 2002 as potential drug in a variety of diseases, ranging from Parkinson, sclerosis, cognition disorders, dementia, cancer or metabolic disorders, among others.

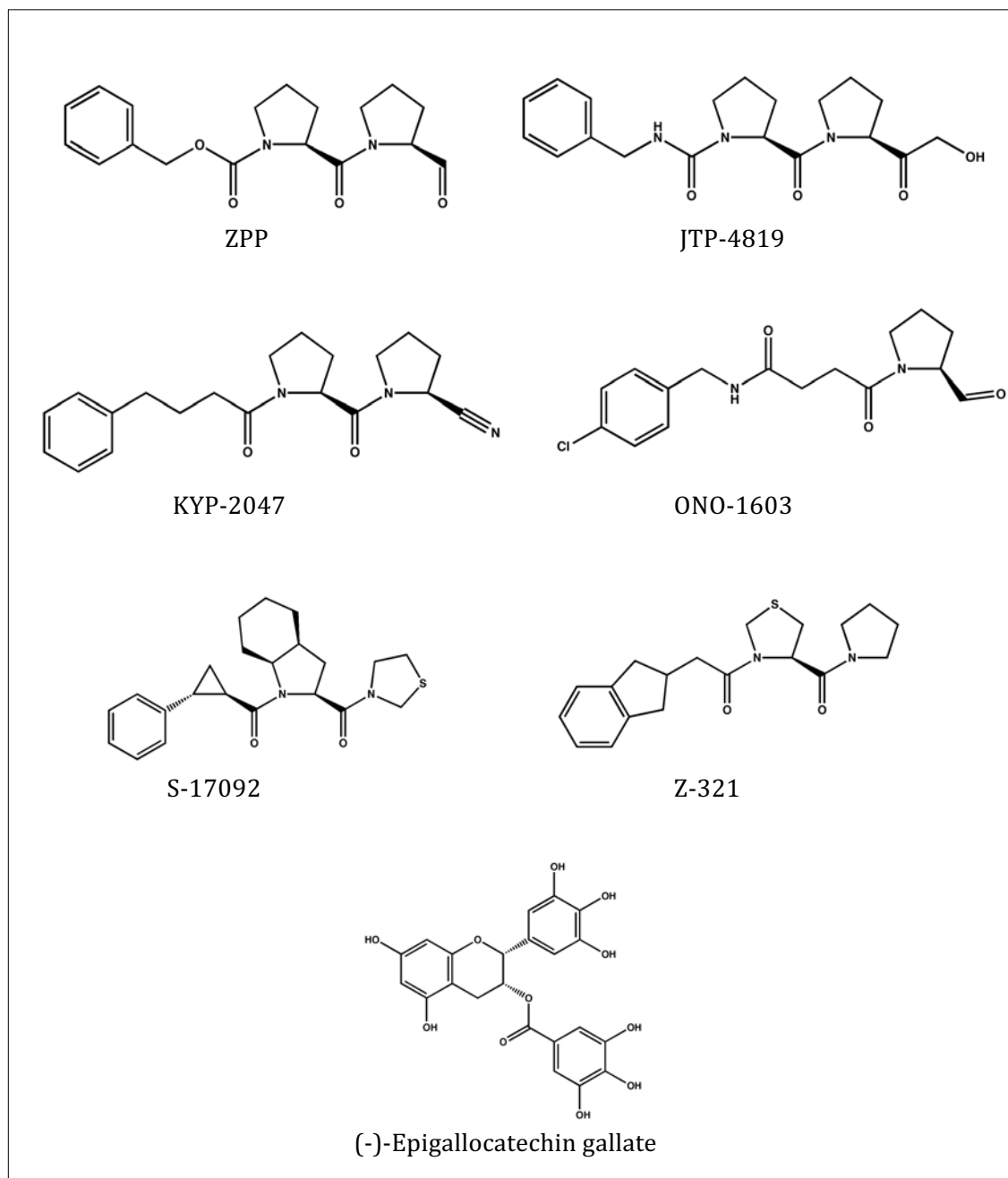


Figure i.6.: POP inhibitors.

4.3. Cathepsins L and B

Chapter four of results is devoted to these proteases.

4.3.1. The proteins

Cathepsin L (EC 3.4.22.15) and cathepsin B (EC 3.4.22.1) are cysteine endopeptidases belonging to subfamily A of the C1 family (clan CA).

4.3.1.1. Structure

Both proteases are expressed in the form of zymogens, called procathepsins. The propeptide has a stability role and is autoprocessed in acidic pH.^[118]

Regarding to Cathepsin L, its proenzyme is around 30 KDa, and the first crystal structure was elucidated by Coulombe et al in 1996.^[119] It consists of two domains: the first one is predominantly α -helical while the second is mostly β -sheet. The active site cleft, containing Cys25 and His163 is located in between of both domains. It is protected by a segment of the propeptide, which is oriented in the opposite sense to the natural substrate. The prosegment not only interacts with the mature enzyme by the active site cleft, but also through a loop in the β -sheet domain named as prosegment binding loop. This contact is principally driven by hydrogen bonds and hydrophobic interactions.^[119]

Cathepsin B structure was revealed before, in 1991, by Musil et al.^[120] As well as for other members of the family, it is formed by two domains, with the active site located between them in a polar cleft.

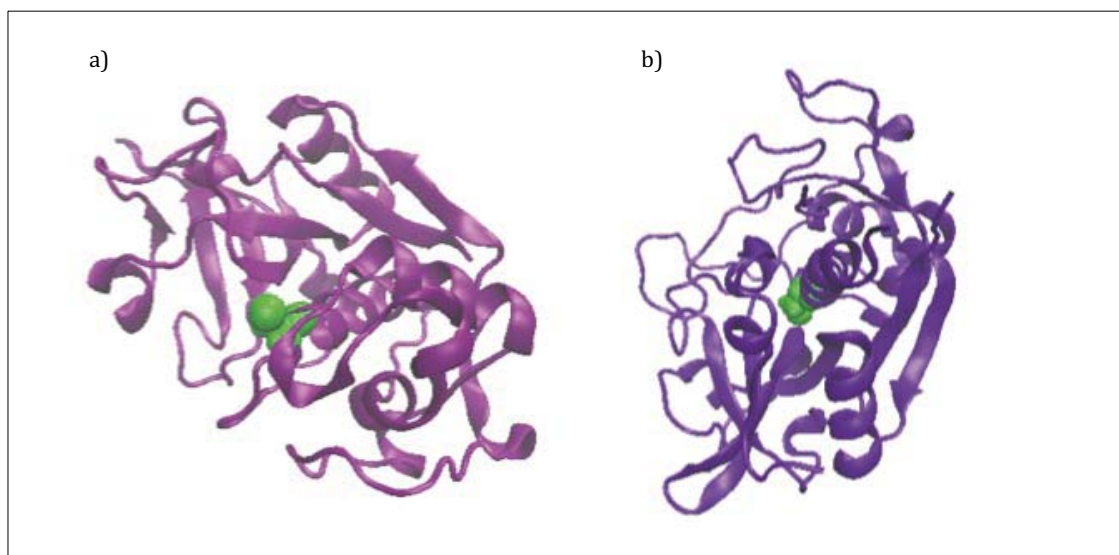


Figure i.7.: Crystallographic structures of a) cathepsin L; b) cathepsin B. Cysteine residue of the active center is shown in green for both proteases.

4.3.1.2. Expression

They are ubiquitously expressed, though expression levels vary depending on the tissue.^[121] Besides, their expression is highly regulated at all the cell cycle stages.

Cathepsins are particularly concentrated in the lysosomes. (2,4,5_article). But, in cytotoxic T lymphocytes and natural killer cells, the cysteine proteases are rerouted from lysosomes to the cell surface.^[122]

4.3.1.3. Function

Cathepsin B hydrolyses a variety of substrates, with preference for those containing basic residues or phenylalanine at P2 and not-bulky side chain amino acids in P1.^[123] On the contrary, cathepsin L shows a more restrictive substrate profile. It cleaves peptide bonds formed by a hydrophobic amino acid in P2 and a polar residue in P1. ^[124]

Cathepsins were thought to be housekeeping enzymes, due to their specific accumulation in lysosomes. However, further investigations revealed other functions.

Cathepsins process proteins in the nucleus and intracellular organelles (hormone secretory granules). Thus, these proteases are implicated in a variety of processes: such as protein turnover, enzyme activation, hormone maturation, bone remodelling, antigen presentation and epidermal homeostasis.^[122, 125-127]

In addition, degradation of the extracellular matrix (ECM) is also carried out by cathepsins. For example, cathepsin B degrades three of the most important basement components of the ECM.^[128]

4.3.2. Cathepsins L and B and their therapeutic application

Cathepsins L and B play a key role in a number of human diseases, such as cancer and rheumatoid arthritis.

4.3.2.1. Cancer

In cancer cells, cathepsins are redirected to the cell surface,^[122] and in some types of cancer, these proteases are also secreted to the exterior.^[122, 128, 129] Cathepsins are active in low pH, and since the extracellular microenvironment of tumors is generally acidic, the proteolytic activity is retained after translocation.^[122] Once outside the cellular compartment, cathepsins degrade the basement and extracellular matrix proteins,^[122, 128, 129] thereby promoting angiogenesis and tumor invasive growth^[128] and spread.^[128, 129]

Apart from their cellular location change, the expression and activity of cathepsins are also modified in cancer cells.

The malignancy of the tumor is correlated with the cathepsin levels, with higher expression in the advanced tumorigenic tissues.^[122, 128, 129]

The proteolytic activity is concentrated on the edges of the islet carcinomas, and evident in angiogenic islets and tumors.^[128]

The relevance of cathepsins in cancer was further demonstrated by knockdown experiments. Cathepsins L and B knockdown mice showed a decrease in cell proliferation of 58% of reduction (cathepsin L) and 44% (cathepsin B).^[130]

Based on the scientific evidences, cathepsins L and B are potential therapeutic targets in cancer treatment.

4.3.2.2. Rheumatoid arthritis

High levels of cathepsins were found in synovial fluids of patients suffering from inflammatory arthritis.^[131] The proteases were detected in the early stages of the disease, suggesting a proteolytic role in the joint destruction at the very beginning of the condition.

The induction for their secretion was attributed to the action of cytokines.^[132] Tumor necrosis factor- α (TNF- α) and platelet-derived growth factor (PDGF) stimulated cathepsin B secretion, while basic fibroblast growth factor (bFGF) increased cathepsin L excretion. Finally interferon-gamma (IFN- γ) was an inducer for both proteases.

Later, a second mechanism for the release induction of cathepsins was described. The ATP-gated P2X(7) receptor (P2X(7)R) was already described as a potential therapeutic target in the treatment of inflammatory diseases, since their antagonists inhibited the secretion of proinflammatory cytokines. In further experiments, it was discovered that after activation of the receptor by ATP, cathepsins were rerouted in a fast manner from the lysosomes to the extracellular space, independently of the presence of cytokines.^[133] Furthermore, receptor antagonists were able to arrest the secretion of the cysteine proteases.

In summary, cathepsin secretion is induced in inflammation tissues. Once outside the cell, the cysteine proteases degrade the extracellular matrix, provoking destruction of the tissue. For that reason, cathepsins are potential targets in the amelioration of rheumatoid arthritis damages.

4.3.3. Cathepsins L and B inhibitors

In vitro treatment of human cancer cells^[134] with cathepsin inhibitors showed a reduction in the replication of cells. This was confirmed with *in vivo* treatment of mice tumors, where cathepsin inhibitors afforded a reduction of the volume^[128], invasion^[122] and growth of tumors and vessel density.^[128, 130]

Literature is full of cathepsins L and B inhibitors, but none of them has reached the market. Here the two molecules that are actually under active development are discussed.^[32]

VBY-891 and VBY-825 were registered both in 2012.

The first one, which is in phase I and whose structure is not disclosed, is being co-developed by Leo and Virobay as a cathepsins L, B, S, F and K inhibitor.

The second compound (figure i.7.), belongs to Virobay, and is in preclinical stage. It is described as a cathepsins L, B and S inactivator for the treatment of tumors by angiogenesis inhibition.

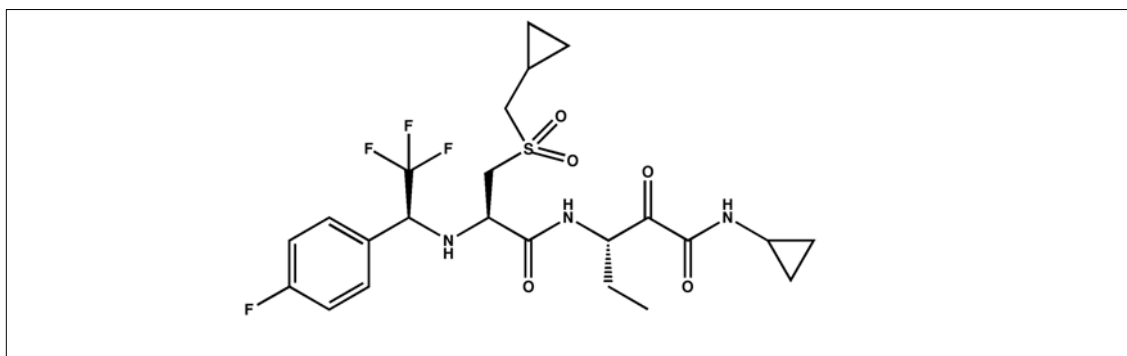


Figure i.8.: Compound VBY-825, a cathepsin L and B inhibitor under active development.

The disagreement between the number of cathepsins L and B inhibitor could be attributed to two causes. First, as previously said, these cysteine proteases are not validated, thus, problems related to the connexion between enzyme inhibition and the benefit in the disease may appear. Second, cysteine proteases are a family with similar active site, and therefore, finding of selective and potent inhibitors is an extremely difficult mission.

OBJECTIVES

Objectives

Proteases participate in a large number of biological pathways. While their activity can be a single and remote action, in most of the cases proteolytic events take part of complex networks. The protease signalling is irreversible and highly regulated and failures either on the activity or expression of proteases lead to aberrant signalling pathways. Thus, proteases are involved in several diseases and therefore are important drug targets.

For that reason, our main goal is to find protease inhibitors as therapeutic agents. We have focused our efforts in four proteolytic enzymes: Dipeptidyl Peptidase IV, Prolyl Oligopeptidase and Cathepsins L and B.

The first one, DPP IV, is a validated target for the treatment of type II Diabetes Mellitus. POP is a candidate target in mental diseases such as schizophrenia, bipolar disorder and Parkinson's disease. Finally, cathepsins L and B are believed to be crucial in tumor growth and metastasis.

In order to find inhibitors for the mentioned proteases, three strategies have been selected: first, screening of medicinal plant extracts; second, high throughput screening; and third, characterization of a combinatorial chemistry library.

Thus, to accomplish our goal, the following four objectives have been established:

1. To express DPP IV and perform an NMR study.
 - a) Set-up an expression system for the recombinant production of human DPP IV.
 - b) Analyse DPP IV by nuclear magnetic resonance and characterize the effect of inhibitors.
2. To discover DPP IV inhibitors from herbal origin.
 - a) Select plants with sugar-blood lowering properties.
 - b) Isolate DPP IV inhibitors from plant extracts and characterize the inhibition mechanism.
3. To discover POP inhibitors by high throughput screening.
 - a) Set-up a fluorescence polarization assay for the detection of POP active-site binders.
 - b) Screen non-toxic compounds against POP and validate them as inhibitors.
 - c) Characterize the inhibition mechanism of POP inhibitors.
4. To identify cathepsins L and B inhibitors from a combinatorial chemistry library
 - a) Screen a library, specially designed for the inhibition of cysteine proteases.
 - b) Characterize the inhibition mechanism of cathepsins inhibitors

RESULTS

The results obtained during the thesis are divided into four chapters. First two refer to the work devoted to DPP IV, the third to POP and the last one is related to cathepsins L and B. Chapters included in results sections are the following:

Chapter 1: DPP IV Labeling methodology and the interaction with its inhibitors

Chapter 2: Discovery of DPP IV inhibitors from plant extracts

Chapter 3: Discovery of POP inhibitors by High Throughput Screening

Chapter 4: Characterization of peptidyl aryl vinyl sulfones as Cathepsins L and B inhibitors

Chapter 1:
DPP IV Labeling methodology
and the interaction with its
inhibitors

CHAPTER 1 CONTEXT**1.1. DPP IV RECOMBINANT EXPRESSION IN INSECT CELLS**

- 1.1.1. Production of baculovirus
 - 1.1.1.1. Recombinant bacmids
 - 1.1.1.2. Recombinant baculovirus
 - 1.1.1.3. Amplification of recombinant baculovirus
- 1.1.2. DPP IV expression and purification
 - 1.1.2.1. Evaluation and optimization of expression conditions
 - 1.1.2.2. Purification set-up

1.2. DPP IV CHARACTERIZATION

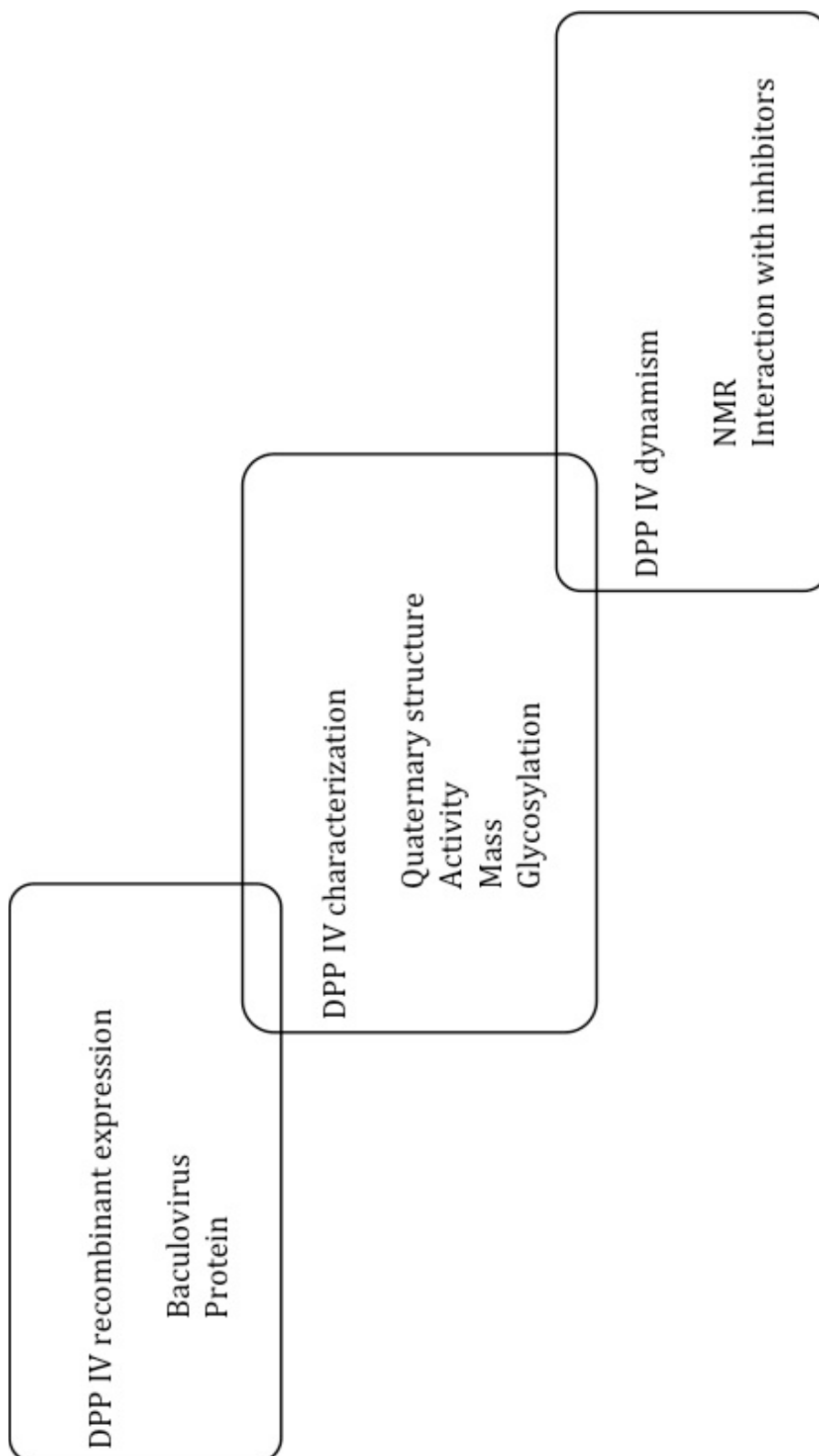
- 1.2.1. Quaternary structure
- 1.2.2. Activity
- 1.2.3. Mass of the protein
 - 1.2.3.1. ESI
 - 1.2.3.2. MALDI
- 1.2.4. Glycosylation pattern
 - 1.2.4.1. Glycosylated DPP IV detection
 - 1.2.4.2. Glycosylated DPP IV characterization: experiment planning
 - 1.2.4.3. Glycosylated DPP IV characterization: sample treatment
 - 1.2.4.4. Glycosylated DPP IV characterization: sample analysis

1.3. DPP IV DYNAMISM

- 1.3.1. Nuclear Magnetic Resonance (NMR)
 - 1.3.1.1. NMR applied to protein study
 - 1.3.1.2. Study of large proteins by NMR
- 1.3.2. DPP IV study by NMR
 - 1.3.2.1. Methyl-¹³C methionine selective labeling
 - 1.3.2.2. Methyl-TROSY experiment
- 1.3.3. Signal assignment: site-directed mutagenesis
 - 1.3.3.1. Site-directed mutagenesis
 - 1.3.3.2. Mutant DPP IV selective labeling
- 1.3.4. NMR study of DPP IV interaction with its inhibitors
 - 1.3.4.1. Interaction of DPP IV with Isoleucine-Thiazolidide (P32/98)
 - 1.3.4.2. Interaction of DPP IV with NVP DPP 728
 - 1.3.4.3. Interaction of DPP IV with berberine

CHAPTER 1 OVERVIEW

Experimental scheme



Chapter 1 context

The DPP IV crystallographic structure was solved in 2003^[135], and there are several examples in the Protein DataBase, either the protein alone or in complex with inhibitors. However, DPP IV dynamism has not been studied yet. The protein dynamic behavior is crucial in the understanding of several properties, even in ligand binding affinities. Prolyl oligopeptidase (POP), a closely related protein to DPP IV, has revealed a high dynamic equilibrium between two conformations^[101]. In basis of this finding, it was planned to evaluate the dynamism of DPP IV.

In order to achieve this objective, Nuclear Magnetic Resonance (NMR) was selected. It is a powerful tool that affords structural information while the protein is in solution, thus allowing the free movement of regions or even domains. This technique requires the protein labeling with isotopes, and therefore, an expression system for DPP IV was needed to allow its study by NMR.

DPP IV is a homodimeric serine protease anchored to the plasmatic membrane. The ectodomain of the protease comprises residues 29 to 766 and contains the active-site center of the protein. Besides of the membrane-bound protein, DPP IV is also present in a soluble form, starting from residue serine 39. ^[136]

Mammalian DPP IV has a complex glycosylation pattern ^[137] which accounts for 12 to 20% of the total protein molecular weight. ^[138] Although these glycoside modifications are not affecting the dimerization and activity of the enzyme ^[138], attempts to express it in prokaryotic cells have failed. On the contrary, previous bibliography has shown that the full DPP IV, as well as the ectodomain ^[139] and the soluble form ^[138] are active and stable when expressed in eukaryotic cells.

Based on that, a eukaryotic system for the expression of the ectodomain of DPP IV was selected. Yeast represents the most simple eukaryote organism for protein expression. Yeast cultures allow easy transfection of genetic material and yields tend to be high when compared to other eukaryotic systems. However, earlier experiments in DPP IV expression in *Pichia pastoris* have evidenced that a significant amount of protein was retained in the periplasmic space of the yeast.^[140]

On the contrary, mammalian cells represent the most complicated system. Mammalian cultures afford low yields, but proteins are similar to the human homologous.

In the middle, insect cells supposed the best option. The most common cell lines are *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Hi5). Genetic manipulation of insect cells was possible by means of the baculovirus technique, a tool that has been first reported in 1983 by Smith and co-workers ^[141] and has been extensively used since then in order to obtain a variety of proteins ^[142]. Baculovirus is the term used for the virus species that affect insect cells. They normally occur in nature, but taking advantage of their infectivity and the facility to insert genetic material into insect cells, they have been genetically manipulated in order to be used as vectors. The term baculovirus expression system refers to the use of engineered insect cell virus for recombinant protein expression.

Regarding the protein properties, the glycosylation pattern produced by insect cells is less complex than the one performed by mammalian cells. For example, insect cells are unable

to add terminal sialylated N-glycans. However, insect cell glycosylation is similar enough to the one in mammals to produce an active and stable DPP IV protein, as reported in the bibliography [135, 143]. Thus, the advantage of insect cells versus mammalian cells relies on the simplicity of insect cell culture growing and the superior protein yield. [144]

The following chapter is devoted to DPP IV.

First, the set-up of protein expression in insect cells is detailed. After, the protein characterization is explained. Finally, DPP IV NMR experiments are explained.

1.1. Dipeptidyl Peptidase IV recombinant expression in insect cells

1.1.1. Production of baculovirus

Baculovirus protein expression system is based on the modification of the genetic material of the baculovirus (bacmid). In recombinant bacmids, the cDNA of the protein of interest is inserted under the control of polyhedrin, a strong promoter. This allows the expression of the target protein in the very-late phase of infection (24h-96h post-infection).

Genetic manipulation of the bacmid can be achieved by a set of techniques, but the most straightforward one is site-specific transposition (figure 1.1.).

The cDNA of the target protein is cloned into a carrier plasmid (pFastBac), surrounded by the mini-*att*Tn7 elements. The mini-*att*Tn7 sequences will permit the transposition of the cDNA into the parental bacmid by site-specific transposition. *Escherichia coli* (*E. coli*) Max Efficiency DH10 Bac cells carry a bacmid with an engineered lacZ α gene containing the mini-*att*Tn7 transposition sites, which allow the insertion of mini-Tn7 transposons. Besides, these cells contain a helper plasmid, which codifies for transposition enzymes.

After the transformation of Max Efficiency DH10 Bac *E. coli* cells with the carrier plasmid, transposase enzymes recognize the mini-*att*Tn7 sequences and insert the cDNA of the protein into the parental bacmid. This transposition leads to an ineffective expression of β -galactosidase protein (codified by the lacZ α gene). In the presence of X-Gal and IPTG, colonies where transposition has taken place will be white, compared to the blue colonies where the lacZ α gene remains intact.

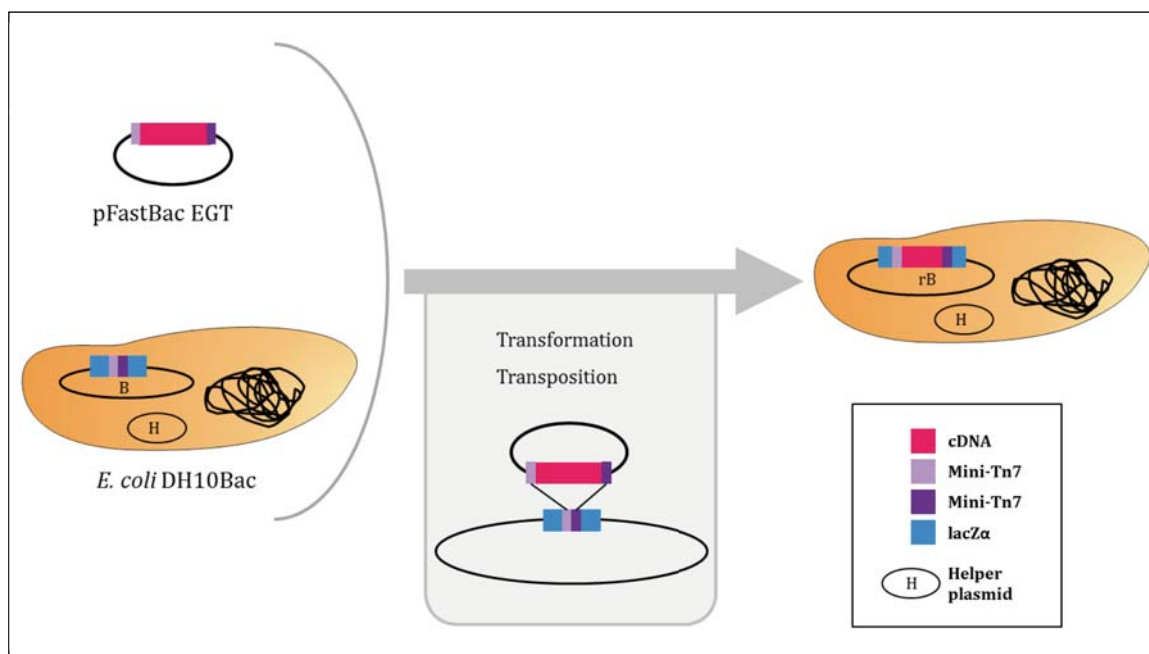


Figure 1.1.: Insertion of the cDNA (in pink) inside the vector is performed by transposition through the Mini-*att*Tn7 sites (purple) and the action of the transposase codified by the helper plasmid. Modified bacmid results in the insertion of the cDNA in the lacZ α gene disrupting the expression of β -galactosidase.

1.1.1.1. Recombinant bacmids

DPP IV cDNA from uterus leiomyosarcoma was provided in the pCMV-SPORT6 plasmid. Regarding the carrier plasmid, two were selected: pFastBac EGT-N and pFastBac EGT-C (figure 1.2.). Vectors were acquired from Prof. Arie Geerlof, from the EMBL.

The main difference between them is that pFastBac EGT-N contains the His-tag before the MCS site, whereas the pFastBac EGT-C has it after the MCS site. By this, N-terminal or C-terminal Histidine-tagged protein can be obtained. Besides, in the case of pFastBac EGT-N, a 3C protease cleaving site is placed between the His-tag and the N-terminus of the protein, allowing the possibility of obtaining a non-tagged protein, which may be useful in case the His-tag alters the structure and/or activity of DPP IV.

Both carrier plasmids contain the mini-*att*Tn7 transposition elements, the polyhedrin promoter, the EGT leader (a sequence that will lead the protein through the endoplasmic reticulum and the Golgi apparatus to the extracellular environment), the SV40 poly A sequence (to obtain a functional mRNA), ampicillin and gentamicin resistant genes, Histidine-tag (which allows affinity protein purification), a pUC origin (for plasmid replication) and a Multiple Cloning Site (MCS) with several unique sites for restriction enzymes.

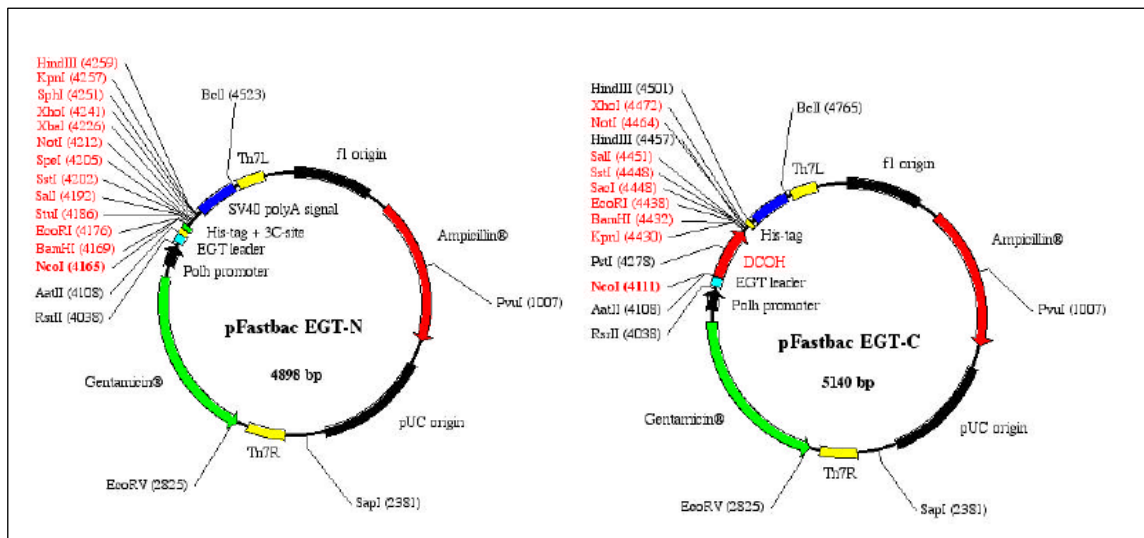


Figure 1.2.: pFastBac plasmids selected as carrier vectors for the baculovirus-insect cells DPP IV expression.

Expression of full-length DPP IV in insect cells reported in the bibliography resulted in truncation of the firsts 29 residues [143, 145]. For that reason, DPP IV sequence corresponding to residues 39 to 766 was amplified (figure 1.3.). Primers used in the DNA amplification were engineered in order to have a melting temperature around 65°C, preserve the Open Reading Frame (ORF) and contained the endonuclease restriction enzyme sites for cleavage that allow the subsequent insertion into the pFastBac plasmid. The selected restriction enzymes did not cleave DPP IV DNA sequence and cut the carrier plasmids in a single site. For pFastBac EGT-N, BamHI and SpeI were used, whereas for pFastBac EGT-C AatII and XhoI were chosen.

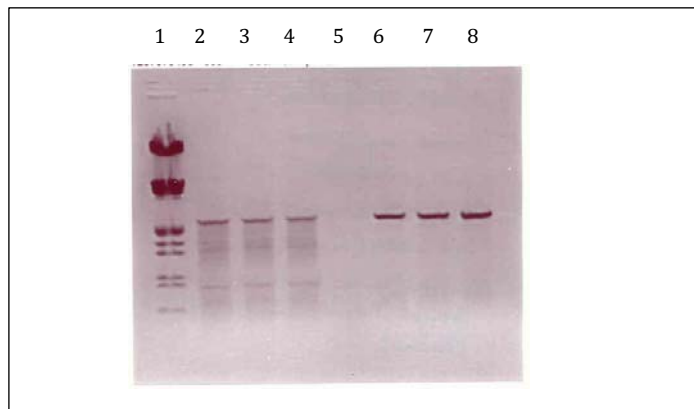


Figure 1.3.: DPP IV (39-766) sequence was successfully amplified with Pfu polymerase.
 Lane 1: Molecular Weight Marker III (0.12-21.2 kbp) (Roche)
 Lanes 2-4: hDPP IV (39-766) DNA amplification for pFastBac EGT-N
 Lanes 6-8: hDPP IV (39-766) DNA amplification for pFastBac EGT-C

An extra step previous to the cloning of the DPP IV (39-766) DNA into pFastBac was planned. It consisted in the cloning of DPP IV (39-766) into pScript AMP SK (+) plasmid. This plasmid allows the insertion of blunt DNA fragments in one step. Since the insertion point is located in the *LacZ α* gene, selection of positive colonies can be done by colour screening.

Ligation of pScript AMP SK (+) with the Taq amplified DPP IV (39-766) DNA sequences was performed and *E. coli* cells were transformed. Several colonies were picked up, and the plasmid was extracted and analyzed for DPP IV presence, both by restriction enzyme digestion followed by agarose gel and also by sequencing PCR. Unfortunately, none of the purified plasmids were positive for DPP IV presence

Then, direct cloning of DPP IV (39-766) DNA sequences into the pFastBac EGT plasmids was performed. pFastBac EGT-N was digested with BamHI and SpeI restriction enzymes and pFastBac EGT-C with Aat II and Xho I restriction enzymes. Ligation of fragments into pFastBac EGT-N and -C was performed with T4 DNA Ligase using an approximate ratio of 3:1 (insert:vector). Transformation of XL-Gold *E. coli* cells was done, but no colony was obtained.

At that point, a new strategy was planned. It consisted in the use of pGEM-T Easy plasmid (figure 1.4.) as an extra step before the final ligation into pFastBac EGT-N and -C. This plasmid contains a poly-T tail in the MCS. The tails allow the ligation of Taq polymerase amplified DNA fragments, since this enzyme leave poly-A tails. As a result, the *lacZ* gene will be interrupted, permitting the identification of gene ligation by colony colour visualization.

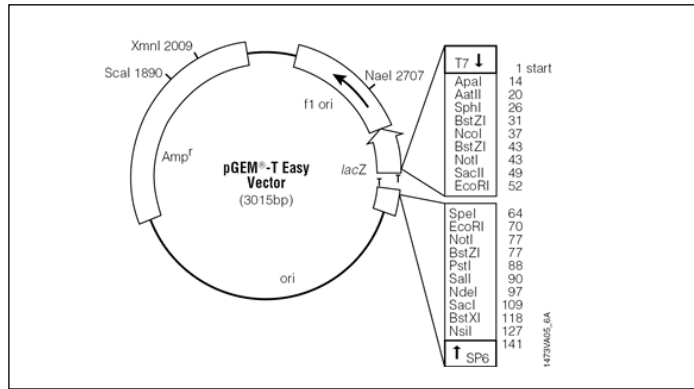


Figure 1.4.: pGEM-T Easy plasmid

After ligation of the Taq amplified DPP IV sequences into the pGEM-T Easy vector, several colonies for each construct were selected. Plasmid sequencing showed positive colonies for both constructs. Plasmids were then digested with BamHI and Spe I for pFastBac-N or Aat II and Xho I for pFastBac-C and the DPP IV DNA fragment was purified by an agarose gel.

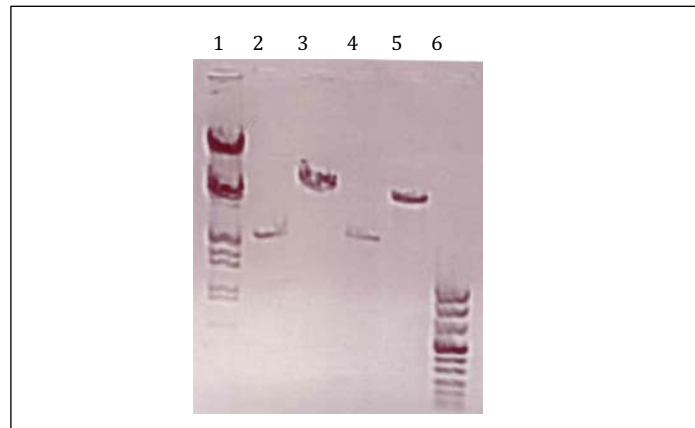


Figure 1.5.: pGEM-T Easy plasmid

- Lane 1: Molecular Weight Marker III (0.12-21.2 kbp) (Roche)
- Lane 2: hDPP IV (39-766) DNA for pFastBac EGT-N RE cut
- Lane 3: pFastBac EGT-N RE cut
- Lane 4: hDPP IV (39-766) DNA for pFastBac EGT-C RE cut
- Lane 5: pFastBac EGT-C RE cut
- Lane 6: Molecular Weight Marker VII (0.081-8.57 kbp) (Roche)

Then, ligation of DPP IV DNA sequences into their corresponding pFastBac EGT-N or -C was done. Transformation of XL-Gold *E. coli* cells gave colonies into the Agar plates. Plasmids were extracted and analyzed by sequencing PCR, demonstrating the ligation of DPP IV 39-766 residues cDNA.

Next step was transposition of the DPP IV cDNA fragment into the bacmid. Transformation of Max-Efficiency DH10 Bac cells with the donor plasmids (pFastBac EGT-N or -C containing DPP IV) was performed. In the following day, colonies were observed for blue/white colour. Only white colonies, which were unable to produce the LacZ α gene because of the transposition of the DPP IV fragments, were selected. After, the recombinant bacmids were purified and analyzed by sequencing PCR.

1.1.1.2. Recombinant baculovirus

Last step in the production of recombinant baculoviruses was the transfection of the recombinant bacmids in insect cells. For this purpose, Sf9 cells were chosen. Sf9 cells are derived from ovarian tissue of the moth *Spodoptera frugiperda* and are the main option when producing baculovirus for recombinant protein expression. Transfection was carried out in 6-well plates by incubating the recombinant bacmids with a lipidic solution that enables the entry of the genetic material into the eukaryotic cells. After, cell culture plates were incubated at 27°C for three days. During this time, baculovirus were produced by transfected cells, yielding population 1 (P1).

Steps performed until this point are outlined in figure 1.6.

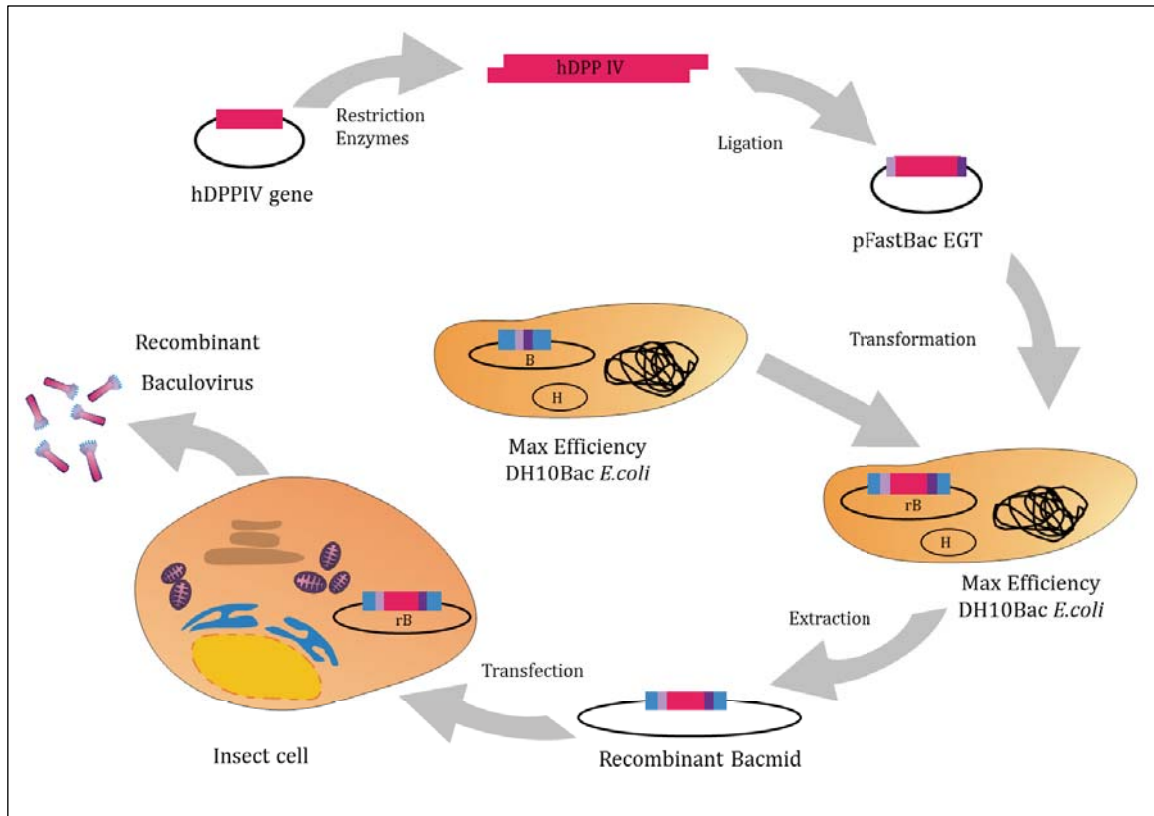


Figure 1.6.: Recombinant baculovirus production scheme.

1.1.1.3. Recombinant baculovirus amplification

In order to gain potency in virulence, which is required for an optimum protein expression, the virus population has to be amplified. Sf9 cells, grown in solution, were infected with an aliquot of P1 baculovirus. After four days, the new population of virus (P2) was collected from the media. A final amplification was performed in the same way, but infecting with a P2 aliquot in order to obtain the P3 baculovirus population. Cell viability was lower for the P2 infection (50%) compared to the one after P1 infection (86%), demonstrating the increase on virulence in the subsequent baculovirus populations.

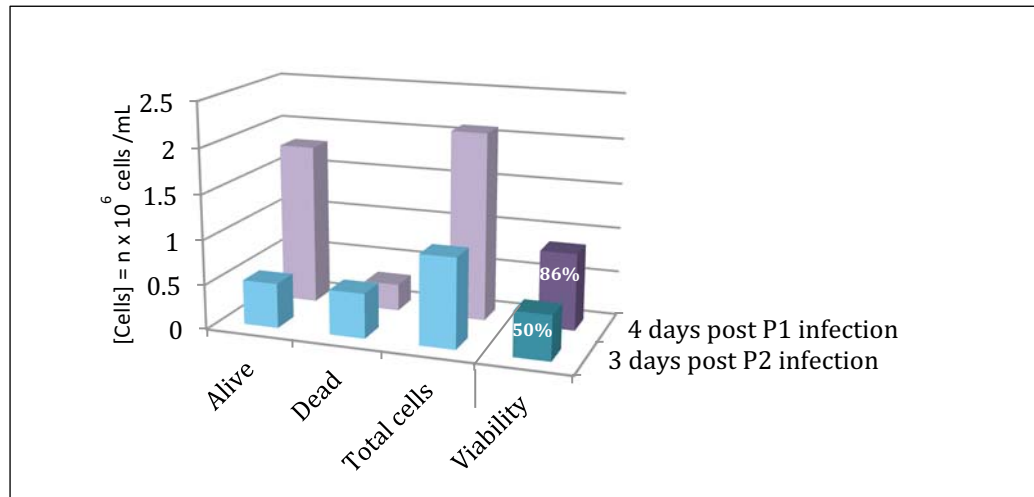


Figure 1.7.: Recombinant baculovirus amplification (Cterm His-tag DPP IV virus)

At this stage, two amplified P3 populations of baculovirus were obtained. Each P3 population corresponded to a DPP IV construct, with an N-terminal His-tag (from now on: Nterm His-tag DPP IV) or C-terminal His-tag (from now on: Cterm His-tag DPP IV).

1.1.2. DPP IV expression and purification

The general protocol for expression and purification is illustrated in figure 1.8. In each step, parameters that were modified are highlighted.

Insect cells are infected with baculovirus and incubated at 27°C in constant agitation.

The DPP IV cDNA is preceded by the EGT leader, a sequence for the protein extracellular secretion. For this reason, the media is collected for DPP IV purification. Histidine tag allows the purification by an affinity chromatography (nickel column). However, a previous step has to be performed before, since the media, where DPP IV is, contains chelating elements that would interfere with the column avoiding the binding of His-tagged protein. In order to remove these chelating elements, a step of automated diafiltration was added. The media is placed in a chamber, and with the help of a peristaltic pump, it is passed through a 0.2 µM filter. The system is completely closed, so as the medium is filtered, the same volume of buffer is added to the chamber. After several rounds all chelating elements are filtered. Then, the addition of buffer is stopped, allowing volume reduction. As a result, the final media has a small volume and will not interfere with the nickel column.

Finally, DPPIV purity, stability and activity were checked. Optimization of both, expression and purification was performed at the same time. Detailed information is given in the following sections.

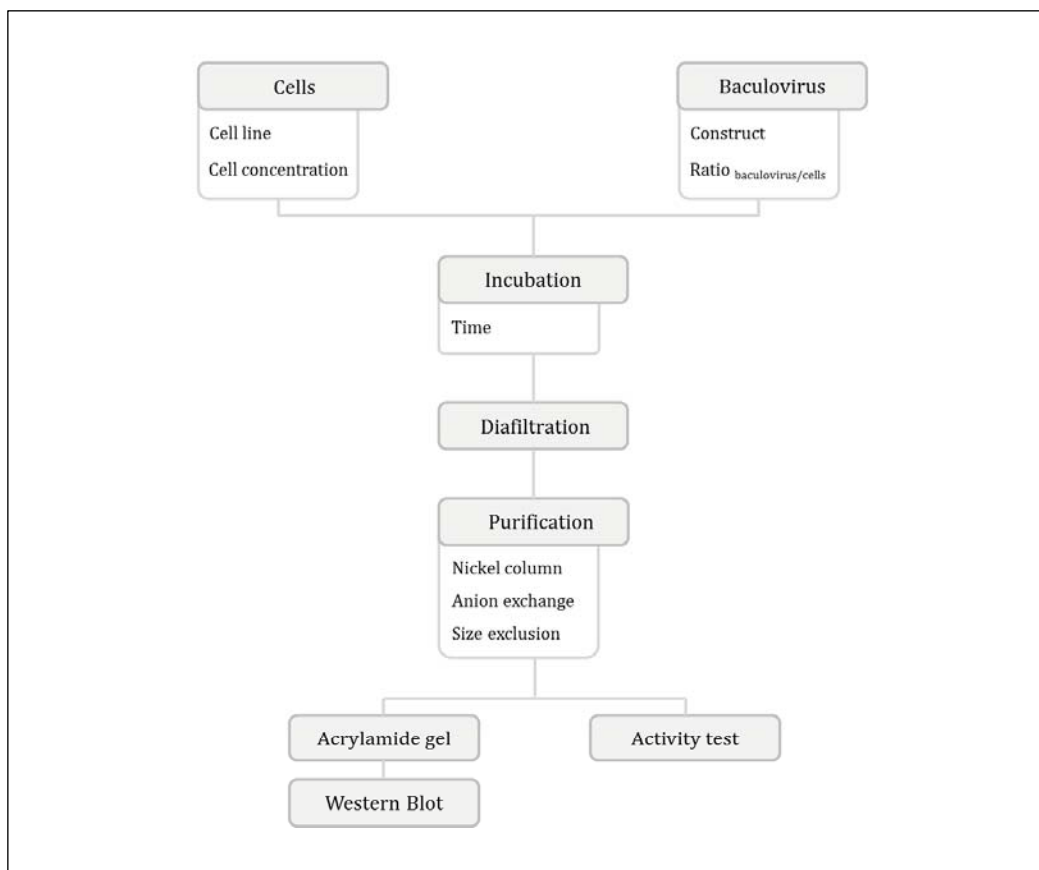


Figure 1.8.: Scheme for DPP IV expression and purification.

1.1.2.1. Evaluation and optimization of expression conditions

Optimization of the expression is crucial for standardization of the process. The evaluation of the conditions was based in the obtaining of active protein, and the best one was selected according to the maximum yield. Future NMR experiments would require a considerably large amount of protein, thus high yield is a requisite with priority.

The setting up of DPP IV expression consisted in trials where several parameters were combined in a different manner. Those parameters consisted on cell line (Sf9 or Hi5), baculovirus construct (Nterm or Cterm His-tag DPP IV), initial cell concentration ($1 - 3 \times 10^6$ cells/mL), baculovirus multiplicity of infection (MOI), and incubation time (2-4 days). (table 1.1.). The decision of the best combination was based on the results from the acrylamide gels and the activity tests.

| Code | Cell line | Construct | Initial [Cell] (cells/mL) | V _{baculovirus} /V _{culture} | V _{baculovirus} (mL)/cell x 10 ⁹ | Days |
|------|-----------|------------|---------------------------|--|--|------|
| P1 | Hi5 | Ct His-tag | 1 x 10 ⁶ | 3 mL / 1 L | 3.0 | 2 |
| P2 | Sf9 | Nt His-tag | 1 x 10 ⁶ | 1 mL / 250 mL | 4.0 | 2 |
| P3 | Hi5 | Ct His-tag | 1 x 10 ⁶ | 1 mL / 250 mL | 4.0 | 2 |
| P4 | Hi5 | Nt His-tag | 1 x 10 ⁶ | 1 mL / 120 mL | 8.3 | 2 |
| P5 | Sf9 | Ct His-tag | 1 x 10 ⁶ | 1 mL / 250 mL | 4.0 | 2 |
| P6 | Sf9 | Nt His-tag | 1 x 10 ⁶ | 1 mL / 250 mL | 4.0 | 3 |
| P7 | Sf9 | Ct His-tag | 0,8 x 10 ⁶ | 1 mL / 250 mL | 5.0 | 3 |
| P8 | Hi5 | Ct His-tag | 0,9 x 10 ⁶ | 1 mL / 250 mL | 4.4 | 3 |
| P9 | Sf9 | Nt His-tag | 1,3 x 10 ⁶ | 1 mL / 250 mL | 3.1 | 3 |
| P10 | Sf9 | Nt His-tag | 1,5 x 10 ⁶ | 5 mL / 250 mL | 13.3 | 3 |
| P11 | Sf9 | Ct His-tag | 1,3 x 10 ⁶ | 5 mL / 250 mL | 15.4 | 3 |
| P12 | Sf9 | Ct His-tag | 2,5 x 10 ⁶ | 5 mL / 250 mL | 8.0 | 3 |
| P13 | Sf9 | Ct His-tag | 1 x 10 ⁶ | 4 mL / 250 mL | 16.0 | 3 |
| P14 | Hi5 | Ct His-tag | 1 x 10 ⁶ | 4 mL / 250 mL | 16.0 | 3 |
| P15 | Sf9 | Ct His-tag | 2,2 x 10 ⁶ | 4 mL / 250 mL | 7.3 | 4 |
| P16 | Sf9 | Ct His-tag | 2 x 10 ⁶ | 5 mL / 250 mL | 10.0 | 4 |
| P17 | Sf9 | Ct His-tag | 3 x 10 ⁶ | 5 mL / 250 mL | 6.7 | 4 |
| P18 | Sf9 | Ct His-tag | 3,3 x 10 ⁶ | 5 mL / 250 mL | 6.1 | 4 |
| P19 | Sf9 | Ct His-tag | 3 x 10 ⁶ | 5 mL / 250 mL | 6.7 | 4 |
| P20 | Sf9 | Ct His-tag | 3 x 10 ⁶ | 5 mL / 250 mL | 6.7 | 4 |

Table 1.1.: DPP IV expression trials and conditions used in each one.

CELL LINE

Two cell lines were tested: Sf9 cells, from *Spodoptera frugiperda*, and Hi5, from *Trichoplusia ni*. While Sf9 are reported to be the common cell line in order to obtain and amplify the baculovirus, Hi5 cells are recommended for the expression of secreted proteins, because they allow higher yields. In both cases, DPP IV was detected by western blot. However, in all of the expressions where Hi5 cells were used (P1, P3, P4, P8, P14) a DPP IV fragment was also present (figure 1.9.).

The origin of the fragment could be attributed to either a degradation process or to an inefficient expression of the proper full DPP IV. The fact that part of the expressed protein was actually a fragment diminished the overall yield. Furthermore, it would imply a second purification step, affecting also the yield of pure dimeric DPP IV and increasing the total time of purification. On the contrary, DPP IV expressed in Sf9 cells had the appropriate size. Sf9 cells, compared to Hi5 are easier to harvest and scale-up in suspension culture, since they don't need supplements (Hi5 cells require Heparin) and are able to grow in a higher cell density, which affects the quantity of protein obtained in each expression.

Based on that evidences, Sf9 cells were selected for the following purifications.

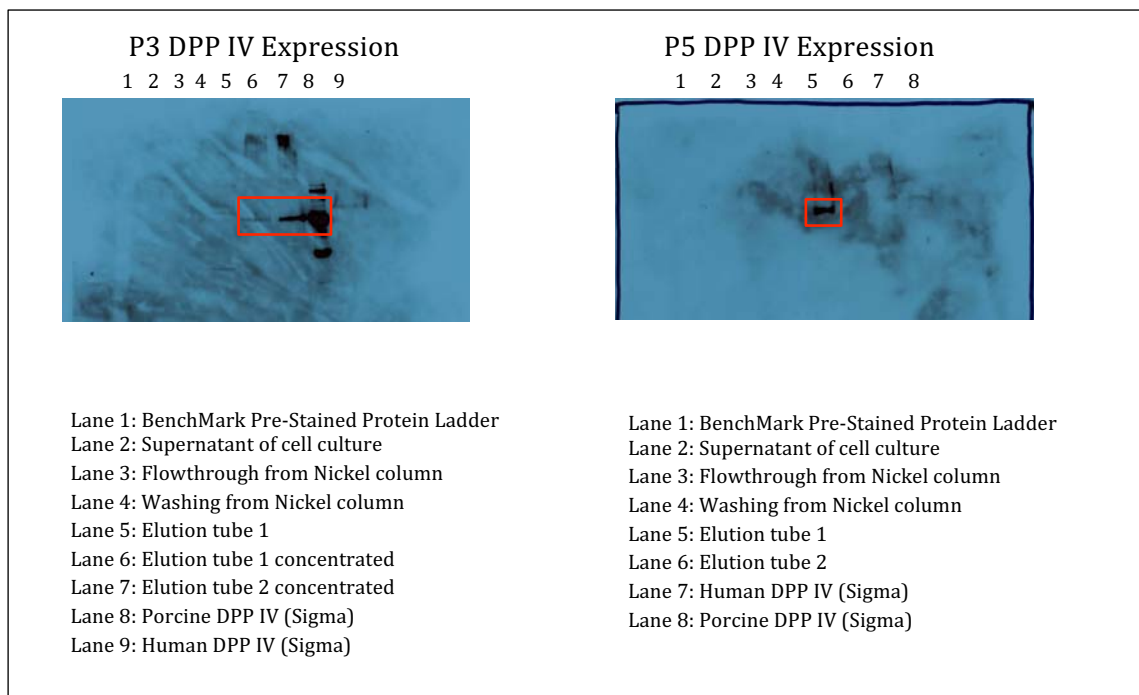


Figure 1.9.: Western blots of DPP IV expression. P3 DPP IV corresponds to the expression in Hi5 cells, while P5 is in Sf9 cells. The band corresponding to DPP IV is framed in red.

BACULOVIRUS CONSTRUCT

Respect to the construct, DPP IV could have the His-tag either in the N-terminal or in the C-terminal. The objective was to obtain both versions of the protein in order to select which one was best for the diverse techniques that were planned to test. The two constructs were nicely expressed in Sf9 cells, presenting the appropriate size. Then, each construct was tested in order to assess the proteolytic activity of the enzyme. While the Cterm construct showed activity ranging from 1 to 80 Fluorescence Units/ng_{protein}, the Nterm construct showed no activity at all (table 1.2.). It has to be noted that the Fluorescence Units was an arbitrary number used in the setting-up of the conditions and that DPP IV was not completely pure. The Nterm construct could be inactive due to an inappropriate folding, which would strongly affect the quality of the protein, making it an unviable sample for structural study. One hypothesis could be that this construct could not homodimerize. It has been previously reported that monomeric DPP IV is not active. Based on the results from the activity tests, Nterm construct was discarded.

CELL CONCENTRATION

Initial cell concentration is a parameter that is recommended between $1-3 \times 10^6$ cells/mL for Sf9 cells. On the one hand, with higher cell concentration, higher protein yield is expected to be obtained. However, crowding affects cell growing, diminishing the available nutrients, and protein yield may be lower. Cell density was in the low range in the first's expressions, and it was gradually incremented until the maximum allowed for Sf9 cells was reached. Since purification was not totally optimized, decision of the initial density was made based on the visual appreciation of the ratio between the band of DPP IV in the acrylamide gel versus the other bands in the elution lane. In figure 1.10. three conditions

are exemplified. In P11, the initial cell density was 1.3×10^6 cells/mL, in P16 2×10^6 cells/mL and in P20 3×10^6 cells/mL. Despite the fact that incubation time was lower in P11, it can be distinguished that DPP IV band was more intense with the higher cellular concentration (P20). It was concluded then that the maximum cell density, such as in P20, offered a high DPP IV yield.

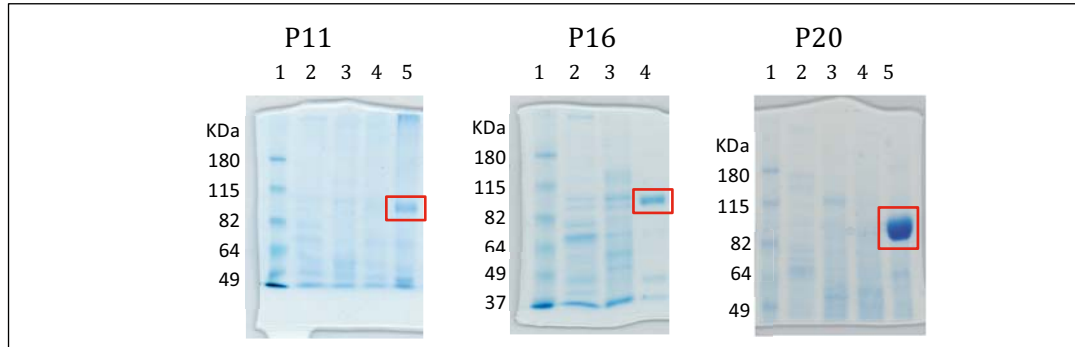


Figure 1.10.

P11 and P20

Lane 1: BenchMark Pre-Stained Protein Ladder

Lane 2: Supernatant

Lane 3: Flowthrough

Lane 4: Washing

Lane 5: Elution

DPP IV bands are framed in red.

P16

Lane 1: BenchMark Pre-Stained Protein Ladder

Lane 2: Flowthrough

Lane 3: Washing

Lane 4: Elution

BACULOVIRUS/CELL RATIO

Baculovirus multiplicity of infection (MOI) is the ratio between the number of infective particles and the number of target cells. Higher MOI imply a high concentration of virus versus cell density. Typically, a MOI higher than 8 assures a total infection of cells. In protein expression, recommended MOI values are between 1 (60% infected cells) and 5 (99% infected cells). Higher MOI assure complete infection, and imply that a larger number of cells receive at least one virus particle, allowing a simultaneous infection. However, too high MOI may cause the early death of cells before they have used the nutrients in the media in order to produce the target protein. On the contrary, if the MOI is too low, a large percentage of cells won't be infected. This population of cells will expand better and may deplete the nutrients in the media, leaving very little for infected cells. For that reason, expression trials at different ratios have to be done for each baculovirus in order to determine the ratio that gives the highest protein yield.

But before, for a baculovirus stock the number of infective particles has to be defined. Plaque assays in which either alive or dead cells are stained are the principal option to determine the number of virions per millilitre. In the case of baculovirus, the most extended technique is based on Neutral Red. In this assay, wells of Sf9 cells are infected at a set of virus dilutions. After few days, a mixture of Neutral red in agarose media is overlaid to each well. Neutral red dye stains alive cells, while dead cells remain clear. The number of clear plaques are counted and the pfu (plaques forming units)/mL of a given baculovirus stock is calculated. This number, which varies from stock to stock, describes the concentration of active particles.

Once determined the pfu/mL concentration with the plaque assay, the MOI number is defined by an ensemble of expression trials. While the plaque assay has to be done for each baculovirus stock, the MOI number is constant for a construct. It implies just a plaque assay every time a new batch of baculovirus is obtained.

Neutral red plaque assay was performed three times and in none of the trials pfu/mL could be determined. Cellular death was visible, but plaques were not clearly defined. Visual assays are useful when the virus particle is very infective, in the way that a single virion can expand rapidly to the neighbour cells and so on. However, in the baculovirus stocks that were obtained, particles were no so infective, or needed a higher incubation time to expand to the neighbourhood. This impossibilities the calculation of pfu/mL by visual techniques, but not the protein expression. For this type of slow-propagating virus, MOI used should be higher compared to the most aggressive virus.

However, since pfu/mL couldn't be calculated by visual plaque assays, MOI couldn't be determined. Then, another parameter was needed in order to assess this ratio. At that moment, two parameters were defined. One was the volume of baculovirus stock per volume of cell culture. The other was also the volume of baculovirus stock, but against the cell initial density. This second parameter was added since cell concentration was also being modified during optimization and had to be into account.

The main disadvantage of these parameters is that the optimum ratio for protein expression is different in each batch. Then it had to be slightly optimized for each baculovirus lot. In order to readjust, viability and the elution peak in the nickel column were observed.

Finally, ratios of 5-10 mL of baculovirus per 250 of Sf9 cell culture at 3.0×10^6 cells/mL were determined as the best conditions. For fresh baculovirus stocks, 5 mL was enough, but frozen aliquots lost infectivity with time, and volume had to be increased.

INCUBATION TIME

Incubation time is the last parameter that was optimized. DPP IV gene was engineered to be under the control of the polyhedrin promotor, which is activated in the very late phase of the baculovirus infection cycle (24-96 hours post-infection). First expression trials consisted in collection of cell supernatant after 48 hours. By increasing the incubation time up to 4 days, DPP IV yield was incremented. Larger incubation times were not tested as nutrients are depleted and secreted protein may be degraded.

SUMMARY OF EXPRESSION CONDITIONS

| | |
|------------------------|--|
| Cell line | Sf9 cells |
| Baculovirus construct | Cterm His-tag DPP IV |
| Cell concentration | 3×10^6 cells/mL |
| Baculovirus/Cell ratio | 5-10 mL _{P3 Baculovirus} / 250 mL _{cell culture} |
| Incubation time | 4 days |

Table 1.2.: DPP IV expression final conditions.

1.1.2.2. DPP IV purification

DIAFILTRATION

Purification of a DPP IV was envisaged as a simple process, but difficulties were encountered. In nickel affinity columns, His-tagged proteins are attached to the matrix and are eluted with competition with imidazole. First hurdle was the removal of chelating elements present in the insect cell culture media. These molecules avoid the binding of DPP IV to the column matrix, and DPP IV would be directly eluted with the flowthrough. Dilution of media would decrease their concentration, but after having tried this option, no protein was detected in the elution. Then, a diafiltration step was decided to be included before the nickel column. Diafiltration is an automated process that consists on a constant flow passing of solution A through a membrane, while solution B is added. Smaller particles than the cut-off of the membrane pass through the filter-membrane and then, are discarded. By this, chelating elements were eliminated. At the same time, since a solution B is constantly added, a buffer exchange is done. In this case, solution A was the filtered-supernatant of cell culture and solution B was the binding buffer for the affinity nickel column. Diafiltration allowed then, the removal of the chelating elements in the media and the reduction of the total volume, which simplified the following purification.

PURIFICATION CONDITIONS

| Code | Cell line | Construct | Purification | Band in gel | Purity | Activity |
|------|-----------|------------|-----------------------------|-------------|--------|-----------------------------|
| P1 | Hi5 | Ct His-tag | Ni column | Fragment | Poor | Not tested |
| P2 | Sf9 | Nt His-tag | Ni column | Correct | Poor | Not tested |
| P3 | Hi5 | Ct His-tag | Ni column | Fragment | Poor | Not tested |
| P4 | Hi5 | Nt His-tag | Ni column | Fragment | Poor | Not tested |
| P5 | Sf9 | Ct His-tag | Ni column | Correct | Poor | Not tested |
| P6 | Sf9 | Nt His-tag | Ni column + SD200 | Correct | Poor | Not tested |
| P7 | Sf9 | Ct His-tag | Ni column + SD200 | Correct | Poor | Active |
| P8 | Hi5 | Ct His-tag | Ni column + SD200 | Fragment | Poor | Active |
| P9 | Sf9 | Nt His-tag | Ni column | Correct | Poor | Inactive |
| P10 | Sf9 | Nt His-tag | Ni column | Correct | Poor | Inactive |
| | | | Ni column (His-tag cleaved) | Correct | Poor | Inactive |
| P11 | Sf9 | Ct His-tag | Ni column + SD200 | Correct | Poor | Active |
| P12 | Sf9 | Ct His-tag | Ni column + SD200 | Correct | Poor | Active |
| P13 | Sf9 | Ct His-tag | Ni column + SD200 | Correct | Poor | Active |
| P14 | Hi5 | Ct His-tag | Ni column + SD200 | Fragment | Poor | Active |
| P15 | Sf9 | Ct His-tag | Ni column + SD200 | Correct | Poor | Active |
| P16 | Sf9 | Ct His-tag | Ni column + DEAE FF | Correct | Medium | Not tested |
| | | | Ni column + SD75/SD200 | Correct | Medium | Not tested |
| P17 | Sf9 | Ct His-tag | Ni column + DEAE FF | Correct | Medium | Not tested |
| | | | Ni column + SO12 | Correct | Medium | Not tested |
| P18 | Sf9 | Ct His-tag | Ni column + SO6 | Correct | Medium | Not tested |
| P19 | Sf9 | Ct His-tag | Ni column* | Correct | Good | Not tested |
| P20 | Sf9 | Ct His-tag | Ni column* | Correct | Good | Active |
| | | | | | | 2 Units/mg _{DPPIV} |

Unit = $\mu\text{mols product/min}$

Table 1.3.: DPP IV expression trials and purification conditions used in each case.

After diafiltration, affinity nickel column was done. However, DPP IV was not fully pure. Then, size-exclusion column was used (table 1.3.).

Superdex 200 was selected as first option because its range (Between 10 and 600 KDa). Unfortunately, it did not allow the purification of DPP IV from impurities. (Figure 1.11. P15)

Then, anion-exchange purification was used. Elution was performed with growing salt concentration. Different gradients were tried and mainly all impurities were removed, except for a small protein, as seen in the acrylamide gel (figure 1.11. P17)

First attempts to remove this smaller protein with size-exclusion columns failed, although a set of columns was used (combination of Superdex 75 and Superdex 200, as well as Superose 12 and Superose 6). It was concluded that, despite being small in size as seen by the acrylamide gels, the impurity was a bigger aggregate in native state and it was eluted with DPP IV in size-exclusion columns.

Since any of the columns allowed the obtaining of pure DPP IV, the conditions used in the affinity column were fine-tuned. Nickel purification was optimized, and by increasing the imidazole concentration in the washing step from 20 to 50 mM, impurities were finally removed in one single step and DPP IV was eluted in a good purity standard (figure 1.11. P19). Purifications where these conditions were used are marked with "*" in table 1.3.

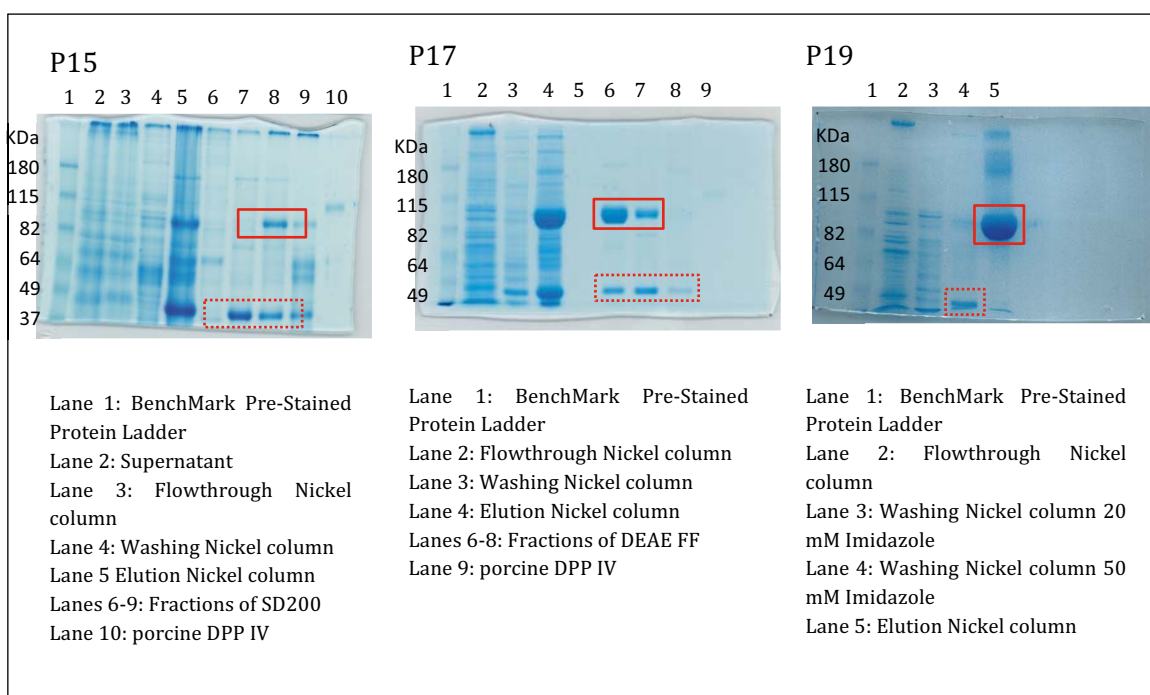


Figure 1.11.: Acrylamide gels belonging to expression trials P15 (Nickel column + SD200), P17 (Nickel column + Anion exchange) and P19 (Nickel column, optimized imidazole concentration). DPP IV band is marked with a red solid frame, while impurities are framed in dotted frames.

SUMMARY OF CONDITIONS

Column: Nickel column

Binding Buffer: 50 mM Tris-HCl; pH= 7.9; 400 mM NaCl

Washing Buffer: 50 mM Tris-HCl; pH= 7.9; 400 mM NaCl; 50 mM Imidazole

Elution Buffer: 50 mM Tris-HCl; pH= 7.9; 400 mM NaCl; 200 mM Imidazole

1.2. Dipeptidyl Peptidase IV characterization

After the optimization of the expression and purification of DPP IV, the protein was characterized. Dimeric state, activity, mass and glycosylation pattern of DPP IV were assessed.

1.2.1. Quaternary structure

Dimeric state of the protein was proven by size-exclusion column (Superdex 200). The chromatogram showed one peak corresponding to the approximate size of a DPP IV dimer (220KDa).

In order to complement size-exclusion chromatography, native gel electrophoresis was done. It consists on running proteins in non-denaturing conditions. Buffers are SDS free, sample buffer does not contain DTT or any other reducing agent and samples are not heated, thus, proteins maintain their structure. In native gel electrophoresis, then, proteins are separated depending on their cross-section area. Then, monomers, dimers and higher order species are separated.

In the case of DPP IV, a set of acrylamide percentages, ranging from 4% to 7.5%, was used, and in all of the native gels only one species was identified.

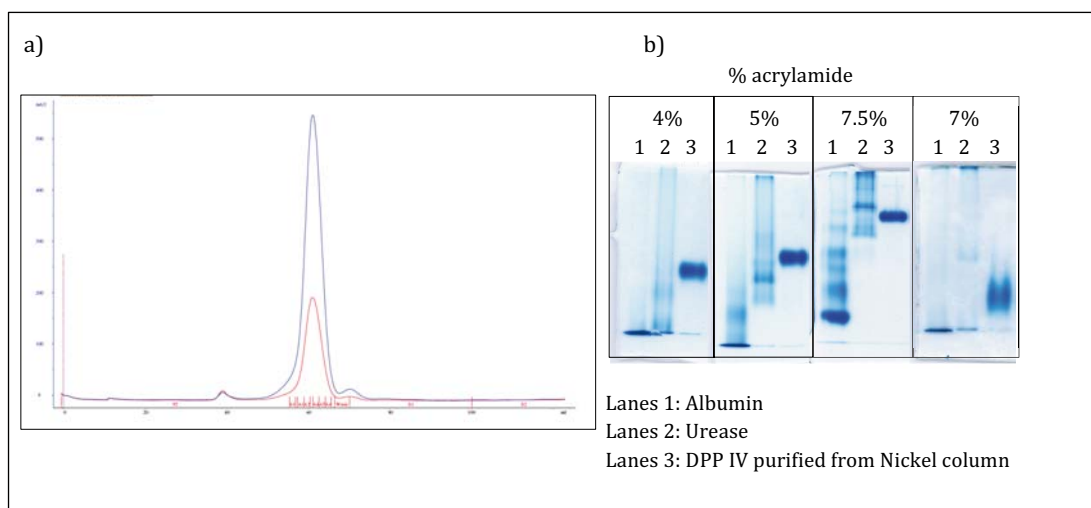


Figure 1.12.: a) SD200 chromatogram of pure DPP IV; b) Native gel electrophoresis.

1.2.2. Activity

As observed in the preliminary expression trials, DPP IV presented proteolytic activity. It was measured as the fluorescence produced upon cleavage of substrate H-Gly-Pro-AMC, releasing the aminomethylcoumarin group (AMC), which emits fluorescence at 485 nm when excited at 360 nm. In order to evaluate the affinity of DPP IV for its substrate and the specific activity of the enzyme, kinetic experiments were performed.

First kinetic consisted on a subset of enzymatic assays where concentration of both, substrate and protein were variable. Fluorescence values were represented against time, giving as a result the progress-curve plot. From that representation (figure 1.13.), enzyme concentration and time linearity were determined. The enzyme concentration selected

was the one with fluorescence values significantly different from the blank sample. As well, this concentration had a fair window in the linear time to calculate inhibition constants in the future experiments. Each batch of DPP IV had a different optimum concentration, but all of them ranged between 0.5-5 μM .

The linear time was more or less constant in all of the expressions, and it was set to 20 minutes.

Once determined these two parameters (DPP IV concentration and linear time), velocities for each substrate concentration were calculated. Representation of velocities in front of substrate concentration (Michaelis-Menten plot) allowed the estimation of K_M value. This value, which is an intrinsic for a given protein and substrate pair, is related to the affinity between the two molecules, and therefore, is constant among all batches. K_M corresponds to the substrate concentration that affords half of the maximum velocity. In the DPP IV/H-Gly-Pro-AMC system, K_M was defined at 45 μM .

Finally, specific activity was calculated with a second kinetic, where only one condition, with fixed DPP IV and substrate concentration is required. DPP IV concentration was the obtained in the first kinetic experiment, while substrate concentration was 50 μM , since values around the K_M are recommended. Apart from this condition, a standard curve of AMC was included in the assay. Representation of fluorescence versus μmols of AMC was fitted to a one-order equation. Then, fluorescence values of the enzymatic reaction at time 0 as well as at 20 minutes, were measured. After, transformation of fluorescence values to μmols of AMC was done with the standard curve equation and velocities ($\mu\text{mols}_{\text{Substrate}}/\text{min}$) were obtained.

Velocity is related with specific activity by mass correction. First, one $\mu\text{mol}_{\text{Substrate}}/\text{min}$ was defined as one Unit, and the specific activity was calculated as the $\text{Units}_{\text{Substrate}}$ per mg of DPP IV. Each batch of DPP IV had its own specific activity, and it ranged from 1-5 $\text{Units}_{\text{Substrate}}/\text{mg}_{\text{DPP IV}}$.

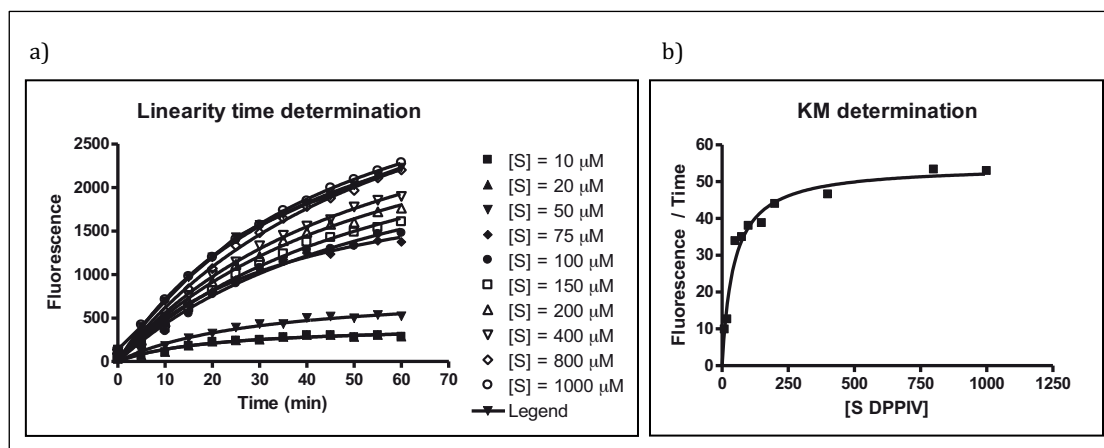


Figure 1.13.: a) Determination of lineal time for DPP IV proteolytic activity; b) K_M determination of H-Gly-Pro-AMC for DPP IV.

SUMMARY OF DPP IV ACTIVITY

DPP IV concentration: 0.5-5 μM

Linearity time: 20 min

$K_M = 45 \mu\text{M}$

Specific activity = 1-5 $\text{Units}_{\text{Substrate}}/\text{mg}_{\text{DPP IV}}$

1.2.3. Mass of the protein

1.2.3.1. Electro-spray ionization (ESI)

First attempt for mass estimation was based in the observation of the full denatured DPP IV. Heating of DPP IV was not a possible denaturing treatment, since it provoked the protein aggregation. For that reason, sample was solubilized with acetonitrile (ACN) and formic acid, with a final concentration of 45 μ M. Then, DPP IV was purified by liquid chromatography, with a Biosuite pPhenyl 1000 10 μ m Reversed-Phase Column and analyzed in an orthogonal acceleration time-of-flight mass spectrometer. Although several trials were performed, only one condition gave a mass around the expected for DPP IV. A mass of 100698.04 \pm 49.85 Da was observed at a retention time of 24.573 minutes. Unfortunately, resolution of mass peaks did not achieve the quality standards required for an optimum characterization. Similar experiments performed in the lab with a protein with the same amino acidic length gave good results, which suggested that the lack of resolution in the case of DPP IV was not related to its large size. In the case of this second protein, it was produced in *E. coli*, and therefore it was not glycosylated. On the contrary, DPP IV, as it is expressed in insect cells, is expected to be glycosylated, and as consequence the sample contains a mixture of DPP IV species glycosylated in a slightly different manner. Such a heterogeneous sample would be a challenge to fully characterize by mass spectrometry.

Then, experiments in native state were carried out. Denatured DPP IV was very prone to precipitate, and it was believed that sample loss may occur during this process. Thus, by keeping the protein in the native state, higher concentration would be available for ionization and further detection. On the other hand, DPP IV would be dimeric and the size of the sample would be around 220 KDa, which is quite large for mass spectrometry detection. Unfortunately, after several trials, none of the experiments afforded a good quality result.

1.2.3.2. Matrix-Assisted Laser Desorption/Ionization (MALDI)

Since ESI did not allowed DPP IV detection, MALDI was used.

Two DPP IV were selected: 0.5 and 0.25 mg/mL.

Matrix used was α -Cyano-4-hydroxycinnamic acid (ACH), and two protein:matrix proportions were tested: 1:1 and 0.5:1.5 (v/v).

Finally, samples were assayed in the presence and absence of 10% formic acid.

Unfortunately, DPP IV was not detected in any of the conditions.

1.2.4. Glycosylation pattern

1.2.4.1. Glycosylated DPP IV detection

First, dyeing of acrylamide gel with Schiff's reagent demonstrated that DPP IV was glycosylated. This evidence was observed both in SDS and non-SDS acrylamide gels (figure 1.14.).

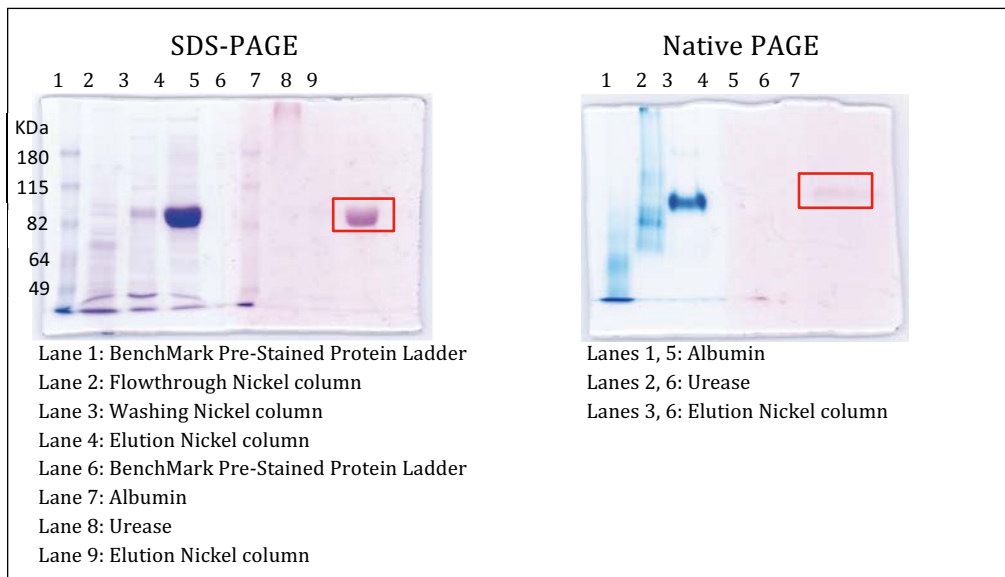


Figure 1.14.: Denaturing and non-denaturing acrylamide gels of DPP IV. For each one, gel on the left is stained with Coomassie blue, and gel on the right is stained with Schiff's reagent. For this second dyeing, DPP IV band is framed in red.

1.2.4.2. Glycosylated DPP IV characterization: experiment planning

Despite DPP IV was well reported in the bibliography, very little was known about DPP IV glycosylation. Characterization of human DPP IV had concluded that it presented a complex glycosylation with terminal $\alpha(2-6)$ -*N*-acetylgalactose, $\alpha(2-6)$ -*N*-acetylgalactosamine and $\alpha(2-3)$ -galactose-linked sialic acid. [137] A comparative study between DPP IV expressed in Sf9 and chinese hamster ovarian (CHO) cells confirmed a different pattern, mainly by the presence of high mannose structures in Sf9 cells versus the high and complex glycosylation of CHO cells. [143] These studies were performed by the incubation of the protein with lectins, a qualitative technique that does not allow the identification of the modified residues. By X-ray crystallography, the nine potential glycosylation points of the soluble DPP IV form were glycosylated when protein was expressed in insect cells. [53] Unfortunately, modeling of the sugar moieties was not possible, most likely because of the poor ordering in the crystal or because glycosylations are heterogeneous between proteins in the crystal. [138]

Our main objective was to characterize both, the point and the type of glycosylation, and the technique that offered the possibility to assess it was mass spectrometry. A previous publication had analyzed both the N-glycosylated peptides from chymotrypsin digestion of DPP IV and the N-glycans. [146] In the case of the N-glycosylated peptides, identification of modified asparagine residues was done by MS/MS analysis. Deglycosylation of peptides

and proteins with the enzyme PNGase F, removes the sugar moiety but it also produces an asparagine to aspartic modification. Taking into account this change seven out of nine potential N-glycosylation points (NXT/S sequences where X ≠ Pro, Asp) were detected. The other two peptides were not detected and glycosylation couldn't be confirmed nor discarded. With respect to the N-glycans, sugar moieties obtained by PNGase treatment were analyzed by MALDI-MS, which allowed the identification of the structures. However, no connection between the glycosylation point and structure was done. Besides, no information about the O-glycosylation in DPP IV was known to date.

In order to pursue our objective, a strategy of differential enzymatic or chemical deglycosylation with trypsinization of DPP IV was designed. The advantage of this strategy relied in the application of combinations of glycosidases, which left different patterns in the N- and O-glycosylated peptides. The scheme of sample treatment is showed in figure 1.15.

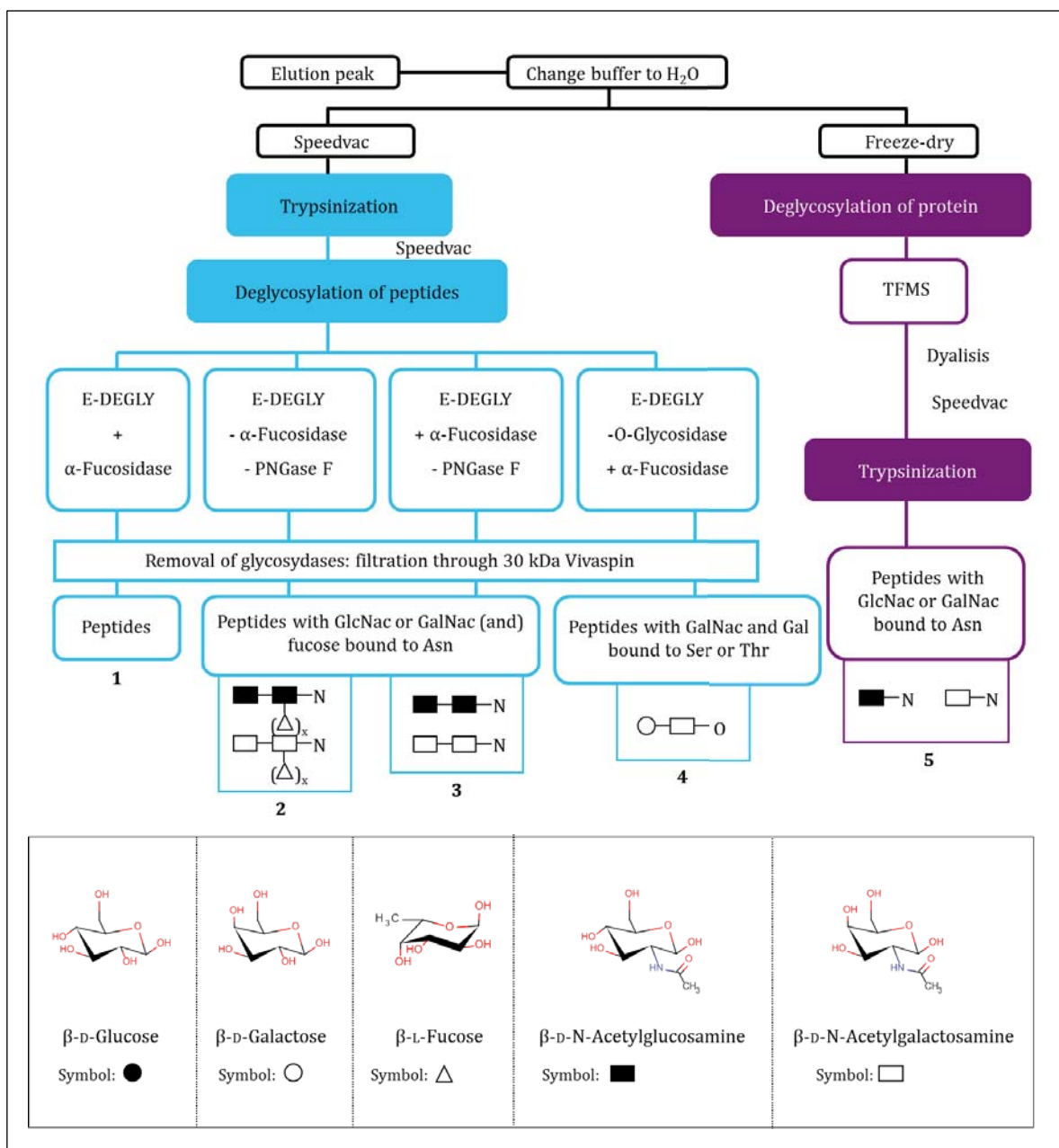


Figure 1.15.: Scheme that summarizes the strategies followed in order to describe DPP IV glycosylation.

1.2.4.3. Glycosylated DPP IV characterization: sample treatment

The E-DEGLY kit contained the following glycosidases: PNGase F, O-Glycosidase, α -2(3,6,8,9)-Neuraminidase, β -1 \rightarrow 4-Galactosidase and β -N-Acetylglucosaminidase. To this set of enzymes, α -fucosidase was added. In case 1, Deglycosylated Tryptic Peptides (DTP1), would present no sugar moiety and the only modification would be the asparagine to aspartic modification. In DTP2 the absence of PNGase F would leave two units of β -N-Acetylglucosamine or β -N-Acetylgalactosamine with fucose moieties. DTP3 would be identical to DTP2 with the exemption that fucose moieties would not be present thanks to the addition of α -fucosidase. DTP4, because of the lack of O-Glycosidase, would still carry the sugar modifications in Serine or Threonine. Finally, case 5 consisted on a chemical deglycosylation with trifluoromethanesulfonic acid (TFMS). This treatment degrades all glycosidic moieties except for the innermost Asparagine-linked β -N-Acetylglucosamine or β -N-Acetylgalactosamine.

1.2.4.4. Glycosylated DPP IV characterization: sample analysis

After sample treatment, it was analyzed by means of Electro-spray ionization (ESI) in a Fourier transform ion cyclotron resonance instrument (FT-ICR). This instrument allows a better resolution compared to traditional spectrometers, which in the case of a complex case like a heterogeneous protein is crucial. It also offers the possibility of performing Collision Induced Dissociation (CID) in the Linear Ion Trap (LIT), which permits the MS/MS identification. In figure 1.16, the scheme of analysis is depicted. Analysis of samples was performed in collaboration with the Mass Spectrometry core facility of IRB Barcelona.

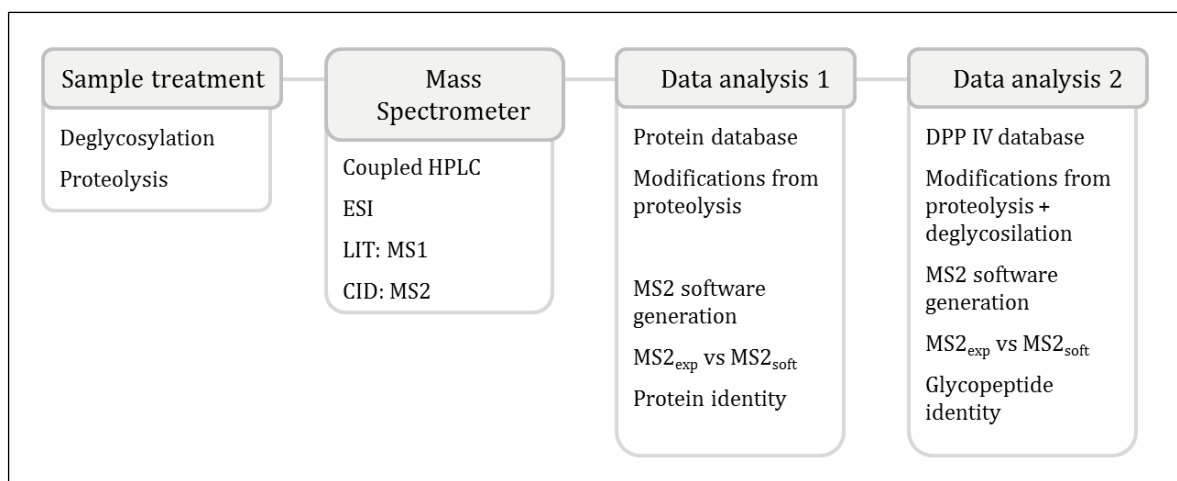


Figure 1.16.: Analysis of MS data.

DATA ANALYSIS 1: Protein identification

First, DTP were purified in the MS-coupled HPLC and peptides were ionized. Ions with charge equal or higher than 2 and with a steep intensity over m/z were trapped in the LIT (only the 6 ions corresponding to the more intense peaks). CID was done to the ions and MS2 was detected.

In a parallel way, data analysis 1 was carried out. MS1_{exp} were compared to a Protein database (Uniprot or SwissProt) using a search engine (SEQUEST or Mascot, respectively),

and taking into account the modifications that peptides suffer during proteolysis. In this case, DPP IV was digested with Trypsin (which cleaves after lysine and arginines). In trypsinization cysteine residues suffer carbamidomethylation (which was set as a static modification), methionines may be oxidized (which was set as a dynamic modification) and N-terminal glutamines of peptides can cyclize to piroglutamate (which was set as dynamic). The software calculates the $MS1_{soft}$ of modified peptides from all the proteins contained in the database. If mass of $MS1_{exp}$ is found to be equal to $MS1_{soft}$, the protein is identified. Then, the software proceeds to calculate the $MS2_{soft}$ for each $MS1_{soft}$. Comparison of $MS2_{exp}$ to $MS2_{soft}$, is used in order to confirm the peptide sequence. False Discovery Rate (FDR) is a percentage number that informs about the similarity between $MS2_{exp}$ and $MS2_{soft}$. When FDR is 99% or higher, the peptide identity is confirmed. If FDR is lower than 95% identity is not reliable.

By data analysis 1, sample was identified as DPP IV, as expected, with both search engines. Slight differences in coverage were observed using SEQUEST or Mascot, and were attributed to the database.

DATA ANALYSIS 2: Post-translational modifications identification

Then, second data analysis started. It allows the identification of post-translational modification in the protein. In this case, the workflow was exactly the same but with minor differences. First, the database is reduced to the protein identified in the previous data analysis, in this case DPP IV. Second, only SEQUEST was used as a search engine. Third, and more important, other modifications were added into the software in order to calculate MS1 of peptides. Those modifications were customized for each sample treatment (1 to 5) depending on the enzymatic or chemical treatment they had suffered. These modifications were set up as dynamic.

In DTP 1, the modification produced by the sample treatment was a substitution of asparagine to aspartic. With this analysis condition three modified peptides were detected with an $FDR \geq 99\%$. These peptides contained Asn 150, Asn 321 Asn 520 respectively. In DTP 5, the modification was an extra mass corresponding to an asparagine-linked N-Acetyl Hexose. Two peptides were identified, containing Asn 150 and Asn 520. These residues were also found in the first sample treatment analysis, reassuring the glycosylation of these residues. With respect to the other treatments, no coincidence between $MS2_{exp}$ and $MS2_{soft}$ was obtained. In the case of treatment 4, the one related to the O-glycosylations, serine 59 was found to be modified with an Acetyl Hexose, but the FDR value was lower than 90%.

ALTERNATIVE SAMPLE TREATMENT

The fact that no glycosylation modifications were found did not imply that there were no such modifications. The main obstacle was that protein coverage was extremely low (20-30%). In order to increase it, it was planned to perform a double proteolytic digestion with Trypsin and Glutamyl endopeptidase (Glu-C), which cleaves after glutamic and aspartic residues. Treatment with a second proteolytic enzyme allowed increasing the coverage up to 40%. Deglycosylated Tryptic/Glu-C peptides (DTCP) ranging from 1 to 5 potentially had the same glycosylations as DTP 1 to 5, and therefore, software treatment was analogous (with the exemption of the extra modification corresponding to the Glu-C). Unfortunately, no new glycosylation modifications were observed with a FDR higher than 99%. Potential

serine and threonine residues susceptible of being glycosylated, as observed with FDR lower than 95% were: Ser 59,Thr 186,Ser 669, Thr 570, Ser686, Thr687, Ser690.

CONCLUSIONS

Results obtained suggested that sample preparation should be modified in order to increment coverage. Enrichment of peptides and selection of other proteolytic enzymes would help to increase the number. However, these experiments required investing time and protein. Taking into account the results that could have been obtained in a thorough study against the time that was needed, it was decided to preferentially intend the efforts for another project.

1.3. DPP IV dynamism

The dipeptidyl peptidase IV structure is characterized by X-ray crystallization.^[53] A search in the Protein DataBase of human DPP IV structures gives a total of 79 entries. All of them have been obtained by X-ray crystallization, and two of them offer a good resolution between 1.5 and 2 Å. Those structures provide essential information in order to understand the behaviour of DPP IV inhibitors and the interaction of the protease with other proteins. Key residues in the contact surfaces are identified and new molecules with higher affinity for DPP IV can be engineered in base of this information.

However, X-ray provides rigid and static structures, a frozen state of the protein in a particular moment. It is nowadays well accepted by the scientific community that proteins are dynamic and that this mobility influences biological processes by changes in the affinity and binding time window.^[147]

In the case of DPP IV, there is no study about a possible dynamism. However, previous studies on prolyl oligopeptidase (POP), a closely related protein to DPP IV, have demonstrated the presence of POP conformers.^[148] Due to the high structure homology between these proteins it was hypothesized that DPP IV could feature dynamism as well.

1.3.1. Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is a powerful technique that allows the study of proteins in solution. NMR study of POP had revealed a dynamic equilibrium between two conformers. Besides, the equilibrium can be shifted to the less dynamic conformer by addition of POP inhibitors.^[101] Given the potential of NMR to provide dynamic information, and the positive results obtained with POP, this technique was selected for the study of DPP IV dynamism.

1.3.1.1. NMR applied to protein study

NMR is based on the magnetic properties of nuclei atoms. In the field of proteins, NMR has been used in the elucidation of tridimensional structures^[149], as well as in the study of protein interactions (protein-protein and protein-ligand).^[150]

Upon binding of protein with an interaction partner, the protein structure experiences changes, either in a localized or global manner. Nuclei properties are altered by these structural modifications, and NMR allows the observation of changes by the measure of parameters such as coupling constants or chemical shifts in 'magnetically active' nuclei (typically: ¹H, ¹⁵N and ¹³C).

NMR structure study can be focused in the observation of the protein (receptor) or the partner (ligand). In the first case, proteins have to be isotopically enriched. The most common expression system for stable expression of labeled proteins is *E. coli*, since it is versatile and yields are normally high. However, as a prokaryote, *E. coli* is unable to produce post-translational modifications. In order to overcome this hurdle, a series of labeling techniques in eukaryote has been developed.^[151] In yeast, *Pichia pastoris*^[152] and *Kluyveromyces lactis*^[153], are common species, being the second one superior in terms of protein expression reproducibility. Expression of labeled proteins is also possible in insect^[154] and mammalian cells.^[155] The more similar the expression system, the more

structural resemblance will have the labeled expressed protein to the human one. However, yield is inversely affected.

Recently, cell-free (also named as *in vitro*) protein synthesis has drawn attention for its application in NMR. Lysates of *E. coli* complemented with glutathione buffer and the enzyme disulfide isomerase, has been successful in the expression of properly folded ^{15}N labeled eukaryotic proteins.^[156] Also, lysates of wheat germ have been developed.^[157] However, cell-free technique is still not suitable for higher post-translational modifications.

Besides of the isotopically labeling of proteins, NMR requires high receptor concentrations. In some cases, proteins tend to aggregate, which complicates, and even avoids, their study by this technique.

Another aspect related to protein stability is the time needed for NMR experiments. Typically, these experiments are performed during several hours or days, in which the protein may precipitate.

In order to overcome the stability problems, potent high field magnets (800 and 900 MHz) and cooled cryogenic probes.

All these advances in the application of NMR for the study of proteins are patent in bibliography. The technique has been useful in structure elucidation^[158], either as a unique tool^[159] or in combination with other, such as molecular dynamics^[160] or circular dichroism spectroscopy.^[161], in protein dynamism investigation^[162-165], in the study of post-translational modifications effects in structure and binding properties^[166, 167], and in ligand binding,^[168] among many other examples.

1.3.1.2. Study of large proteins by NMR

Study of large proteins by NMR represents a challenge.

First hurdle is related with signal lifetime during acquisition. Signal relaxation, which ultimately causes signal lost, is a phenomena produced by the 'active' nuclei. Thus, with a large number of nuclei interacting among them, a faster relaxation is observed. For that reason, large proteins relax faster than smaller ones, resulting in low signal sensitivity.

Sample preparation can be tailored to avoid fast relaxation. This is accomplished by sample perdeuteration. ^2D is an 'inactive' nuclei in contraposition to ^1H , and thus, protein perdeuteration allows reduction of nuclei interaction and subsequent fast relaxation.

Besides, to counterbalance relaxation in large proteins, a number of experiments have been developed. Transverse relaxation optimized spectroscopy (TROSY)-based pulse sequences represent the main strategy for the study of these proteins.^[169] The high sensitivity of TROSY experiments relies in the selection of the slow-relaxing signal component, in spite of the slow and fast component interconversion performed by traditional HSQC experiments.

A second hurdle associated with the study of large proteins by NMR is the large number of overlapping signals in the spectra. Interpretation of crowded spectra is extremely difficult or even impossible.

As a solution, samples are selectively labeled. Several selective labeling methods are available in *E. coli*, and are based in the administration of isotopically labeled amino acids

or precursors to auxotrophic cell strains (cells where a particular amino acid biosynthetic pathway is disrupted). Thus, cells are grown in the presence of the corresponding labeled molecule that is incorporated in the protein. For example, supplementation of Leu/Val-auxotrophic *E. coli* cells with Methyl-¹³C- α -keto isovaleric acid, leads to Methyl-¹³C-Leu and Methyl-¹³C-Val labeled protein, and ¹⁵N-indole is used for ¹⁵N-Tryptophan labeling of proteins. Signal number is considerably reduced by selective labeling, but since residue type is distributed along the tridimensional structure, the subset of NMR peaks are, generally, a representation of the overall protein structure.

Finally, for spectra interpretation, conventional methods used in NMR for residue-signal assignment cannot be translated to large proteins. Thus, signals are assigned by specific-site mutagenesis, which affords a signal disappearance in the spectra.

Regarding the tridimensional structure determination of large proteins, it is practically impossible by means of NMR. The backbone signal overlapping is inherent of these types of proteins, and then, backbone assignment cannot be done.

However, protein dynamism can be described in large proteins. Conformational changes, which occur in the millisecond (ms) time-scale (slow), affect the relaxation. Thus, protein dynamism can be monitored by changes on the transverse relaxation parameter (T_2). It has to be specified that protein should be perdeuterated in order to avoid fast signal relaxation.

1.3.2. DPP IV study by NMR

DPP IV is around 220 KDa, and each monomer contains 728 residues (expressed form is 39-766). Since a uniform labeling would provide incoherent information, selective labeling was then, the solution.

Methyl-¹³C-methionine labeling has been previously used for the study of dynamism in proteins with good results. In the case of POP, this strategy afforded high quality in the signal to noise ratio and well-resolved signal distribution in the spectra window. Regarding DPP IV, it has 14 methionine residues, distributed along the structure.

Given 1) the positive outcome of this labeling in POP study by NMR, 2) the fact that POP and DPP IV have similar tridimensional structures, and 3) that with only 14 methionine residues, spectra overcrowding was not expected, it was decided to apply the same labeling methodology in the DPP IV study.

However, while for *E. coli* there are several methodologies, based in the addition of labeled molecules in the culture, labeling strategies in insect cells consist on the use of labeled media, either custom or house-made. The first group of media is expensive, and the second is difficult to optimize and protein yields vary from batch to batch.

Therefore, for the study of DPP IV by NMR, an intermediate option was selected. Depleted Sf-900 II SFM is an insect cell media that contains all the nutrients needed for optimum protein expression except for Methionine and Cystine. Thus, media is completed by the

addition of only two reagents, which avoids batch to batch incoherences. This depletion allowed the supplement with labeled methionine instead of the cold aminoacid.

The main disadvantage of insect cells is that perdeuteration is not possible. Then, relaxation experiments, which would afford information about a possible conformational equilibrium, cannot be performed. As a solution, it was proposed to study the Chemical Shift Perturbation (CSP) of Met signals in the spectra upon binding of small molecules. As observed in POP, specific spectra signals suffer a displacement in the chemical shift and/or a change in the signal intensity. Subsequent site-specific mutagenesis experiments sheds light on the identification of flexible protein regions.

1.3.2.1 Methyl-¹³C Methionine selective labeling

A selective labeling consisting in Methyl-¹³C Methionine was selected for the NMR study of DPP IV. This labeling is achieved in insect cells by the use of depleted Sf-900 II SFM supplemented with ¹³C-Methyl Methionine. Complete media was obtained by pre-warming at 37°C and then supplementing with 1 g/L of ¹³C-Methyl Methionine and 0.150 g/L of L-Cystine. Sf9 cells were grown in the usual culture media (Sf-900 III SFM) and the day of infection, cells were harvested and resuspended in the labeled media at a final concentration of 3 x 10⁶ cells /mL. Then, baculovirus was added. Incubation, diafiltration and purification were performed in the same way as the non-labeled protein. After elution, DPP IV buffer was exchanged to 50 mM D11 Tris pH = 7.9, 50 mM NaCl in 90% D₂O/10%H₂O. Protein was finally concentrated until reaching a value close to 100 μM.

1.3.2.2 Methyl-TROSY experiment

DPP IV, as a large protein, relaxes quite fast compared to small or medium-sized proteins. For that reason, the Methyl-TROSY experiment [170] was selected.

Methyl-TROSY experiment consists on a HMQC sequence optimized for large proteins. Due to a differential magnetic transference, the interference between intra-methyl dipolar interactions is reduced, and then, sensitivity is increased. [170]

Experiments were performed at 600 and 800 MHz Bruker Digital Avance NMR instruments fitted with triple-resonance z-axis gradient cryoprobes.

Before and after each Methyl-TROSY experiment, monodimensional proton NMR spectra were recorded in order to evaluate sample integrity. Data were processed with Topsin 2.0.

As observed in the spectra of figure 1.17., signals were nicely observed for labeled DPP IV, in both magnetic field instruments, B800 and B600.

From the potential 14 methionines, 11 signals are detected. The table inserted in figure 1.17. contains the chemical shifts of the ¹³C-Methyl Methionine signals of DPP IV.

| Signal number | $\nu(F2)$ [ppm] | $\nu(F2)$ [ppm] |
|---------------|-----------------|-----------------|
| 1 | 2,03 | 13,59 |
| 2 | 2,14 | 14,34 |
| 3 | 2,18 | 14,60 |
| 4 | 2,13 | 14,54 |
| 5 | 2,12 | 14,87 |
| 6 | 2,04 | 15,30 |
| 7 | 1,85 | 14,28 |
| 8 | 1,28 | 14,80 |
| 9 | 1,06 | 13,64 |
| 10 | 0,75 | 12,94 |
| 11 | 2,12 | 15,69 |

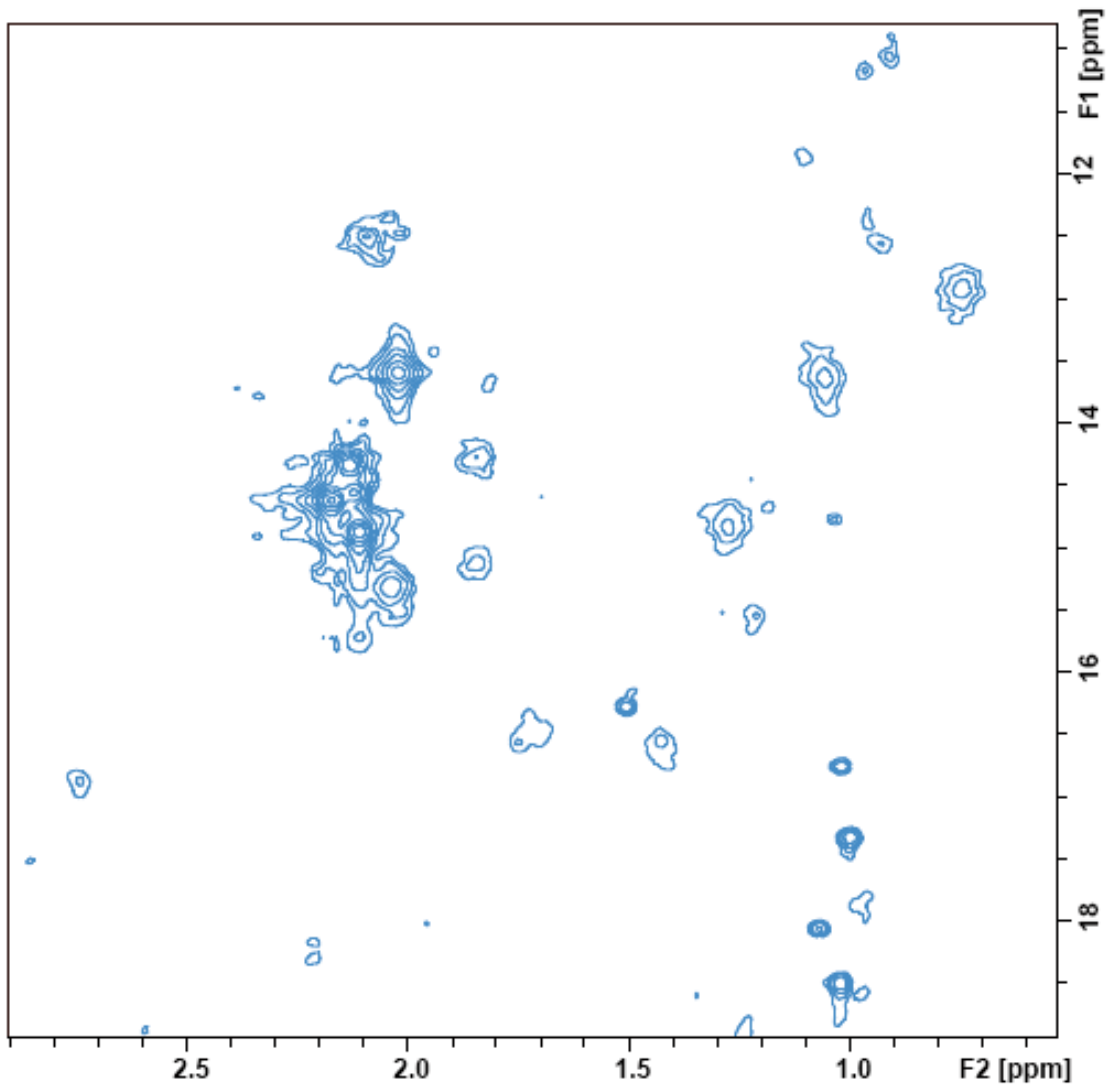
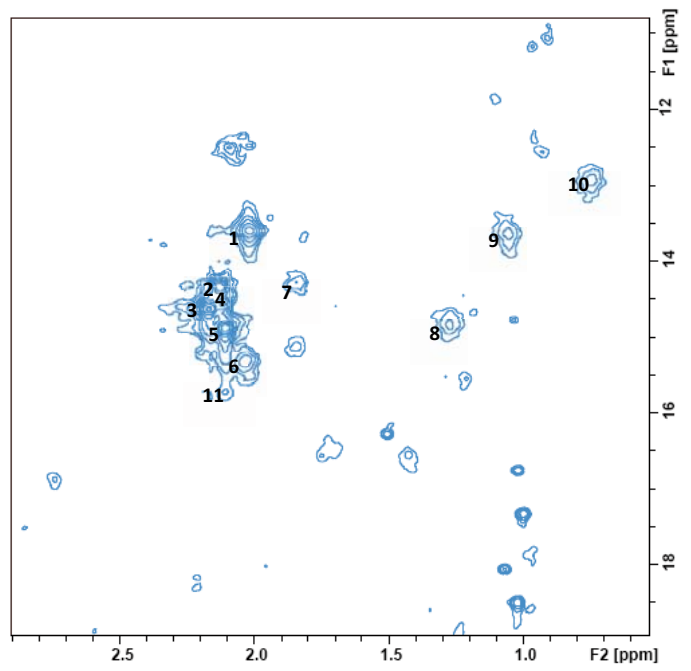


Figure 1.17.: Methyl-TROSY spectra corresponding to methyl- ^{13}C methionine labeled DPP IV (39-766)

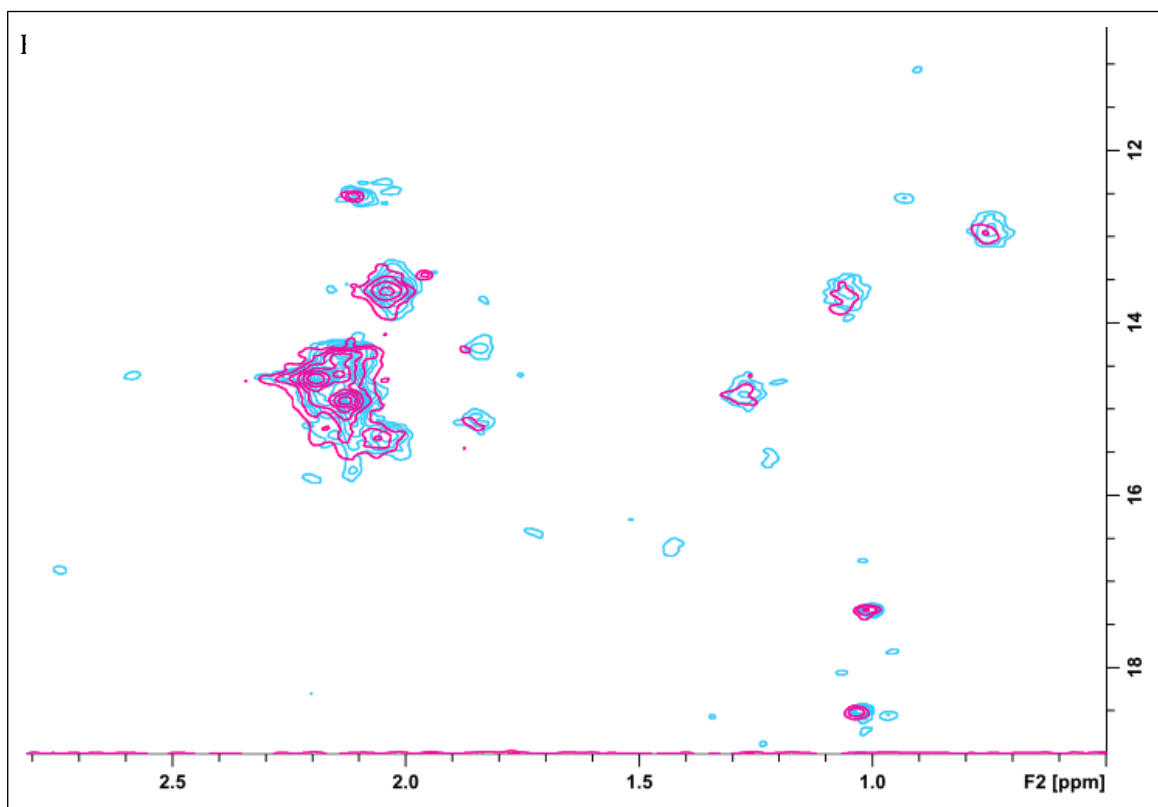


Figure 1.18.: Methyl-TROSY spectra of methyl- ^{13}C methionine labeled DPP IV (39-766) in B600 (pink) and B800 (blue) MHz magnets.

1.3.3. Signal assignment: Site-directed mutagenesis

1.3.3.1. Site-directed mutagenesis

Signal interpretation in large proteins is only possible by means of site-directed mutagenesis. For that reason, the mutagenesis of methionine residues was planned. From the 14 methionine present in the soluble DPP IV protein expressed, five were selected. This selection was based upon the situation of these residues in the tridimensional crystallographic structure of DPP IV (1H2W).

In figure 1.19. selected methionines are depicted in orange. These residues are: Met425, Met503, Met509, Met591 and Met671. All of them are located in the internal cavity of DPP IV between the two domains. Ser630 from the active site is depicted in green. Then, structural changes in the active site, or close-by, will strongly affect the chemical shift of the ^{13}C -Methyl from these methionine residues.

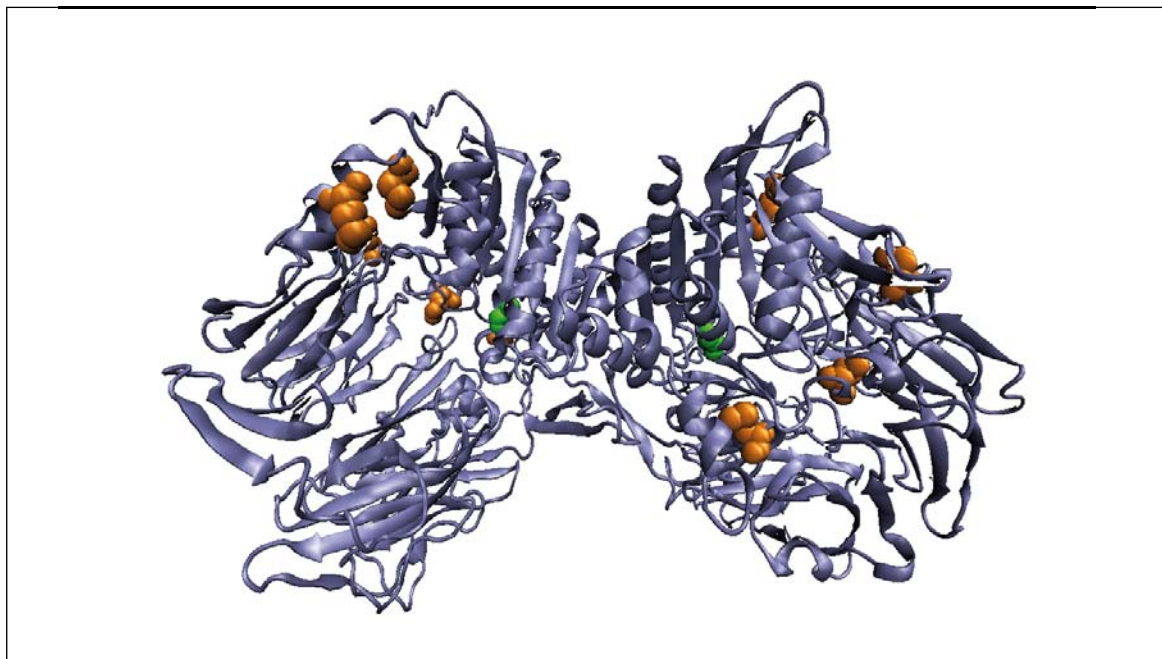


Figure 1.19.: DPP IV crystallographic structure. Methionine residues selected for mutagenesis are highlighted in orange, and Ser630, corresponding to the serine of the active-site, is depicted in green.

Site-directed mutagenesis of selected residues was planned to modify methionine by leucine. Both amino acids are classified as nonpolar (hydrophobic) and have a similar chain length. Furthermore, substitution of methionine with leucine has been proved in POP. In based of this modification, mutated Forward and Reverse primers (mFP and mRP) were designed with the help of the mutagenesis Agilent software.

Mutants were named as M425L, M503L, M509L, M591L and M671L.

Although mutagenesis was expected to be a simple experiment, it turned out into one of the major obstacles during this research.

STRATEGY ONE

First strategy was the use the QuickChange II Mutagenesis kit. It allows the amplification of plasmids with the help of mutated primers that contain the “ctg” codon belonging to Leu instead of “atg” for the forward primer, or “cag” instead of “cat” for the reverse primer. This amplification is performed in one single PCR. After it, parental non-mutated plasmid is digested with the enzyme DpnI. It specifically recognizes methylated DNA, which is produced by *E. coli*, but not by chemical DNA amplification. Therefore, parental plasmid obtained from mini-cultures of *E. coli* would be degraded in front of the new mutated plasmids. A small nick in the DNA chain between the start and the end of the amplification reaction will remain, but after transformation of bacteria with this product, it will be ligated by cells.

Then, pFastBac EGT-C containing wt human DPP IV was submitted to the mutagenesis reaction. This plasmid is already big (around 5000 bp) and DPP IV has incremented its size in 2000 bp more. Although different PCR conditions were assayed (change of concentrations of primers, template, and reagents as well as modification of temperature and time of PCR steps), none of the conditions allowed obtaining bacterial colonies. It was

concluded that because of the large size of the plasmid, the PfuUltra polymerase enzyme could not amplify it completely. A large smearing was detected, a clear signal that the amplification reaction was not giving a single product. A purification step before bacterial transformation was then added in an attempt to enrich the population of the appropriate size plasmid in front of incomplete species. However, it didn't afford colonies.

STRATEGY TWO

At that point, it was decided to change the mutagenesis strategy (figure 1.20.). A first DNA amplification of DPP IV gene was performed in two parts. Part a was the 5' extreme of the mutated gene and was obtained using the wtFP (wild-type Forward Primer) and the mRP. Part b was the 3' extreme of the mutated gene and was obtained by amplification with the mFP and the wtRP. Products of this first PCR reaction were used as templates for the second one. As a result, the full length mutated DPP IV gene was obtained. After cleavage with the restriction enzymes, ligation was performed. Unfortunately, no colonies were obtained. The hypothesis for the failure of this second strategy was that wt primers were not long enough to allow the restriction enzymes to cut. Apart from containing the target site for cleaving, primers must contain additional bp surrounding it since restriction enzymes are endonucleases, but not exonucleases.

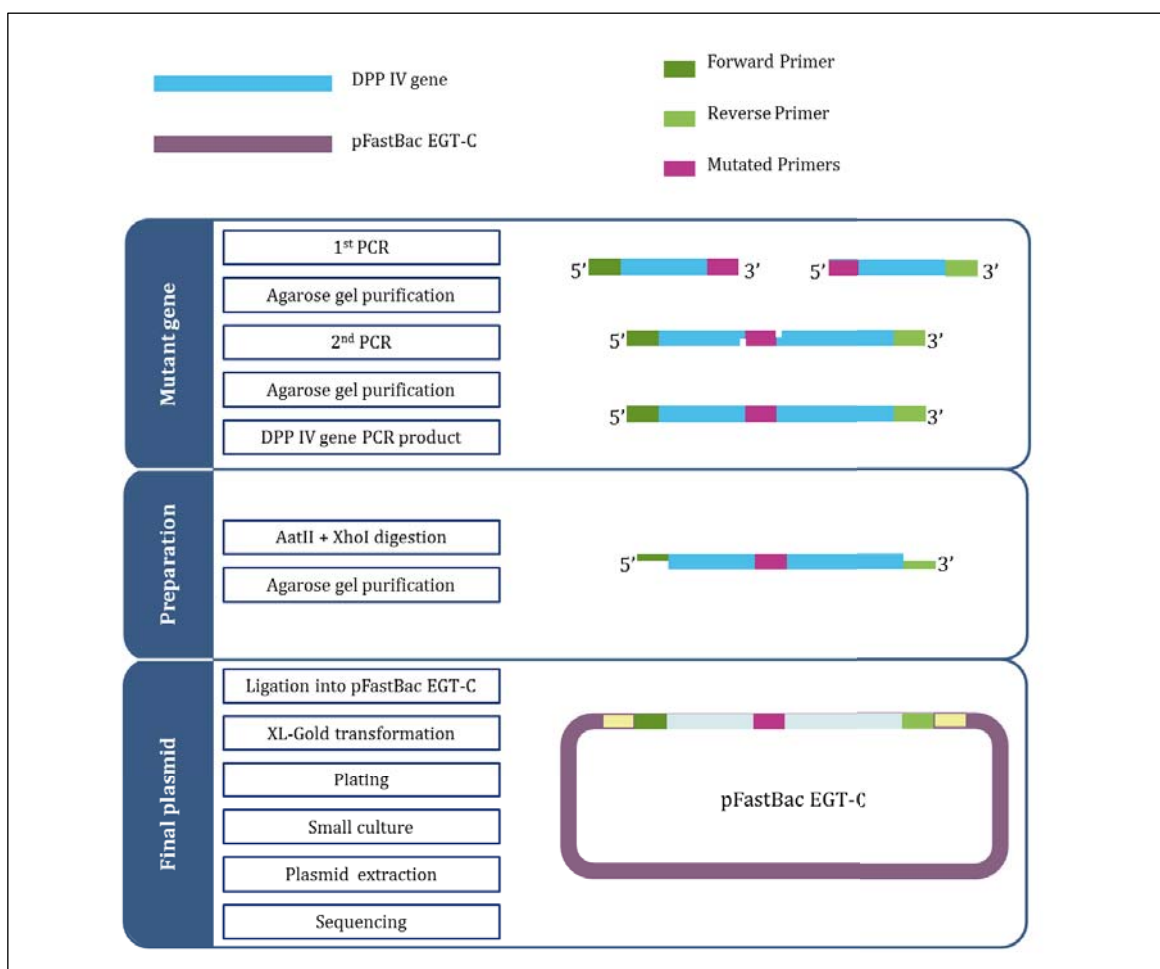


Figure 1.20.: Mutagenesis scheme of the strategy followed for DPP IV mutants.

STRATEGY THREE

As a solution, third strategy consisted in an analog of strategy 2, but with longer versions of primers. Colonies were obtained for all the mutants, but when sequencing, strange results appeared. Using the reverse primer to be able to detect mutation points (all of them are closer to the Cterm of the protein than to the Nterm), amplified sequences were all identified as DPP IV. They also contain the corresponding mutation. However, after the mutated primer, alignment of the sequence with DPP IV gene was not continued. In order to understand why the alignment was not continued, cleavage of mutated pFastBac EGT-C plasmids with the DPP IV surrounding restriction enzymes was done. Purification in agarose gel revealed a smaller mass that coincided with the fragment that aligned with DPP IV. No explanation that could explain this phenomenon was found. And after small trials, this strategy was rejected.

STRATEGY FOUR

Finally, strategy four consisted in going back to QuickChange II Mutagenesis kit but using as a template the pGEM-T Easy containing DPP IV instead of pFastBac EGT-C. The advantage relies in the size of the plasmid, which is 2 kbp smaller. In this case, the DNA amplification reaction worked and colonies were obtained for M425L, M503L, M591L and M671L, but not for M509L. Mutant DPP IV was extracted with the digestion with restriction enzymes, and then, was ligated into the final pFastBac EGT-C. Sequencing of donor plasmids confirmed the constructs M425L, M503L and M671L, but in the case of M591L, the same fragment as in strategy 3 was obtained.

Max Efficiency DH10Bac *E. coli* cells were transformed with mutant DPP IV pFastBac EGT-C plasmids and mutant recombinant bacmids were obtained. Bacmid transfection and baculovirus amplification were performed following the same procedure as for the wild type DPP IV.

SUMMARY OF MUTAGENESIS CONDITIONS

| | |
|------------------------------|--------------------------------|
| Technique | QuickChange II Mutagenesis kit |
| Template | pGEM-T Easy |
| RE digestion | |
| Agarose gel purification | |
| Ligation into pFastBac EGT-C | |

Table 1.4.: DPP IV mutagenesis final strategy

1.3.3.2. Mutant DPP IV selective labeling

The selective labeling of methyl-¹³C methionine, applied before for wild type (wt) protein, was used for the expression of the M503L DPP IV mutant.

The followed protocol for expression and protein purification was the previous described for wt DPP IV.

Then, the Methyl-TROSY experiment was performed at a 600 MHz Bruker Digital avance NMR instrument fitted with triple-resonance z-axis gradient cryoprobe. Data were processed with Topsin 2.0.

Unfortunately, expression yield was significantly low (28 μ M) and spectra resolution was poorer than in the wt DPP IV. Nevertheless, interesting observations were done. First, comparison of the NMR spectra of M503L DPP IV versus the one of the wild type protein was done (figure 1.21.)

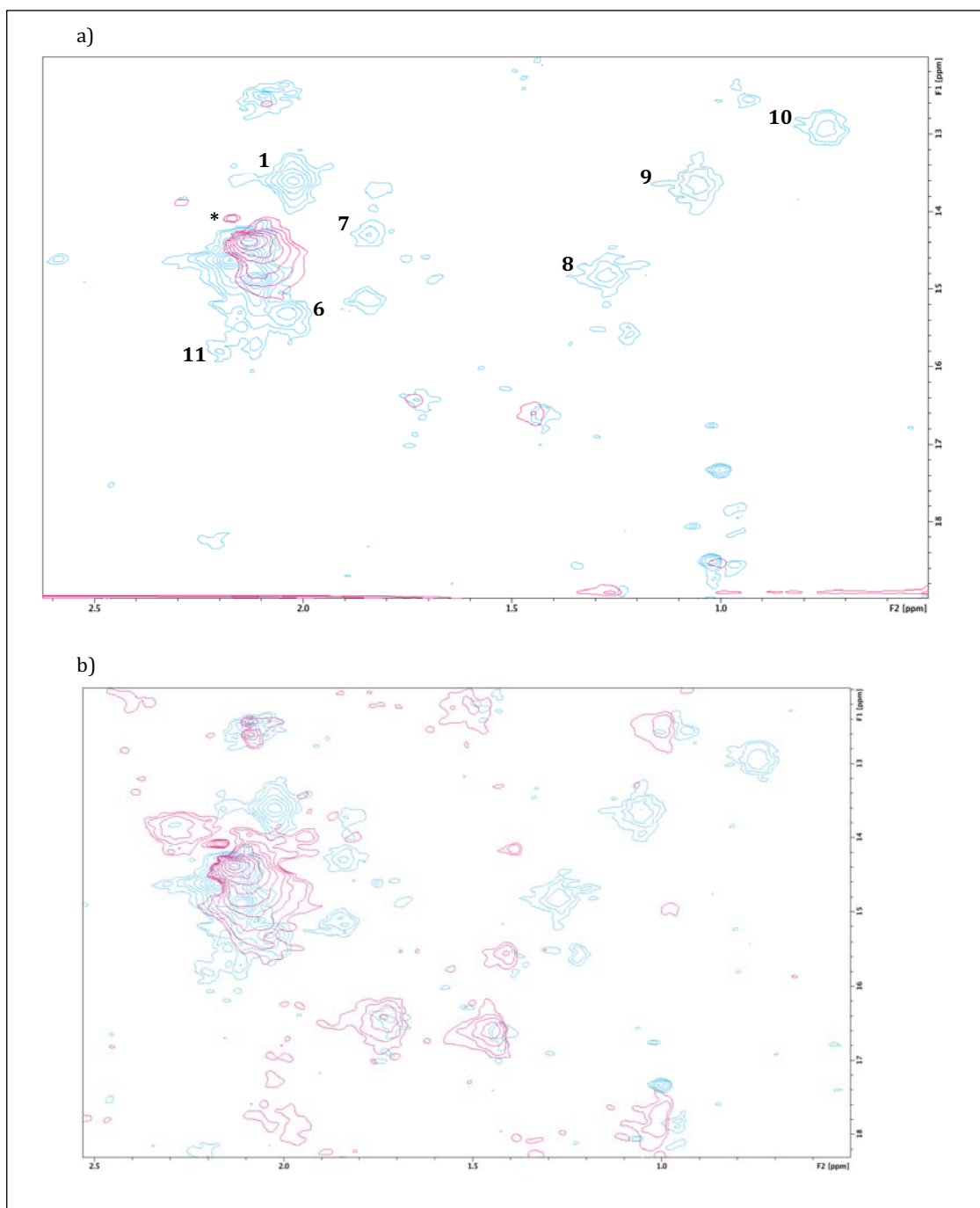


Figure 1.21.: Methyl-TROSY spectra of methyl- 13 C methionine labeled wt DPP IV (blue) and methyl- 13 C methionine labeled M503L DPP IV (pink). In b, the sensitivity of the mutant protein spectra has been increased.

It was observed that signals 1,7, 8, 9, 10 and 11 are no longer detected in the mutant DPP IV NMR spectra. Thus, assignment was not possible, since any of the missing peaks could be the one corresponding to the methionine residue in position 503.

However, an interesting finding was the appearance of a new signal, depicted with an asterisk in figure 1.21. It was hypothesized that mutant M503L presented a slight conformational change respect to the wild type protein, and one of the methionine residues was sensible to this conformational change.

Nevertheless, more experiments should be performed in order to corroborate this finding and identify the residue responsible for the new signal.

1.3.4. NMR study of DPP IV interaction with its inhibitors

As detailed before, POP, a structurally related protein to DPP IV, suffers a displacement in its conformational equilibrium upon binding of ligands. Thus, it was planned to study the CSP (chemical shift perturbation) of DPP IV methionine signals after addition of inhibitors.

Besides, the NMR study of DPP IV with inhibitors planned to address the question if this technique could be used as a tool for identification of competitive versus non-competitive inhibitors. In the case of DPP IV, inhibitors interacting in an allosteric site would affect protein-protein interactions causing undesirable *in vivo* effects. Prediction of non-competitive behaviour of a small molecule in an early stage would greatly help in the drug discovery stage.

Three inhibitors, with different scaffold, were used to study DPP IV interaction with small molecules: Isoleucine-thiazolidide, NVP DPP 728 and Berberine.

1.3.4.1. Interaction of DPP IV with Isoleucine-Thiazolidide (P32/98)

Isoleucine-Thiazolidide (P32/98) is a reported DPP IV inhibitor.^[171]

Calculation of the inhibition constant was performed by an enzymatic assay. First, DPP IV was incubated in presence of a set of inhibitor concentrations at 37°C for 15 min. After, the DPP IV substrate, H-Gly-Pro, was added. Fluorescence values were read at 20 min and percentages of DPP IV inhibition, respect to the positive control, were calculated for each inhibitor concentration. Representation of the % of protein inactivation versus the logarithm of inhibitor concentration afforded a sigmoidal curve. IC₅₀ value is defined as the inflection point. In the case of P32/98, its IC₅₀ value versus DPP IV was set to 517 nM (figure 1.22.).

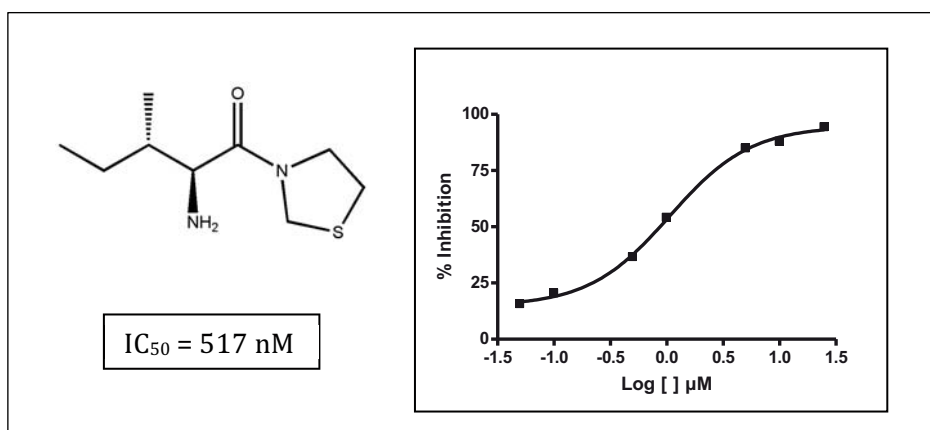


Figure 1.22.: P32/98 structure and IC_{50} curve for DPP IV inhibition.

Regarding its inhibition mechanism, P32/98 is reported as a DPP IV competitive inhibitor.^[172]

Then, the methyl-TROSY experiment of ^{13}C -Methyl methionine labeled DPP IV incubated with P32/98 was performed. The resulting spectra contained all the peaks observed in free DPP IV (figure 1.23. a). The experiment was repeated in a lower field (B600 compared to B800) obtaining the same peaks (figure 1.23. b).

Besides, four additional signals were observed (figure 1.23. a – marked as A-D). For signals A, B and C, the corresponding 1D slices of the peaks were not sharp, and thus, it was suggested a protein origin for them. However, posterior methyl-TROSY experiments of P32/98 alone demonstrated that these signals belonged to the small molecule itself, rather than to the protein (figure 1.24.). The fourth signal (D), whose chemical shift corresponded to 2.16 ppm [H+] and 14.10 ppm [^{13}C], was not present in the spectra of the small molecule, therefore, it was concluded that it was a signature of inhibited DPP IV.

Interestingly, the signal named as “D” is the same that appeared in the M503L DPPIV NMR spectra, named signal “*” (figure 1.25.).

Then, a possible explanation could be that mutant M503L DPP IV adopted a similar conformation to the one of inhibited DPP IV, and the appearance of signal D is the reporter of this minor structural change.

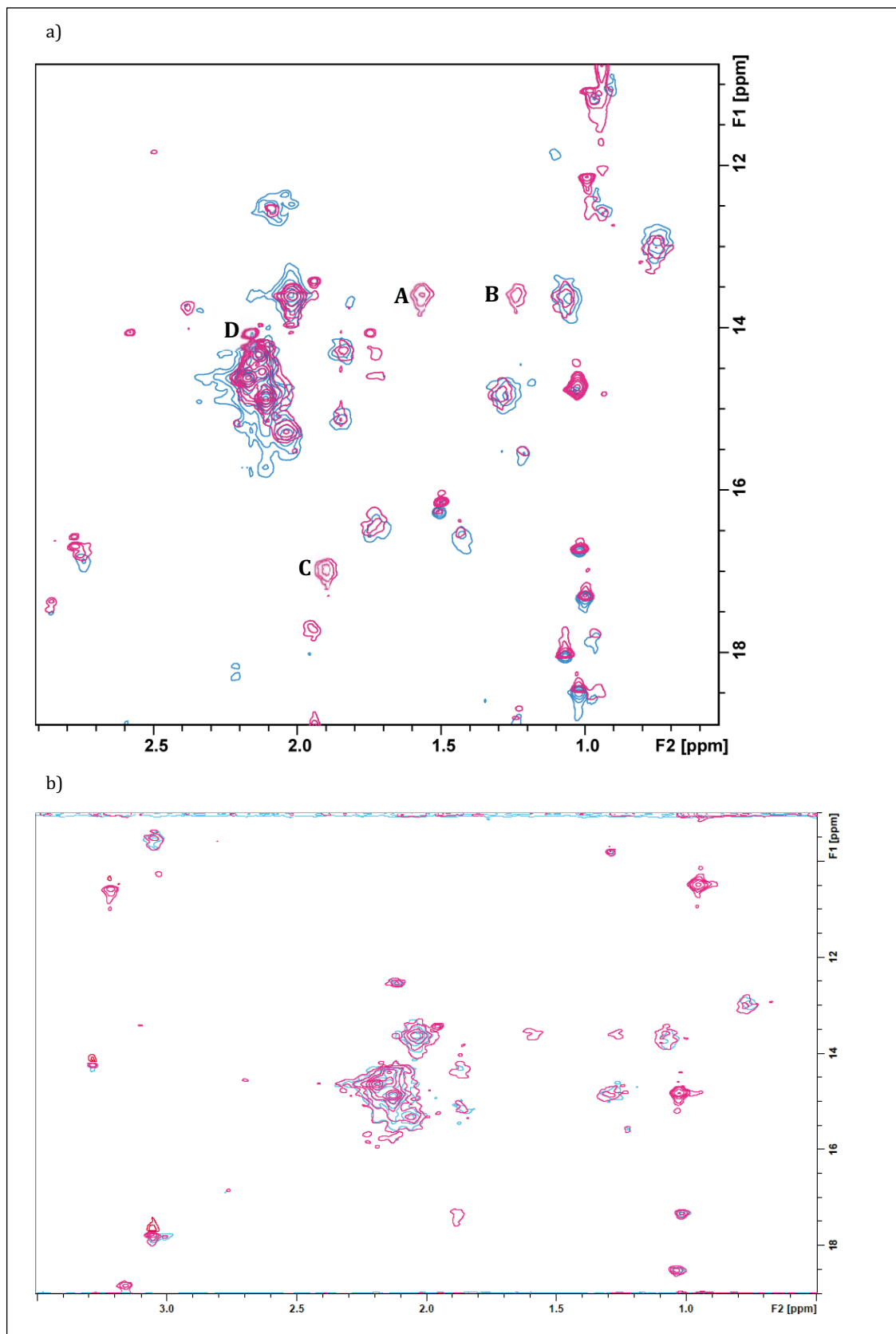


Figure 1.23.: Methyl-TROSY spectra of methyl- ^{13}C methionine labeled DPP IV (39-766) a) Blue: free DPP IV; pink: DPP IV + 10 eq. P32/98; b) DPP IV + P32/98 in B800 (blue) and B600 (pink) MHz equipments.

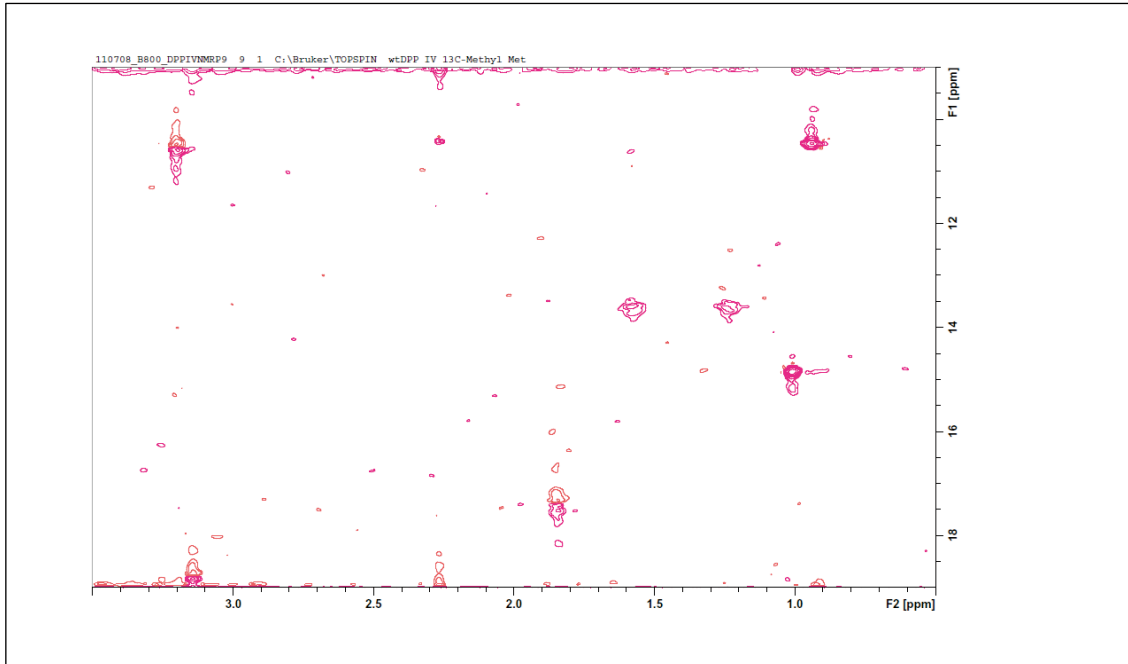


Figure 1.24.: Methyl-TROSY spectra corresponding to P32/98.

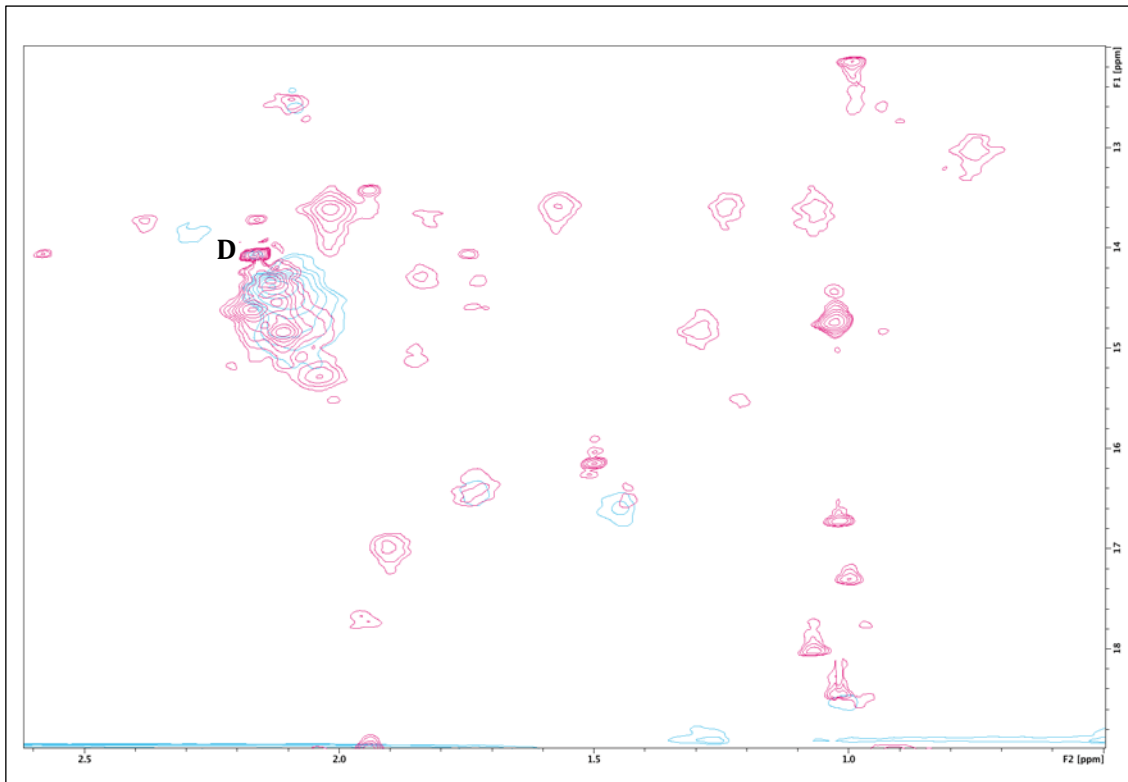


Figure 1.25.: Methyl-TROSY spectra of methyl-¹³C methionine labeled wt DPP IV + 10 eq. P32/98 (pink) and Methyl-TROSY spectra of methyl-¹³C methionine labeled M503L free DPP IV (blue).

Finally, in the same way as for the wt DPP IV, M503L DPP IV was incubated with the canonic inhibitor P32/98 and the Methyl-TROSY experiment was performed (figure 1.26.)

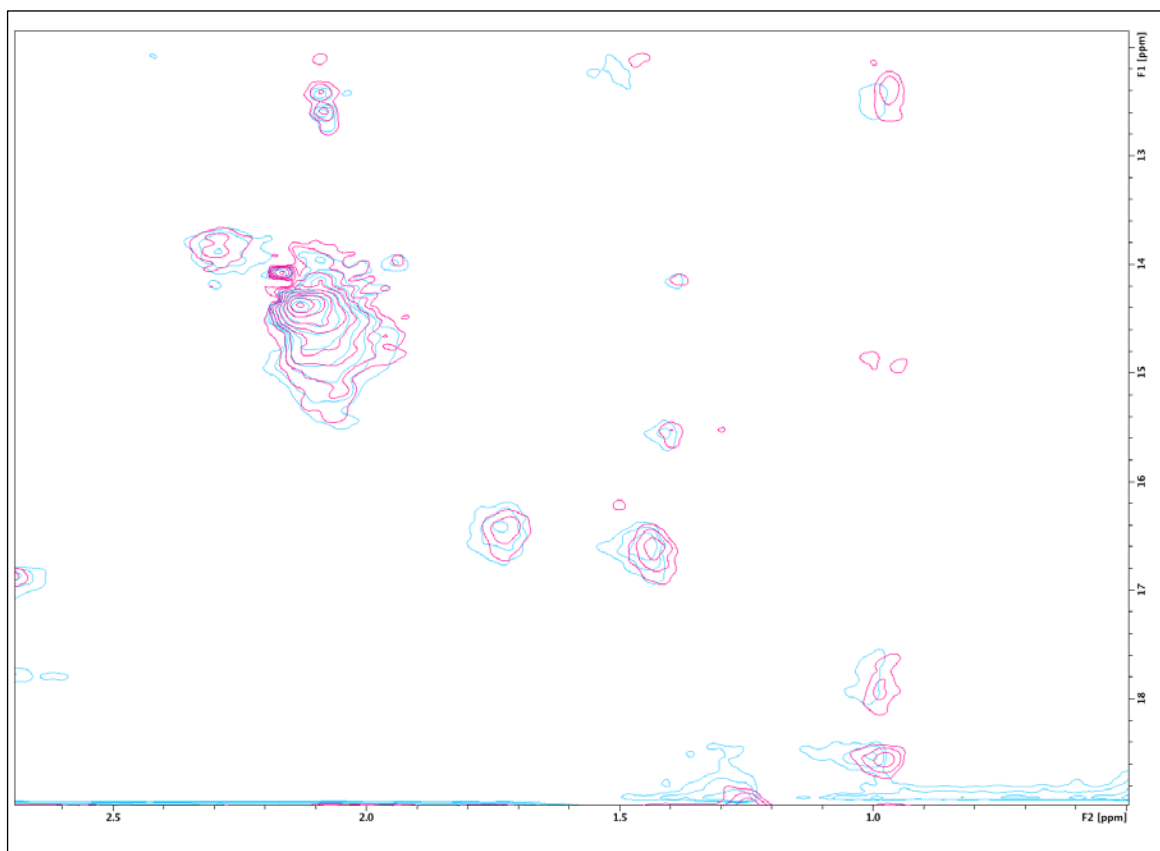


Figure 1.26 Methyl-¹³C methionine labeled M503L DPP IV in blue

In this case, no changes on the spectra belonging to the inhibited protein were detected.

1.3.4.2. Interaction of DPP IV with NVP DPP 728

NVP DPP 728 is a cyanopyrrolidide DPP IV inhibitor discovered by Novartis.^[173] It was characterized as a slow-binding inhibitor that forms a reversible bond with the active-site serine to form an imidate adduct.^[93] Therefore, NVP DPP 728 is a DPP IV competitive inhibitor.

Its potency versus the protease was evaluated, affording an IC_{50} value of 17 nM (figure 1.27.).

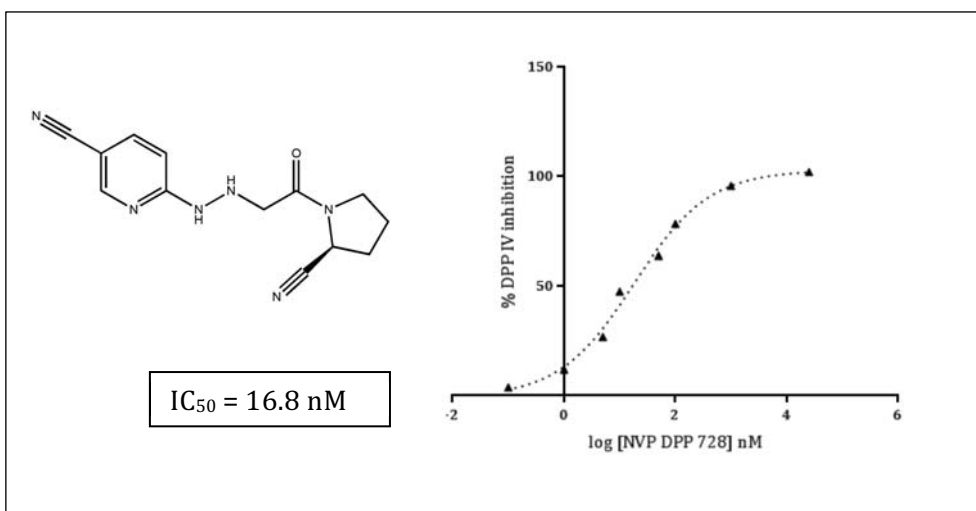


Figure 1.27. NVP DPP 728 structure and IC_{50} curve for DPP IV inhibition.

The effect of NVP DPP 728 on DPP IV, was analyzed by NMR. The methyl-TROSY experiment of ^{13}C -Methyl methionine labeled DPP IV inhibited by this compound was done (figure 1.28.) As a result, the same signal peaks that appeared with DPP IV + P32/98, were observed in this case. The signal denoted as "D" was also present after the action of NVP DPP 728, confirming the hypothesis that the peak was a marker of DPP IV inhibition.

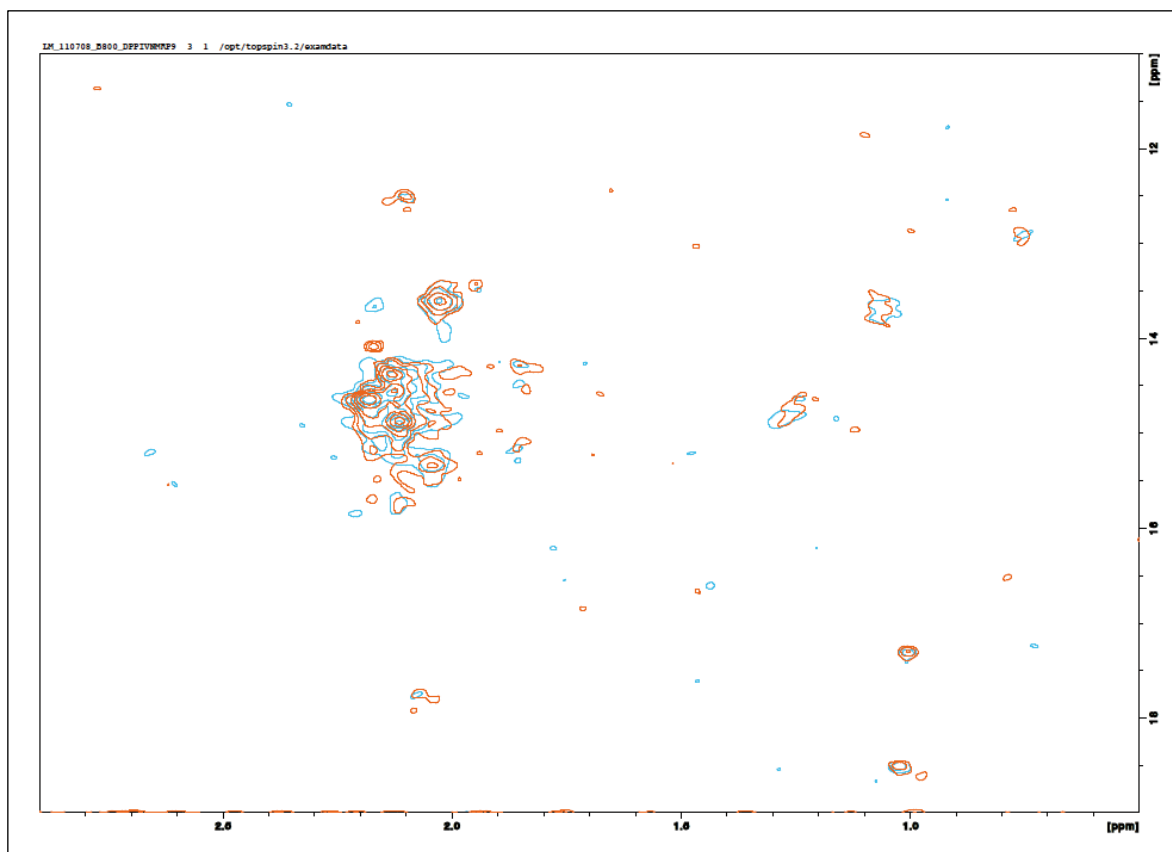


Figure 1.28. Methyl-TROSY spectra of methyl-¹³C methionine labeled DPP IV (39-766) in complex with NVP DPP 728.

1.3.4.3. Interaction of DPP IV with berberine

Berberine is a plant-derived isoquinoline alkaloid described as a natural anti-diabetic compound. Its ability to inhibit DPP IV was described as a possible explanation for its sugar-blood lowering properties.^[174]

Berberine IC₅₀ against DPP IV activity was calculated and set to 140 μM (figure 1.29.), a value 10 times higher than the one reported in the bibliography.^[174]

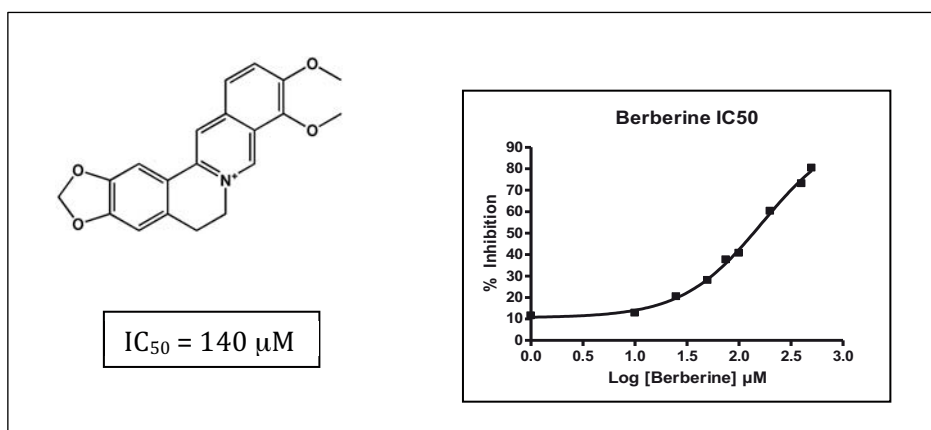


Figure 1.29. Berberine structure and IC₅₀ curve for DPP IV inhibition.

Regarding its mechanism of action, previous crystallization studies described the binding site of berberine.^[174] The molecule is placed in the cavity between the two domains, interacting with Glu205, which is located in a different subsite than the active site serine. In order to fully characterize the mode of inhibition, kinetic experiments were performed. A set of berberine concentrations was assayed against DPP IV activity at different substrate concentration. Initial velocities were calculated for each condition. The representation of velocity inverse versus substrate concentration inverse (Lineweaver-Burk), afforded a series of lines that intersected in the origin (figure 1.30.). Thus, berberine was described as a competitive inhibitor.

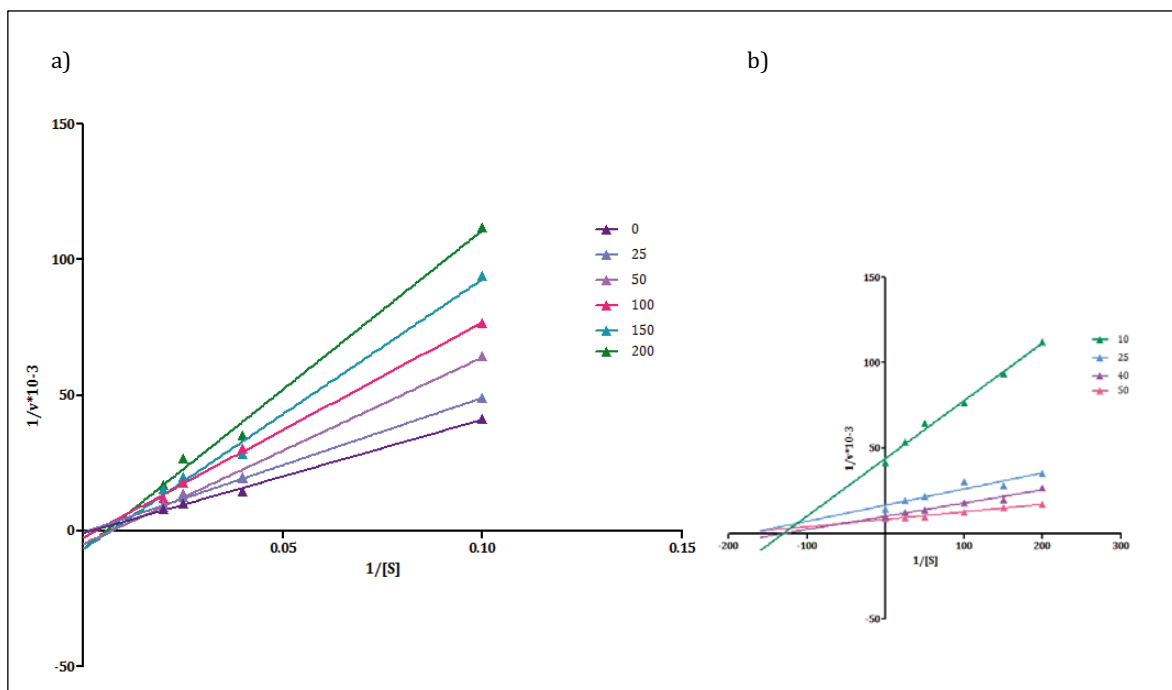


Figure 1.30. a) Lineweaver-Burk representation of berberine. Lines stand for different berberine concentrations; b) Dixon-plot of berberine versus DPP IV. Lines stand for different Substrate concentrations.

Then, the methyl-TROSY experiment of ¹³C-Methyl methionine labeled DPP IV incubated with berberine was performed (figure 1.31.).

In the same way as for the other inhibitors, signals 1 to 11 corresponding to the free DPP IV spectra were observed as well. Signal D, which was characteristic of inhibited DPP IV spectra, also appeared, but in a lesser intensity than with P32/98 and NVP DPP 728.

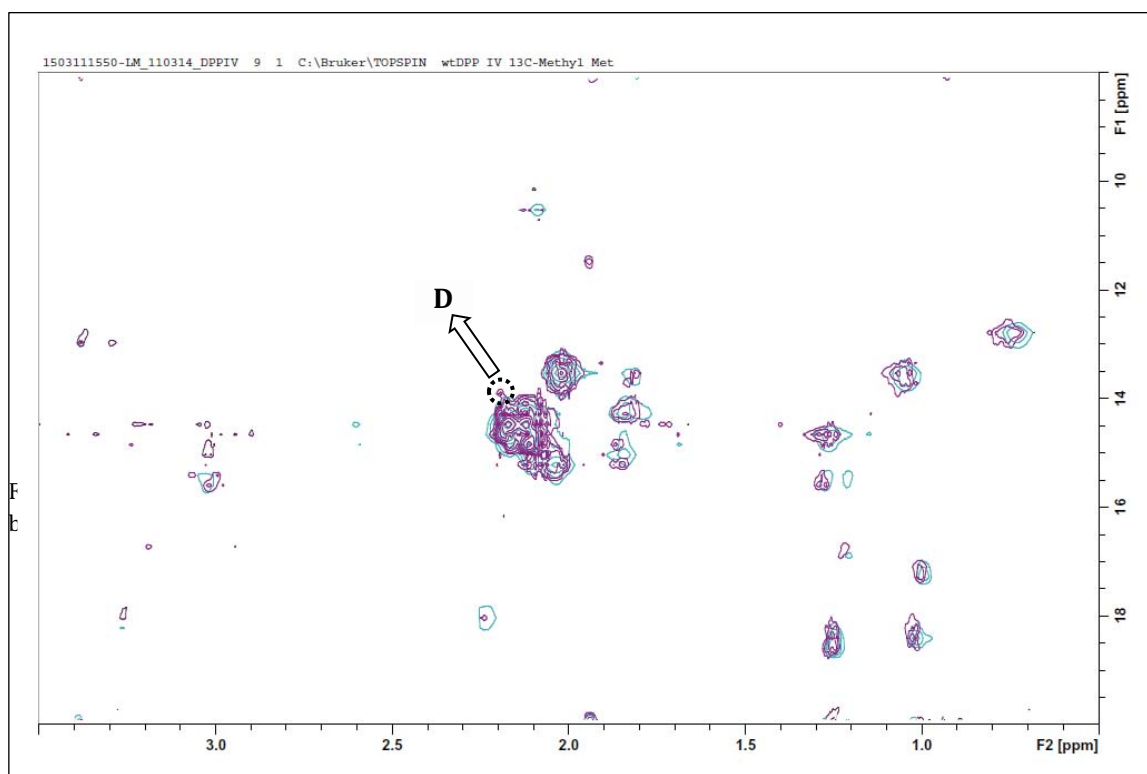


Figure 1.31. Methyl-TROSY spectra of methyl-¹³C methionine labeled DPP IV (39-766) in complex with berberine (purple) versus free DPP IV (blue).

Chapter 1 overview

In this first chapter of results, the recombinant expression of DPP IV was set-up. The use of prokaryotic systems for the expression of the protease was not an option, since DPP IV has post-translational modifications required for a proper structure and activity. Thus, among the eukaryotic systems, insect cells were selected based upon easiness of growth and protein production and the high yields of recombinant proteins. After testing of different protein constructs, insect cell lines and infection and incubation parameters, final conditions were found for an optimum DPP IV expression with moderate yield and high purity.

Once accomplished the DPP IV recombinant expression optimization, the protein was studied by means of NMR in order to obtain information of its dynamism. Since DPP IV is a large protein, a strategy combining selective labeling and the use of TROSY-HSQC experiments was used. From the 14 methionine residues of the protease, 11 of them were detected in the NMR spectra.

Afterwards, assignment of key residues was planned to do. The strategy that was followed was the site-directed mutagenesis, where a methionine residue was substituted by a leucine. Unfortunately, after several attempts, only one mutant was obtained (M503L). Besides, the yield that this mutant afforded was significantly low, and the experiment of NMR did not allowed the assignment of this residue.

Taking advantage of the labeling methodology that was established in the laboratory, study of the inhibitor effect on the NMR spectra of DPP IV was done. Interestingly, the corresponding spectra of DPP IV / inactivator afforded an extra signal.

We believe that it is a consequence of a small structural change that the protease suffers after inhibition. However, further experiments should corroborate this hypothesis.

Chapter 2:
Discovery of DPP IV inhibitors
from plant extracts

CHAPTER 2 CONTEXT**2.1. SELECTION, EXTRACTION AND TEST**

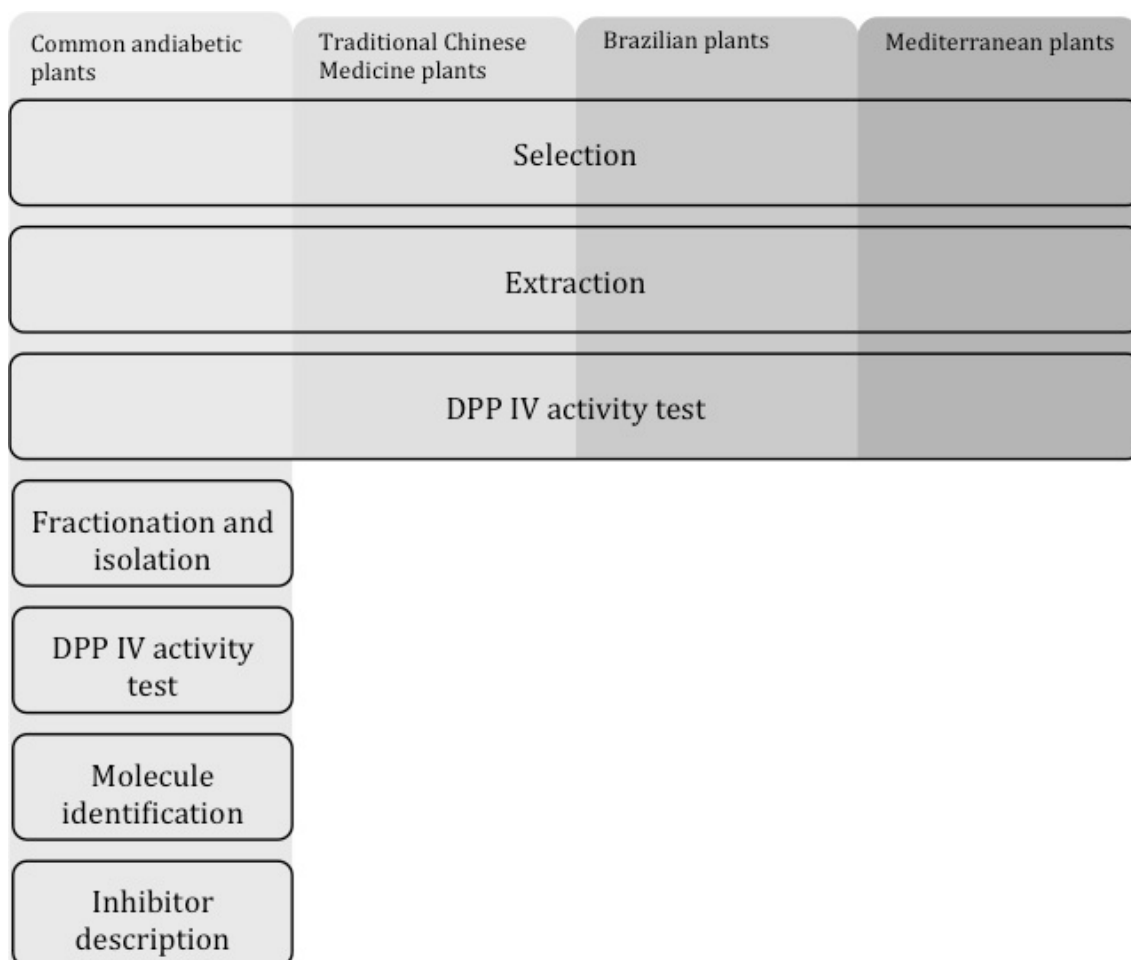
- 2.1.1. Common antidiabetic plants
 - 2.1.1.1. Selection
 - 2.1.1.2. Extraction
 - 2.1.1.3. Plant extracts screening
- 2.1.2. Traditional Chinese Medicinal plants
 - 2.1.2.1. Selection
 - 2.1.2.2. Extraction
 - 2.1.2.3. Plant extracts screening
- 2.1.3. Brazilian plants
 - 2.1.3.1. Selection
 - 2.1.3.2. Extraction
 - 2.1.3.3. Plant extracts screening
- 2.1.4. Mediterranean plants
 - 2.1.4.1. Selection
 - 2.1.4.2. Extraction
 - 2.1.4.3. Plant extracts screening

2.2. *COUTAREA LATIFLORA*

- 2.2.1. Extract fractionation and compound isolation
- 2.2.2. Molecule characterization
- 2.2.3. Inhibitor characterization
 - 2.2.3.1. Study of the inhibition
 - 2.2.3.2. Inhibition mechanism
 - 2.2.3.3. NMR analysis

CHAPTER 2 OVERVIEW

Experimental scheme



Chapter 2 context

First human civilizations found in Nature the cure for a variety of disorders. For that reason, academic scientists have used Nature as a source for drugs and still continue to use it nowadays, together with pharmaceutical companies.

One of the motivations to go back to Nature is the large chemical space it represents. The vast landscape of new scaffolds is a very attractive scenario to look for new leads, which will represent innovative and effective molecules in a diversity of diseases. This distinctive area of chemical space is possible due to the number of differences between biosynthesis and the synthesis performed in medicinal chemistry. [175] In biosynthesis the building blocks are limited, but a high diversity is accomplished by a set of diverse pathways. On the other hand, synthesis takes advantage of the large variety of building blocks that are transformed then by a certain number of reactions. Another principal discrepancy between the two syntheses is that biosynthesis is oxophilic and achieves stereoselectivity by selective enzymes. Meanwhile, medicinal chemists produce preferably low stereochemical molecules with a high proportion of nitrogen atoms. [175] The case of plants is a clear representative of these features. In the process of adaptation to the environment, plants have evolved, synthesizing specific molecules that ensure their survival. By this, all botanical species contain a remarkably large source of chemicals, which may be of medicinal interest.

Apart from diversity in terms of molecule content, Nature is attractive in the field of drug discovery for other reasons. Natural products have been used since centuries ago by different cultures all over the world in order to cure or ameliorate certain diseases and maladies. This implies that these extracts contain effective molecules that reach a certain target and interact with it, with a positive result for the patient. Not only these molecules are effective, but are also safe, in terms of side effects. Natural extracts have been tested by a large population and, with certain exemptions, have no serious side effects. Furthermore, all the traditional treatments are based in oral administration, fact that implies that the responsible molecule or set of molecules can potentially be administered orally. Then, effective, safe and orally bioavailable molecules are present in natural products, and for that reason, they represent a powerful source for the discovery of new drugs.

According to Newmann and Cragg [176], from the almost one thousand small molecules approved drugs from 1981 to 2010, a total of 64 have natural origin. The number is increased when taking into account synthetic molecules with a natural product as a pharmacophore (55 molecules belong to this group, making a total of 119 new chemical entities). Altogether, it accounts for the 11% of the approved drugs during this period of time, a number that reflects the high value of natural sources in drug discovery.

In the field of diabetes, 37 new chemical entities and medical indications have been approved from 1981 to 2010. [176] From these drugs, 12 have a total synthetic origin. Regarding to the serine protease DPP IV, the four inhibitors discovered in this period (alogliptin benzoate, sitagliptin, saxagliptin and vildagliptin) are enclosed in the synthetic group. Linagliptin, which was discovered one year after, in 2011, belongs to the same group as well. The fact that none of the DPP IV inhibitors has a natural origin does not imply that it is not a feasible source, but that this area has not been examined yet. In this

regard, new scaffolds may be found when tested against this protease, increasing the explored chemical space and contributing to a large panorama of DPP IV inhibitors.

Among natural products, plants represent a powerful source in the diabetes treatment. Until 2008, 800 medicinal plants were used in the folk treatment of diabetes, but only 410 had been proved as antidiabetic agents. However, the mechanism of action was proven for 109 extracts. ^[177,178] Regarding the compounds with botanical origin, 85 molecules were found between 2005 and 2010 to have sugar-lowering action. ^[178] None of them has been described as a DPP IV inhibitor, though the mechanism was not examined for all the molecules.

Based on the lack of DPP IV inhibitors with botanical origin and given the potential of medicinal plant extracts as providers of antidiabetic agents, a study for the finding of new DPP IV inhibitors based on this source was planned. Traditional Chinese medicinal plants, Brazilian plants and a group of other antidiabetic plants were selected for this purpose. At the same time, a library of Mediterranean plants that was accessible, was also tested, despite this group was not filtered for sugar-blood disorders.

2.1. Selection, extraction and test

2.1.1. Common antidiabetic plants

In the course of the present study, plants that, according to the popular belief have antidiabetic properties, were tested in order to find new DPP IV inhibitors. The variety of plants that were assayed is described in this section.

2.1.1.1. Selection

Common plants whose extracts were reported in the bibliography as sugar-blood lowering agents were chosen. A total of seven were selected. Three of them (AP-2, AP-3 and AP-6) were acquired in herbalist stores, other three (AP-4, AP-5 and AP-7) were bought at local stores and the last one (AP-1) was collected in a personal garden.

| Reference | Latin | English |
|------------------|----------------------------|----------------|
| AP-1 | <i>Aloe vera</i> | Aloe |
| AP-2 | <i>Vaccinium myrtillus</i> | Blueberry |
| AP-3 | <i>Coutarea latiflora</i> | Copalchi |
| AP-4 | <i>Allium sativum</i> | Garlic |
| AP-5 | <i>Lycium barbarum</i> | Goji |
| AP-6 | <i>Urtica dioica</i> | Ortiga |
| AP-7 | <i>Raphanus sativus</i> | Radish |

Table 2.1.: Selected antidiabetic plants.

2.1.1.2. Extraction

For each plant, aqueous extractions and ethanolic macerations were done.

Aqueous extractions were performed using the Soxhlet system. After 5-6 solvent cycles, the extract was collected and frozen-dried.

Ethanolic macerations were done by incubation of dried grinded plant or cut up fresh plant with pure ethanol in a closed glass recipient. After 30 days at room temperature, the infusions were solvent evaporated. The extract was recovered from the balloon by adding a few milliliters of 50:50 H₂O/Acetonitrile. The solution was later transferred to tubes and finally it was frozen-dried.

2.1.1.3. Plant extracts screening

A stock of 5 mg/mL was done for each extract. Aqueous extracts were solubilized in Milli-Q H₂O, and ethanolic extracts in DMSO. A final concentration of 100 µg/mL was assayed in order to calculate the inhibition potency versus DPP IV. (Figure 2.1.)

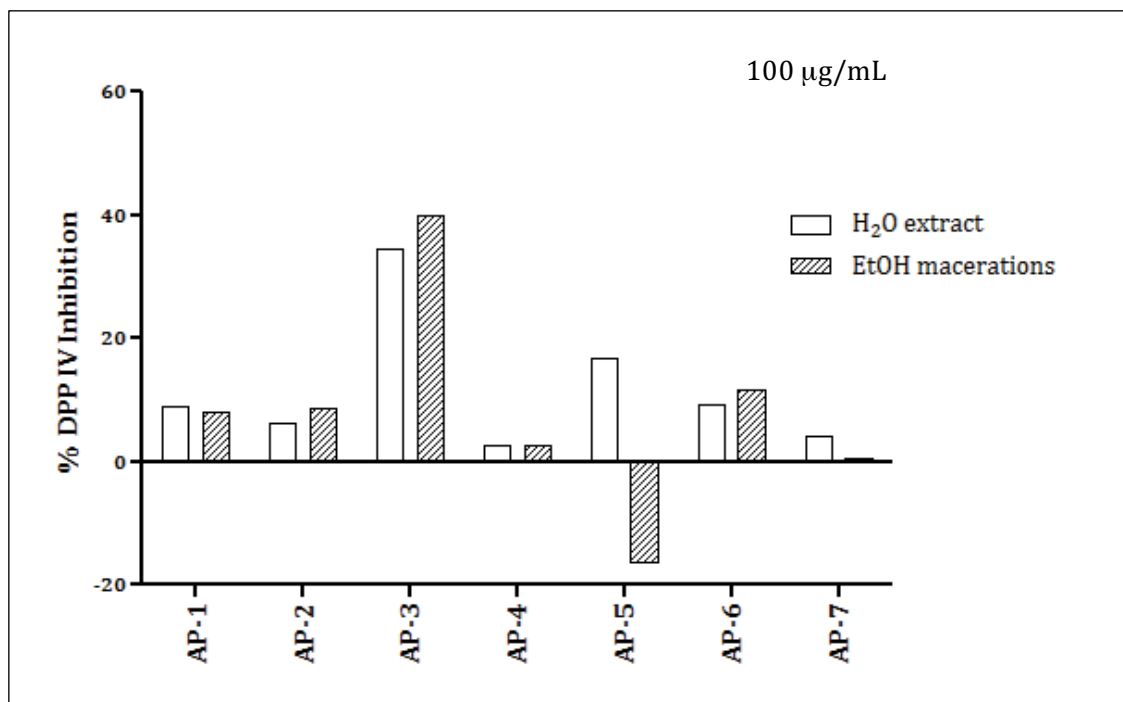


Figure 2.1.: DPP IV inhibition potency for the antidiabetic plant extracts.

Copalchi (AP-3, *Coutarea latiflora*) demonstrated to be a powerful source for DPP IV inhibitors. A posterior assay at a concentration of 500 µg/mL, showed a protease inhibition of 76% for the aqueous extraction and 79% for the ethanolic maceration.

2.1.2. Traditional Chinese Medicinal plants

Traditional Chinese Medicine (TCM) is mainly based (but not solely based) in a large variety of plants and the rich knowledge that exists of their extracts properties. All over the country, around 11,000 plant species are applied in the treatment of diseases. [179] However, some treatments are exclusive of some doctor's family, and the knowledge is only transmitted from generation to generation. This isolation avoids the share of information among the population. For this reason, and others, only 10-15% of the used plants are reported in the medical literature. A deep scientific study of plant extracts would help to increase this percentage.

In TCM, Diabetes is named as *Xiao Ke*, *Xiaokezheng* or *Xiaodanzheng*, and is characterized by a Yin (passive feminine part of the universe) deficiency and dry heat, which ultimately affects Qi (vital energy). [179] Then, TCM treatment for diabetes is based on nourishing Yin and invigorating Qi. [180]

In the diabetes field, Chinese medicines are potent in the treatment of the associated conditions but also in reducing the blood sugar concentration, as demonstrated by the efficacy in single usage. [179] Around 200 species of plants show hypoglycemic effect, [181] which represents an extremely large ensemble of potential active molecules without side-effects.

With this information, a study for the finding of DPP IV inhibitors in Traditional Chinese Medicine was planned.

2.1.2.1. Selection

Picking process was based in the known bibliography ^[178, 180, 181] and 55 plants were selected (table 2.2.). In order to acquire the dried plants, the company Dasherb (formerly Herbasinica) was contacted.

From the 55 plants, 12 were not available. For the resting 43 plants, 11 were already in the lab for a previous project (Plants with code TCM-TT-xx). The remaining 32 plants were acquired and named TCM-LM-01 to TCM-LM-32. (Table 2.3.).

| # | Pinyin | Latin | English |
|----|------------------|--|---------------------------------------|
| 1 | Bai Shao Yao | <i>Radix Paeoniae Lactiflorae</i> | White Peony Root |
| 2 | Bai Zhu | <i>Atractylodes macrocephala</i> | Largehead Atractylodes Rhizome |
| 3 | Ban Xia | <i>Pinellia ternata</i> | Pinellia Tuber |
| 4 | Bei Sha Shen | <i>Glehnia littoralis</i> | Beach silvertop |
| 5 | Calcined Long Gu | <i>Os draconis</i> | Fossils of mammal vertebrae+extremity |
| 6 | Calcined Mu Li | <i>Concha Ostreae</i> | Oyster Shell |
| 7 | Chai Hu | <i>Bupleurum chinense</i> | Chinese Thorowax Root |
| 8 | Chuan Xiong | <i>Radix Ligustici Chuanzixiong</i> | Sichuan Lovage Root |
| 9 | Dan Shen | <i>Radix Salviae Miltiorrhizae</i> | Salvia root |
| 10 | Dan Zhu Ye | <i>Herba Lophatheri Gracilis</i> | Bamboo leaves |
| 11 | Dang Gui | <i>Radix Angelicae Sinensis</i> | Chinese Angelica Root |
| 12 | Du Zhong | <i>Cortex Eucommiae Ulmoidis</i> | Eucommia bark |
| 13 | Fu Ling | <i>Sclerotium Poriae Cocos</i> | White Poria |
| 14 | Fu Pen Zi | <i>Fructus Rubi Chingii</i> | Chinese Raspberry |
| 15 | Fu Shen | <i>Sclerotium Poriae Cocos Pararadicis</i> | Poria with wood root |
| 16 | Gan Cao | <i>Radix Glycyrrhizae Uralensis</i> | Raw Licorice Root |
| 17 | Ge Gen | <i>Radix Puerariae</i> | Lobed Kudzuvine Root (pueraria root) |
| 18 | Geng Mi | <i>Semen Oryzae</i> | Husked uncooked non-glutinous rice |
| 19 | Gou Qi Zi | <i>Fructus Lycii</i> | Chinese Wolfberry Fruit |
| 20 | Gui Zhi | <i>Ramulus Cinnamomi Cassiae</i> | Cinnamon Twig |
| 21 | He Show Wu | <i>Polygonum multiflorum</i> | Fleeceflower Root |
| 22 | Hong Hua | <i>Flos Carthami Tinctorii</i> | Safflower Flower |
| 23 | Hu Zhang | <i>Polygonum cuspidatum</i> | Giant Knotweed Rhizome |
| 24 | Huai Niu Xi | <i>Achyranthes bidentata</i> | Twotooth Achyranthes Root |
| 25 | Huang Bai | <i>Cortex Phellodendri</i> | Amur Cork-Tree Bark |
| 26 | Huang Lian | <i>Rhizoma Coptidis chinensis</i> | Coptis Rhizome |
| 27 | Huang Qi | <i>Radix Astragalus Membranaceus</i> | Milk-Vetch root |
| 28 | Jin Yin Hua | <i>Lonicera japonica</i> | Honeysuckle Flower |
| 29 | Ju Hua | <i>Flos Chrysanthemi Morifolii</i> | Chrysanthemum Flower |
| 30 | Lian Qiao | <i>Forsythia suspensa</i> | Weeping Forsythia Capsule |
| 31 | Lu Gen | <i>Phragmites communis</i> | Reed Rhizome |
| 32 | Mai Men Dong | <i>Tuber Ophiopogonis Japonici</i> | Tuber of Dwarf Lilyturf |
| 33 | Mu Dan Pi | <i>Cortex Moutan Radicis</i> | Cortex of tree peony |
| 34 | Nan Gua Zi | <i>Cucurbita moschata</i> | Cushaw seed |
| 35 | Pao Fu Zi | <i>Aconitum</i> | Ginger processed aconite |
| 36 | Pi Pa Ye | <i>Eriobotrya japonica</i> | Loquat Leaf |
| 37 | Ren Shen | <i>Radix Ginseng - Panax Ginseng</i> | Ginseng |
| 38 | Rou Gui | <i>Cinnamomum cassia</i> | Cassia Bark |
| 39 | Sang Ye | <i>Morus alba</i> | Mulberry Leaf |
| 40 | Shan Yao | <i>Radix Dioscoreae Oppositae</i> | Chinese Yam Rhizome |
| 41 | Shan Zhu Yu | <i>Fructus corni Officinalis</i> | Asiatic cornelian cherry fruit |
| 42 | Sheng Di Huang | <i>Radix Rehmanniae glutinosae</i> | Fresh Rehmannia |
| 43 | Shi Gao | <i>Gypsum</i> | Gypsum |
| 44 | Shu di Huang | <i>Radix Rehmanniae glutinosae conquitae</i> | Cooked Rehmannia root |
| 45 | Tai Zi Shen | <i>Radix Pseudostellariae Heterophyllae</i> | Caryophyllaceous root |
| 46 | Tao Ren | <i>Semen Persicae</i> | Peach Kernel |
| 47 | Tian Hua Fen | <i>Radix Trichosanthis Kirilowii</i> | Trichosantes root |
| 48 | Tian Men Dong | <i>Asparagus cochinchinensis</i> | Cochinchinese Asparagus Root |
| 49 | Tu Si Zi | <i>Semen Cuscutae Chinensis</i> | Chinese Dodder seeds |
| 50 | Wu Wei Zi | <i>Fructus Schisandrae Chinensis</i> | Schizandra fruit |
| 51 | Xuan Shen | <i>Radix Scrophulariae Ningpoensis</i> | Ningpo Figwort Root |
| 52 | Yi Mu Cao | <i>Herba Leonuri Heterophylli</i> | Chinese motherwort |
| 53 | Yu Zhu | <i>Rhizoma Polygonati Odorati</i> | Solomon's Seal Rhizome |
| 54 | Ze Xie | <i>Alisma Orientale</i> | Oriental Waterplantain Rhizome |
| 55 | Zhi Mu | <i>Rhizoma Anemarrhenae Asphodeloidis</i> | Anemarrhena rhizome |

Table 2.2.: 55 Chinese medicinal plants selected.

| Reference | Pinyin | Latin | English |
|-----------|--------------|---|---------------------------------------|
| TCM-LM-01 | Bai Shao Yao | <i>Radix Paeoniae Lactiflorae</i> | White Peony Root |
| TCM-LM-02 | Bai Zhu | <i>Atractylodes macrocephala</i> | Largehead Atractylodes Rhizome |
| TCM-LM-03 | Ban Xia | <i>Pinellia ternata</i> | Pinellia Tuber |
| TCM-LM-04 | Chai Hu | <i>Bupleurum chinense</i> | Chinese Thorowax Root |
| TCM-TT-11 | Chuan Xiong | <i>Radix Ligustici Chuanxiong</i> | Sichuan Lovage Root |
| TCM-LM-05 | Dan Shen | <i>Radix Salviae Miltiorrhizae</i> | Salvia root |
| TCM-LM-06 | Dan Zhu Ye | <i>Herba Lophatheri Gracilis</i> | Bamboo leaves |
| TCM-TT-30 | Dang Gui | <i>Radix Angelicae Sinensis</i> | Chinese Angelica Root |
| TCM-LM-07 | Dao Ya | <i>Fructus Oryzae Germinatus</i> | Rice-grain Sprout |
| TCM-TT-24 | Di Huang | <i>Radix Rehmanniae glutinosae</i> | Fresh Rehmannia |
| TCM-LM-08 | Du Zhong | <i>Cortex Eucommiae Ulmoidis</i> | Eucommia bark |
| TCM-TT-18 | Fu Ling | <i>Sclerotium Poriae Cocos</i> | White Poria |
| TCM-LM-09 | Fu Zi | <i>Aconitum</i> | Ginger processed aconite |
| TCM-LM-10 | Fu Pen Zi | <i>Fructus Rubi Chingii</i> | Chinese Raspberry |
| TCM-LM-11 | Ge Gen | <i>Radix Puerariae</i> | Lobed Kudzuvine Root (pueraria root) |
| TCM-TT-44 | Gou Qi Zi | <i>Fructus Lycii</i> | Chinese Wolfberry Fruit |
| TCM-LM-12 | Gui Zhi | <i>Ramulus Cinnamomi Cassiae</i> | Cinnamon Twig |
| TCM-LM-13 | He Show Wu | <i>Polygonum multiflorum</i> | Fleeceflower Root |
| TCM-LM-14 | Huai Niu Xi | <i>Achyranthes bidentata</i> | Twotooth Achyranthes Root |
| TCM-LM-15 | Huang Bai | <i>Cortex Phellodendri</i> | Amur Cork-Tree Bark |
| TCM-TT-16 | Huang Lian | <i>Rhizoma Coptidis chinensis</i> | Coptis Rhizome |
| TCM-LM-16 | Huang Qi | <i>Radix Astragalus Membranaceus</i> | Milk-Vetch root |
| TCM-LM-17 | Jin Yin Hua | <i>Lonicera japonica</i> | Honeysuckle Flower |
| TCM-TT-39 | Long Gu | <i>Os draconis</i> | Fossils of mammal vertebrae+extremity |
| TCM-LM-18 | Lu Gen | <i>Phragmites communis</i> | Reed Rhizome |
| TCM-LM-19 | Mu Dan Pi | <i>Cortex Moutan Radicis</i> | Cortex of tree peony |
| TCM-LM-20 | Mu Li | <i>Concha Ostreae</i> | Oyster Shell |
| TCM-LM-21 | Pi Pa Ye | <i>Eriobotrya japonica</i> | Loquat Leaf |
| TCM-TT-12 | Ren Shen | <i>Radix Ginseng - Panax Ginseng</i> | Ginseng |
| TCM-LM-22 | Rou Gui | <i>Cinnamomum cassia</i> | Cassia Bark |
| TCM-TT-05 | Sang Ye | <i>Morus alba</i> | Mulberry Leaf |
| TCM-LM-23 | Shan Yao | <i>Radix Dioscoreae Oppositae</i> | Chinese Yam Rhizome |
| TCM-LM-24 | Shan Zhu Yu | <i>Fructus corni Officinalis</i> | Asiatic cornelian cherry fruit |
| TCM-LM-25 | Shi Gao | <i>Gypsum</i> | Gypsum |
| TCM-LM-26 | Tai Zi Shen | <i>Radix Pseudostellariae Heterophyllae</i> | Caryophyllaceous root |
| TCM-LM-27 | Tao Ren | <i>Semen Persicae</i> | Peach Kernel |
| TCM-LM-28 | Tian Hua Fen | <i>Radix Trichosanthis Kirilowii</i> | Trichosantes root |
| TCM-LM-29 | Tu Si Zi | <i>Semen Cuscutae Chinensis</i> | Chinese Dodder seeds |
| TCM-TT-37 | Xuan Shen | <i>Radix Scrophulariae Ningpoensis</i> | Ningpo Figwort Root |
| TCM-LM-30 | Yi Mu Cao | <i>Herba Leonuri Heterophylli</i> | Chinese motherwort |
| TCM-TT-15 | Yu Zhu | <i>Rhizoma Polygonati Odorati</i> | Solomon's Seal Rhizome |
| TCM-LM-31 | Ze Xie | <i>Alisma Orientale</i> | Oriental Waterplantain Rhizome |
| TCM-LM-32 | Zhi Mu | <i>Rhizoma Anemarrhenae Asphodeloidis</i> | Anemarrhena rhizome |

Table 2.3.: 43 Chinese medicinal plants used for the search of DPP IV inhibitors.

2.1.2.2. Extraction

For each dried plant, aqueous extractions and ethanolic macerations were done. Aqueous extractions were performed using the Soxhlet system. The dried grinded plant was placed in a cellulose cartridge, separated from the solvent, which was in the balloon. By heating, the solvent, water in this case, was evaporated and when it entered in contact with the water-cooled condenser, drop by drop fell into the cellulose cartridge. In that moment, extraction took place. Once the Soxhlet was full, the extract fell to the balloon by

a pressure change. Then, the cycle was repeated again. More solvent was evaporated and entered in contact with the extract. Soxhlet system is very efficient since it allows the recovery of a large quantity of plant extract, given that in each round of solvent evaporation only fresh water enters in contact with the dried plant. After an overnight process, which equals to 5-6 solvent cycles, the extract was frozen-dried.

Ethanollic macerations were afforded by incubation of dried grinded plant with pure ethanol in a closed glass recipient. After 30 days at room temperature, the extracts were solvent evaporated. In order to recover the extract from the balloon in the cleanest possible way, a mixture of 50:50 H₂O/Acetonitrile was added. This solution was then transferred to tubes and finally it was frozen-dried.

2.1.2.3. Plant extracts screening

For each extract, a stock of 25 mg/mL was done. Aqueous extracts were solubilized in Milli-Q H₂O, while for ethanolic extracts Dimethyl sulfoxide (DMSO) was used. The DPP IV inhibitory activity of the extracts was calculated using a final concentration of 500 µg/mL (figure 2.2.).

It is noteworthy to point out several aspects.

First, some extracts present negative inhibition. It means that DPP IV was more active. This effect may be explained by the presence of small molecules in the extract that increase the hydrolytic activity of DPP IV. Though no compound has been previously reported as a DPP IV activator, it is a possibility that cannot be discarded. Enzymes have evolved in order to have a controlled activity in a set of cellular conditions. Inhibitors and activators may be able to time-selectively modify the enzyme activity, and DPP IV may have a hotspot in its surface that allows its activation. This hypothesis has to be proved though.

Second, in general, ethanolic extracts are more potent in DPP IV inhibition. Differences between aqueous and ethanolic extracts are visible, but for some extracts, like TCM-LM-04, TCM-LM-17 and TCM-TT-16, percentage of inhibition is similar.

Third, extracts TCM-LM-15 and TCM-TT-16 present high DPP IV inactivation ratio. However, these plant extracts were not further studied since the major component of them is berberine, an already known DPP IV inhibitor.

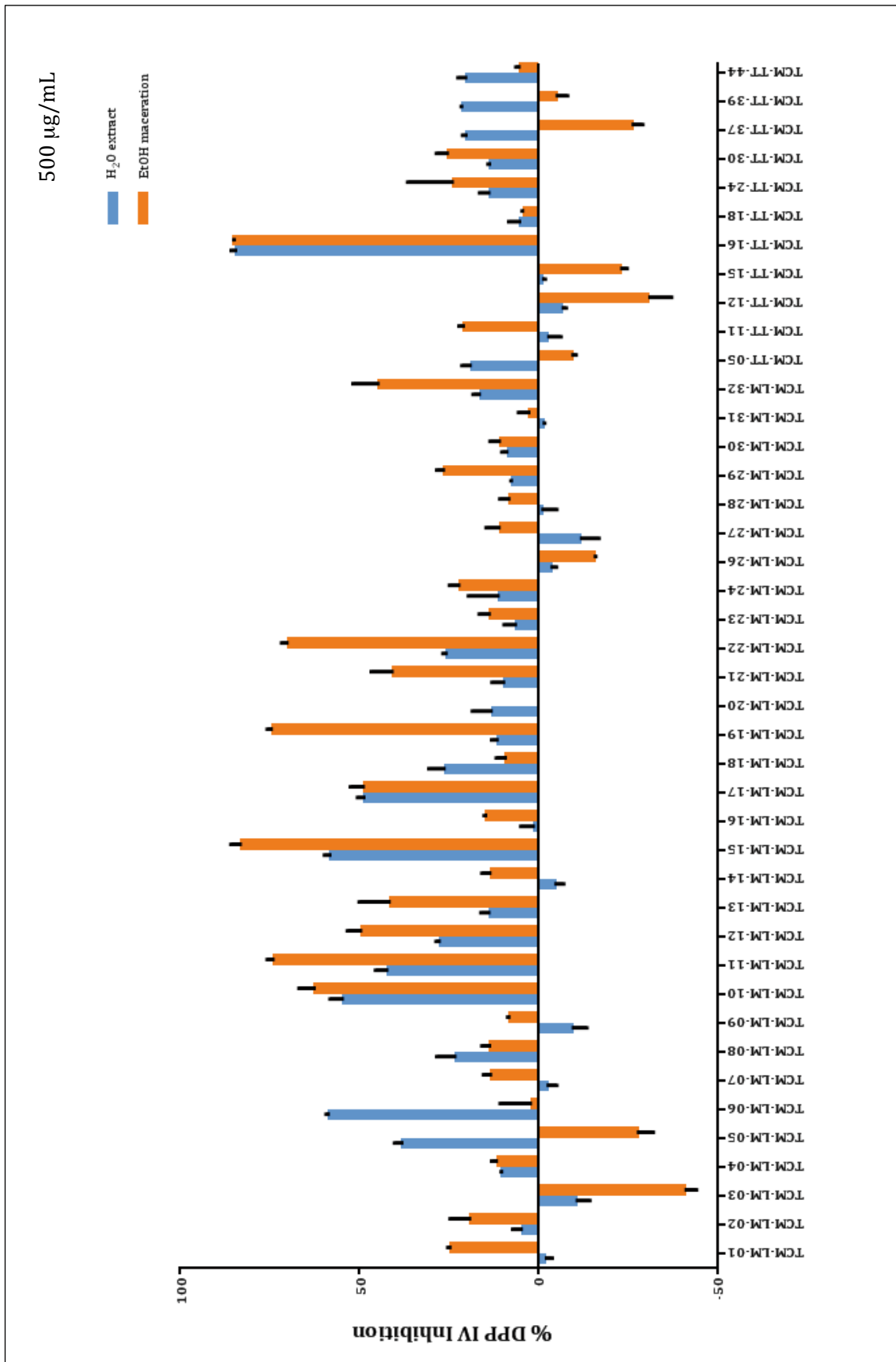


Figure 2.2.: DPP IV inhibition potency for the 43 Chinese medicinal plant extracts.

2.1.3. Brazilian plants

It is estimated that around 50% of the plant species in the world are located in the tropical rainforest. ^[182] These areas are in extreme danger due to the deforestation and efforts are devoted to protect them. Brazil, with a forest area of 477,698,000 hectares (which accounts for a total of more than 50% of the land area), is one of the countries with a richest ecosystem. Brazil contains 6 main biomes: Amazon, Cerrado, Pantanal, Caatinga, Pampa and the Atlantic forest, making of it a relevant place for phytochemical research. Traditional Brazilian medicine has taken advantage of this diversity and is based in the huge number of botanical species. Indigenous people have used it since centuries ago and later, with the immigration of slaves from Africa and settlers from Europe, the flora richness increased even more. Basically, Brazilian herbalists make hydro-alcoholic extractions of plants with the sugarcane juice spirit *cachaça*. The ensemble of plants are used for a set of conditions, like inflammation, allergies and parasitic and infectious diseases, but extracts have proved to be efficacious as well in the nervous and digestive systems, as wound healers and in cancer and cardiovascular diseases.

In the field of diabetes, *Croton cajucara* has proved to be effective as a sugar-blood reducing agent (among other properties). ^[183] However, its use is polemic due to the controversy between studies suggesting hepatotoxicity ^[184] and other demonstrating absence of it.^[185]

In an effort to protect the exploitation of the Brazilian flora, the country has specific federal laws. For example, plant collection is only allowed after authorization of the Brazilian Institute for Environment and Renewable Resources. Besides, the ownership of the Brazilian biodiversity is assigned to the Brazilian indigenous people. This innovative law was set in order to return part of the patent profit to the indigenous people that have allowed the discovery of the drug with their botanical knowledge. It is remarkable to point out that 95% of the patents belong to developed countries, despite the fact that indigenous cultures are promoters of part of their discovery. ^[182] Another measure that the Brazilian government has taken for the “plant-to-drug” protection is that only pharmaceutical compositions are accepted for patenting. Plant extracts are then not a subject for patenting.

Brazilian botanical diversity is exceptionally large and rich. Yet, only a small part of it has been studied. From this percentage of characterized extracts, just a low percentage has been translated into therapeutic patents. Besides, the major part of the investigation in this field is developed by Brazilian research centres. ^[182] Then, the field of Brazilian plants as medicinal drug sources is very interesting, since the actual exploration has covered a small fraction, leaving a huge potential to be unveiled.

With this scenario, a search for DPP IV inhibitors using Brazilian plants as source was envisaged. Given the high variety of biomes, the area of the Pampa, in the south of the country, was the selected one. This work was done in collaboration with Prof. Ionara Dalcol, at the “Universidade Federal de Santa Maria”, in Rio Grande do Sul.

2.1.3.1. Selection

Selection was based in the bibliography. [186, 187] Plants with codes from BP-1 to BP-11 were acquired in local herbalist stores. Plants with codes BP-12 to BP-18 were collected in the area of Caçapava (Rio Grande do Sul). Fresh plants were then dried in the oven and once dried, they were grilled.

| Reference | Portuguese | Latin | English | Part of plant |
|-----------|------------------|---|-----------------------|------------------|
| BP-1 | Alcachofra | <i>Cynara scolimus</i> | Artichoke | Leaves |
| BP-2 | Cascara sagrada | <i>Rhamnus purshiana</i> | Cascara buckthorn | Bark |
| BP-3 | Dente de leão | <i>Taraxacum officinale</i> | Common dandelion | Stalk and leaves |
| BP-4 | Erva de Bicho | <i>Polygonum hydropiper</i> | Water-pepper | Stalk and leaves |
| BP-5 | Erva Doce | <i>Pimpinella anisum L.</i> | Anise | Seed |
| BP-6 | Jambolão | <i>Syzygium jambolanum</i> | Jambul | Leaves |
| BP-7 | Insulina | <i>Cissus sicyoides</i> | Princess vine | Flowers |
| BP-8 | Macela | <i>Achyrocline satureioides</i> | Macela | Stalk and leaves |
| BP-9 | Pata de vaca | <i>Bauhinia forficata</i> | Brazilian orchid tree | Leaves |
| BP-10 | Pfaffia | <i>Pfaffia paniculata</i> | Brazilian ginseng | Bark |
| BP-11 | Stévia | <i>Stevia rebaudiana</i> | Stevia | Leaves |
| BP-12 | Aroeira | <i>Schinus lentiscifolius Anacardiaceae</i> | Pepper tree | Leaves |
| BP-13 | Aroeira | <i>Schinus lentiscifolius Anacardiaceae</i> | Pepper tree | Roots |
| BP-14 | Cancorosa | <i>Maytenus ilicifolia Celastraceae</i> | Cancorosa | Leaves |
| BP-15 | Cancorosa | <i>Maytenus ilicifolia Celastraceae</i> | Cancorosa | Roots |
| BP-16 | Mamica de cadela | <i>Zanthoxylum rhoifolium Rutaceae</i> | "Mamica de cadela" | Bark |
| BP-17 | Murta | <i>Blefarocalix ilicifolia Mirticeae</i> | Murta | Leaves |
| BP-18 | Pitanga | <i>Eugenia uniflora Mistaceae</i> | Brazilian cherry | Leaves |

Table 2.4.: 18 Brazilian medicinal plant parts used for the search of DPP IV inhibitors.

2.1.3.2. Extraction

Hydro-alcoholic extractions (70% Ethanol / 30% H₂O) were done for each plant (or plant section), since this is the solvent used in the traditional Brazilian medicine. Soxhlet apparatus was used as previously for Chinese plants. The dried grinded plant was introduced in the cellulose cartridge, and then part of the solvent was added to it, while the rest was introduced in the balloon. The system was heated for 5 hours. After that, the extract was collected and a fresh batch of solvent was added. It was left overnight without heating. Next morning, the system was heated again and left it for 5 more hours. The second extract was collected and incorporated with the first one. Solvent was evaporated and after, H₂O was added to recover the maximum possible amount of extract. Finally, it was transferred to glass pots and frozen-dried.

2.1.3.3. Plant extracts screening

For each extract, a stock of 25 mg/mL in DMSO was done. A final concentration of 500 µg/mL was assayed against the DPP IV activity and the percentage of inhibition was calculated (figure 2.3.).

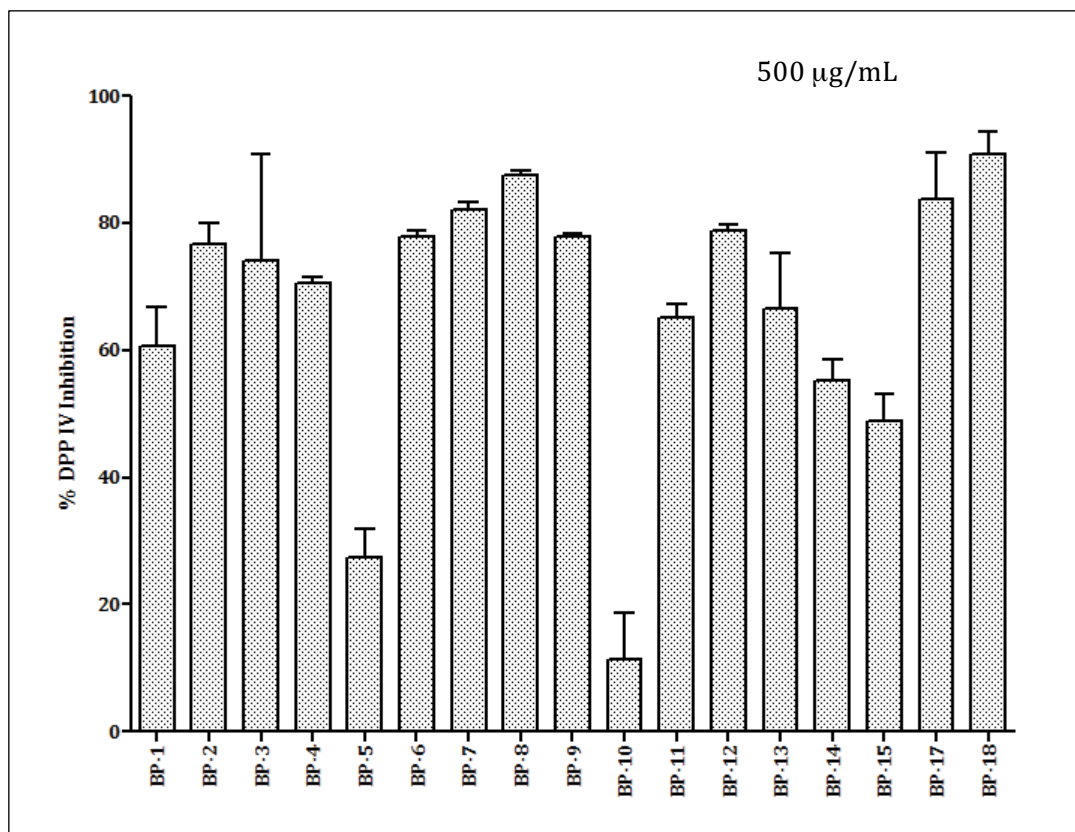


Figure 2.3.: DPP IV inhibition potency for the Brazilian medicinal plant extracts.

Except for BP-3 and BP-10, all the extracts presented a moderate-to-high DPP IV inhibition. The two bests extracts corresponded to Macela (BP-8) and Pitanga (BP-18).

2.1.4. Mediterranean plants

A library of Mediterranean plants was available in the laboratory. Despite that it was a general collection of plants, which was not filtered for diabetes; it was decided to test the extracts against DPP IV activity.

2.1.4.1. Selection

The library consisted in nine plants. From these, all were tested except for MP-5 (Aloe vera), which was previously selected as antidiabetic plant.

| Reference | Catalan | Latin | English |
|-----------|-------------|-------------------------------|-------------------|
| MP-1 | Brotònica | <i>Stachys betonica</i> | Betony |
| MP-2 | Cor de Roca | <i>Hypericum ericoides L.</i> | Hypericum |
| MP-3 | Espígol | <i>Lavandula officinalis</i> | Lavender |
| MP-4 | Menta | <i>Mentha spicata</i> | Mint |
| MP-5 | Àloe vera | <i>Aloe vera</i> | Aloe |
| MP-6 | Sajolida | <i>Satureja hortensis</i> | Summer savory |
| MP-7 | Santònica | <i>Stachys heraclea</i> | Stachys |
| MP-8 | Tila | <i>Tilia cordata</i> | Small-leaved Lime |
| MP-9 | Tomaquera | <i>Solanum lycopersicum</i> | Tomato plant |

Table 2.5.: Mediterranean plants tested for DPP IV inhibition.

2.1.4.2. Extraction

Aqueous extraction with the Soxhlet apparatus and Milli-Q H₂O was done for each plant. The summer students Sílvia Vilaprinyó and Nadja Bertleff performed these extractions.

2.1.4.3. Plant extracts screening

For each extract, a stock of 5 mg/mL in DMSO was done. DPP IV inhibitory activity of infusions was measured at a final concentration of 100 µg/mL of extract (figure 2.4.)

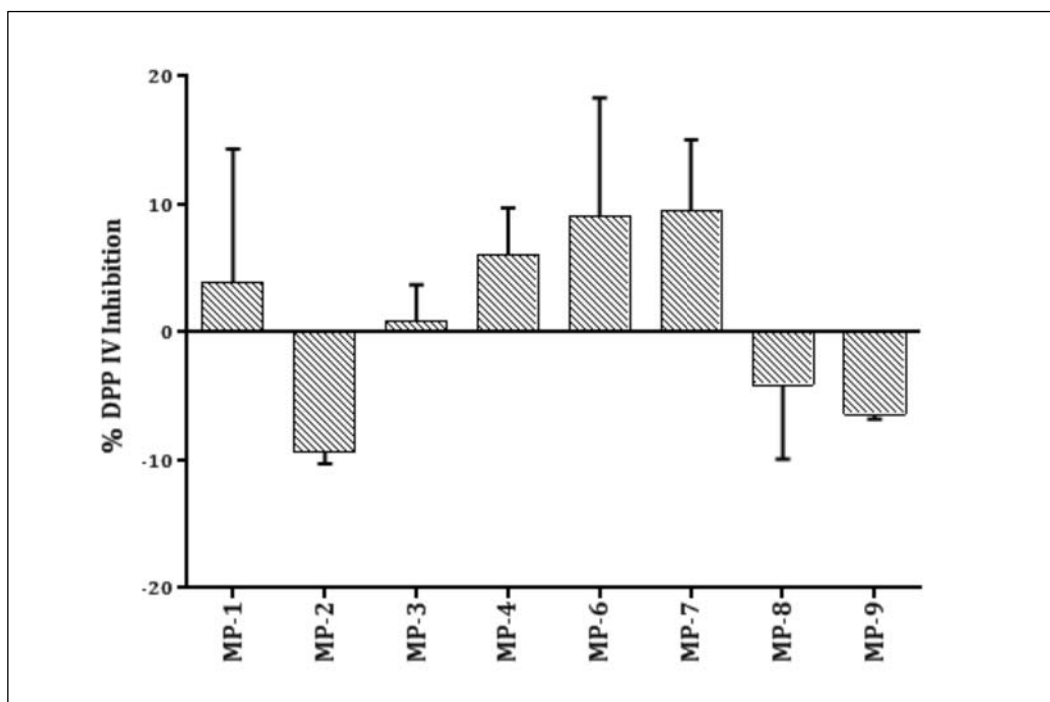


Figure 2.4.: Mediterranean plant extracts parts used for the search of DPP IV inhibitors.

As detected for some Chinese plants extracts, an activation of the protease was observed, being the Hypericum (MP-2) the one exerting a higher increment in the DPP IV activity. Regarding inhibition, Mediterranean plants didn't afford a high DPP IV inactivation. The fact that these plants were not filtered for sugar-blood lowering properties could be a reason for this low efficacy as inhibitor source. Thus, plant selection was proved as a crucial step in drug discovery.

2.2. *Coutarea latiflora*

Coutarea latiflora (also known as *Hintonia latiflora*, Copalchi, Copalquin) (AP-3) demonstrated a high DPP IV inhibition rate, both in the aqueous and the ethanolic extracts (76% and 79%, respectively, for a final concentration of 500 µg/mL). Aimed by this encouraging result, a deeper study of this plant was designed.

The hypoglycemic effect of the Mexican plant *Coutarea latiflora* has been extensively studied. [188-192] Mexican Traditional Medicine is remarkably rich, including an enormous variety of plants with hypoglycemic effects. A thorough review by Andrade-Cetto and Heinrich described 306 Mexican plant species that were used as sugar-blood lowering agents. [193] Among them, Copalchi neoflavonoids and the molecule coutareagenin are defined as the responsible molecules for the effect. Despite the numerous studies, only one of them proved a mechanism of action for the molecules. In it, a group of compounds from Copalchi extract were found to inhibit α -glucosidase. [192] This enzyme is located in the brush-border of the small intestine and is responsible for the carbohydrate hydrolysis, releasing α -glucose. Inhibitors of α -glucosidase delay glucose absorption, and thus, attenuate postprandial hyperglycemia. However, no direct relationship between Copalchi and DPP IV inhibition has been previously found.

Based on the preliminary results, the objective was to identify the active molecule or molecules for DPP IV inhibition present in *Coutarea latiflora* infusion. The ethanolic extract was selected in front of the aqueous one due to the higher observed DPP IV inhibition. In order to find the active principle, the strategy started with the fractionation of the extract, using a reversed-phase column. Then, fractions were tested for DPP IV activity. If inhibition was found, the fraction was analyzed for molecule characterization. These studies were performed by Albert Puigpinós at IRB Barcelona. Last step consisted in a fully characterization of the inhibition mechanism.

2.2.1 Extract fractionation and compound isolation

Purification of the ethanolic extract of *Coutarea latiflora* was performed in a HPLC instrument equipped with a C18 column. The gradient that afforded an optimal resolution was 10-40 % (H₂O, 0.1% TFA-ACN, 0.1% TFA). Each peak was separated and collected for the DPP IV activity assay. Peaks corresponding to retention times 3.79 and 5.10 minutes were positive as DPP IV inhibitors and were named AP-3-a and AP-3-b respectively (figure 2.5.).

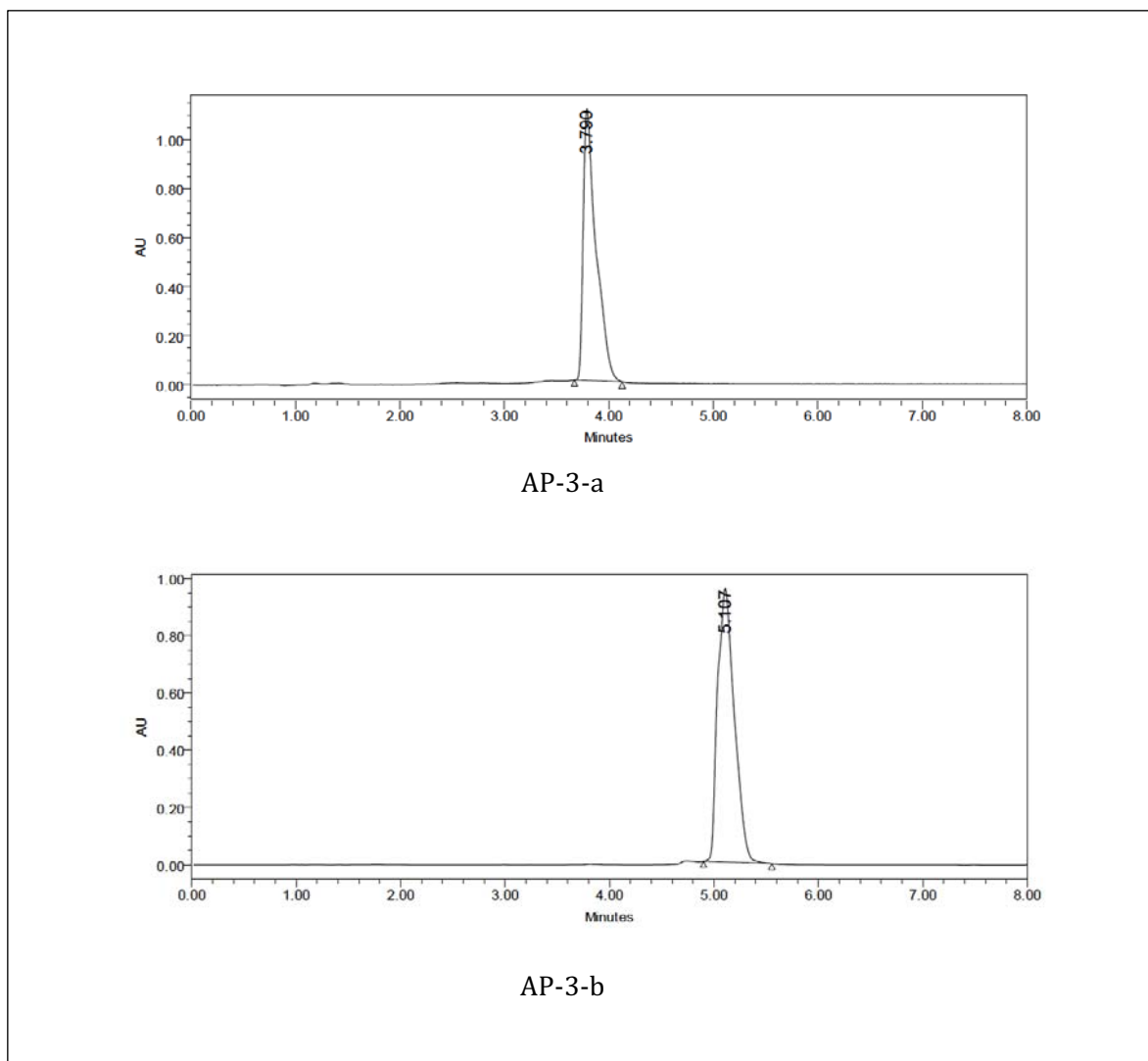


Figure 2.5.: Chromatograms of purified molecules from extract plant AP-3 with DPP IV inhibitory activity.

2.2.2 Molecule characterization

AP-3-a and AP-3-b were characterized by Mass spectrometry and also by Nuclear Magnetic Resonance for their structure elucidation.

AP-3-a presented a mass of $[M+H]^+$: 449.10784, and after a thorough analysis by NMR, the structure was unveiled (figure 2.6.). AP-3-b mass was $[M+H]^+$: 463.12353, and the NMR allowed its structure determination (figure 2.6.). Both molecules are flavonoids and have been previously reported in the bibliography. [189, 191, 194] In fact, the molecule AP-3-b has been named as Coutareagenin. In the large panorama of DPP IV inhibitors, the two discovered small molecules represented a new scaffold of inactivators.

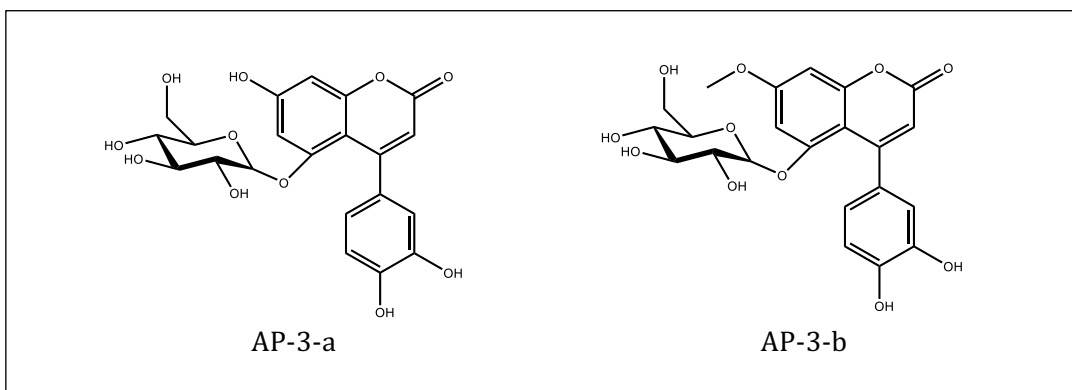


Figure 2.6.: Chemical structures of active molecules from extract plant AP-3 against DPP IV activity.

2.2.3 Inhibitor characterization

This section was devoted to a full study of the mode of DPP IV inhibition of the molecules found in the extract of *Coutarea latiflora*.

2.2.3.1. Study of the inhibition

AP-3-a and AP-3-b IC_{50} values were calculated. DPP IV inhibition rate at a range of diverse molecule concentrations was measured and then represented. AP-3-a IC_{50} was lower (115 μM) than the belonging to AP-3-b, describing a higher DPP IV inactivation (figure 2.7.). For that reason, AP-3-a was selected for further analysis.

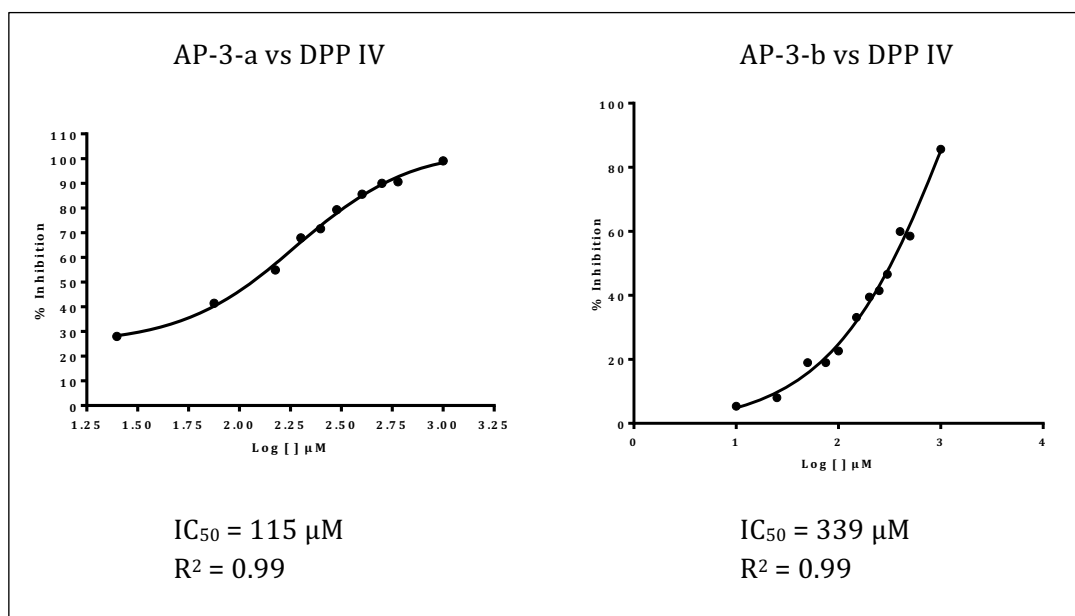


Figure 2.7.: IC_{50} curves DPP IV inhibitors present in the AP-3 ethanolic extract.

It has been previously reported that flavonoids may have tendency to aggregate. [195] Such an aggregation process can affect the enzyme activity. Molecules can trap the protein,

forming a micelle-like structure, thus avoiding the contact with the substrate. In order to prove if AP-3-a was forming micelles, or on the contrary, was inhibiting DPP IV in a selective manner, the inactivation activity was measured in presence of the detergent Triton. Detergents, even in small proportions, disrupt micelles. By comparison of the protease inhibition in presence and absence of detergent, it was concluded that AP-3-a was inhibiting the DPP IV without forming micelles (figure 2.8. a).

Inhibitors may be reversible or irreversible. In medicinal chemistry, the general tendency is to look for reversible inactivators, since fewer side effects are expected. By variation in the AP-3-a /DPP IV pre-incubation times, the time course graphic was obtained. As observed in figure 2.7., DPP IV inhibition is concentration dependent, but pre-incubation time independent (figure 2.8.b). It was concluded then that AP-3-a was inhibiting the protease in a reversible manner.

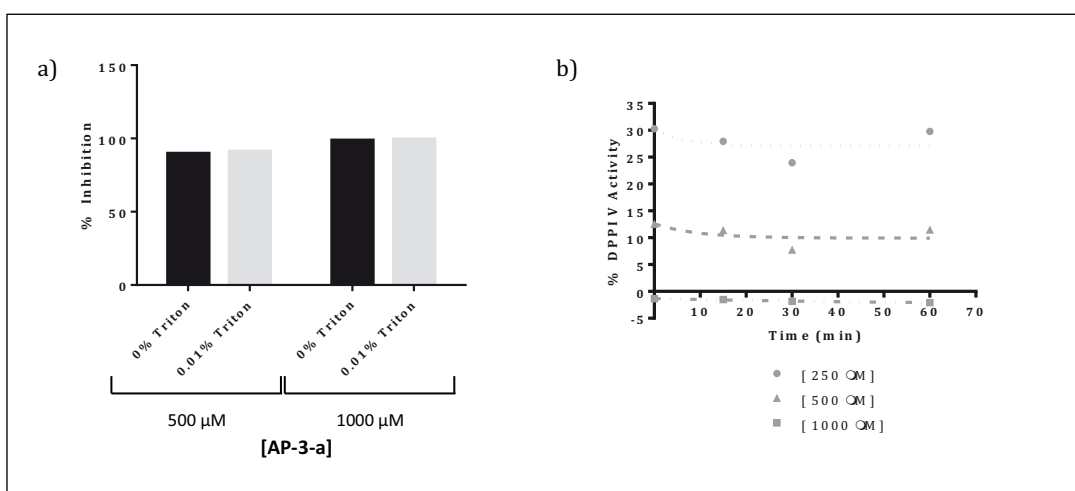


Figure 2.8.: a) DPP IV inhibition performed by AP-3-a in presence or absence of the detergent Triton; b) Time dependence curve for AP-3-a vs DPP IV.

Furthermore, the activity of the aglycone of AP-3-a was assayed. As reported for other plant-derived molecules, the presence of the sugar moiety is not necessary for the interaction with the target, or in other cases, can substantially modify the activity.^[196] Treatment of AP-3-a with TFA under high temperature allowed the removal of the glycosidic group. DPP IV activity was then tested. The aglycone of AP-3-a showed an inhibition rate of 39.77% at 50 μM. AP-3-a presented a similar percentage of DPP IV inactivation at the same concentration (41%). It was demonstrated then that the sugar moiety was not essential for the activity of the compound.

Finally, selectivity of AP-3-a for DPP IV versus the closely related protease POP was tested. In table 2.6. the percentage of protease inhibition for two AP-3-a concentrations is depicted. The molecule had more affinity for DPP IV, but it was a POP inhibitor as well (ratio 1.13). Thus, AP-3-a was not selective for DPP IV. However, POP is located in the CNS, and flavonoids with sugar moieties do not cross the BBB. (18650094) Then, AP-3-a, despite being an *in vitro* POP inhibitor, it is expected not to be active *in vivo*.

Moreover, AP-3-a was identified as a new DPP IV inhibitor, but further modifications in its structure could lead to more potent and selective molecules.

| AP-3-a | % Inhibition | | Ratio |
|-------------|--------------|------|-------|
| | DPP IV | POP | |
| 100 μ M | 43.4 | 38.3 | 1.13 |
| 200 μ M | 67.9 | 60.0 | |

Table 2.6.: DPP IV and POP inhibition by the molecule AP-3-a.

2.2.3.2. Inhibition mechanism

A further analysis of the behavior of AP-3-a as a DPP IV inhibitor was carried out. In order to describe the mechanism of inactivation, kinetic experiments were done.

Different concentrations of AP-3-a were assayed for DPP IV activity at a range of substrate concentrations. For each inhibitor concentration, the progress-curves (Fluorescence vs time) were plotted (one curve/[S]). From these graphics, the linearity time was extracted and set to 20 minutes. Then, velocity rates were calculated for each inhibitor and substrate concentration.

The Lineweaver-Burk linear-transform of the data afforded a set of straight lines without a common intersection point (figure 2.9.). In a zoomed-view of the plot, it was observed that the intersection points were moving to lower substrate concentration (higher $1/[S]$) with higher inhibitor concentrations. This was a clear identification that the compound was not acting in a linear-way, but that it was performing a parabolic inhibition, a behaviour only reported once before for DPP IV.^[197]

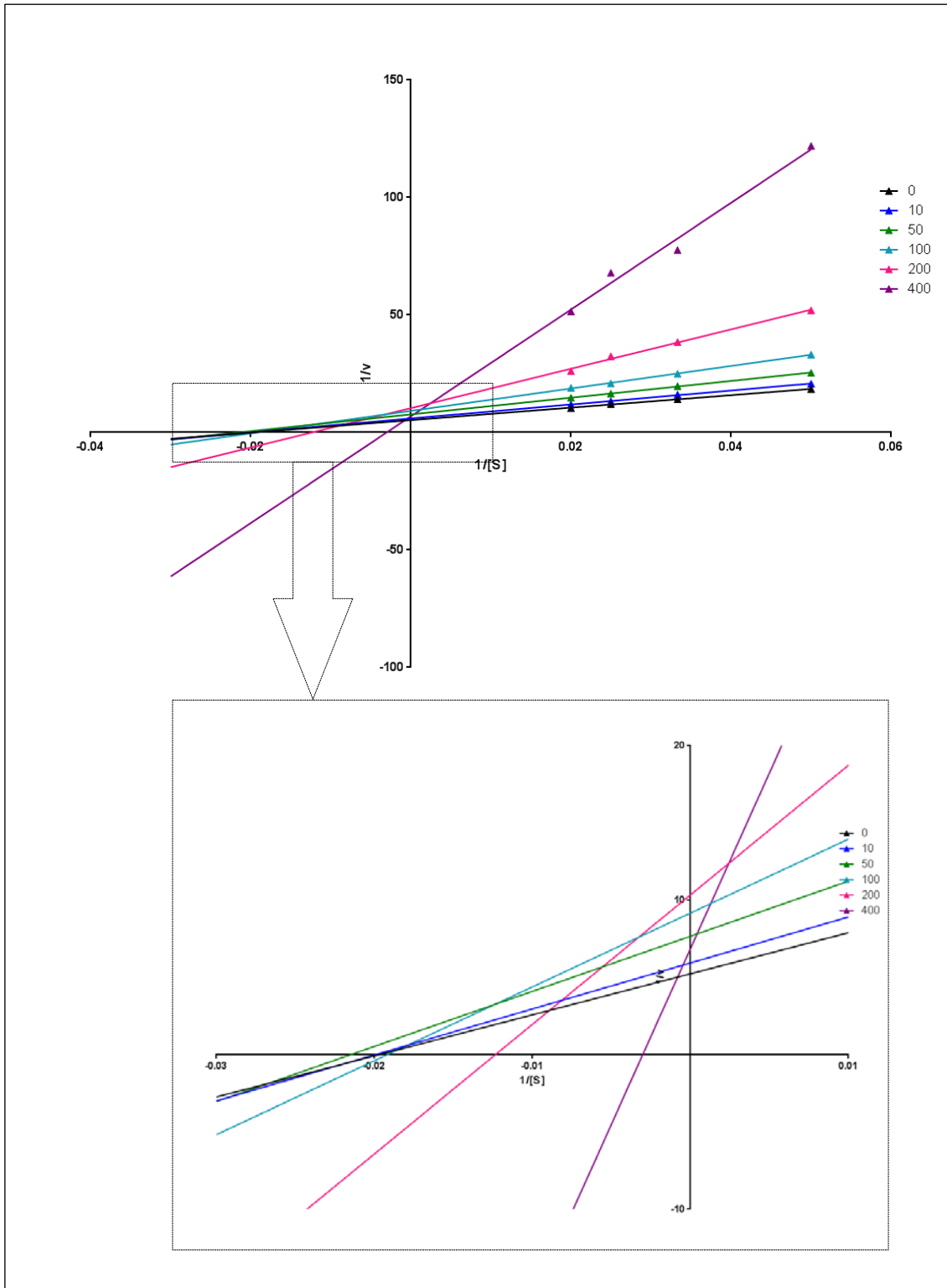


Figure 2.9.: Lineweaver-Burk plot of AP-3-a inhibition of DPP IV. In the square a zoom of the crossing area is depicted.

Linear inhibition consists in a single-site interaction of the inhibitor with the protein. Typically, there are three types of reversible-linear inhibition: competitive, non-competitive and uncompetitive. Competitive inhibitors bind to the same area of the protein where the substrate does, thus avoiding the formation of the ES complex. Non-competitive inhibitors, on the contrary, interact with a different protein area, but modify the biomolecule in such a way that the substrate is unable to bind to it. The last scenario is the uncompetitive inhibitor, where the substrate can interact with the protein despite the presence of the inhibitor, forming an IES complex.

Non-linear inhibition encloses hyperbolic and parabolic inhibition. The first one takes place when the IES complex is not inactive, but product can be formed. Parabolic inhibition is performed when a molecule binds to the protein in two sites, simultaneously or not. The kinetic model of a parabolic inhibitor is represented in figure 2.9.

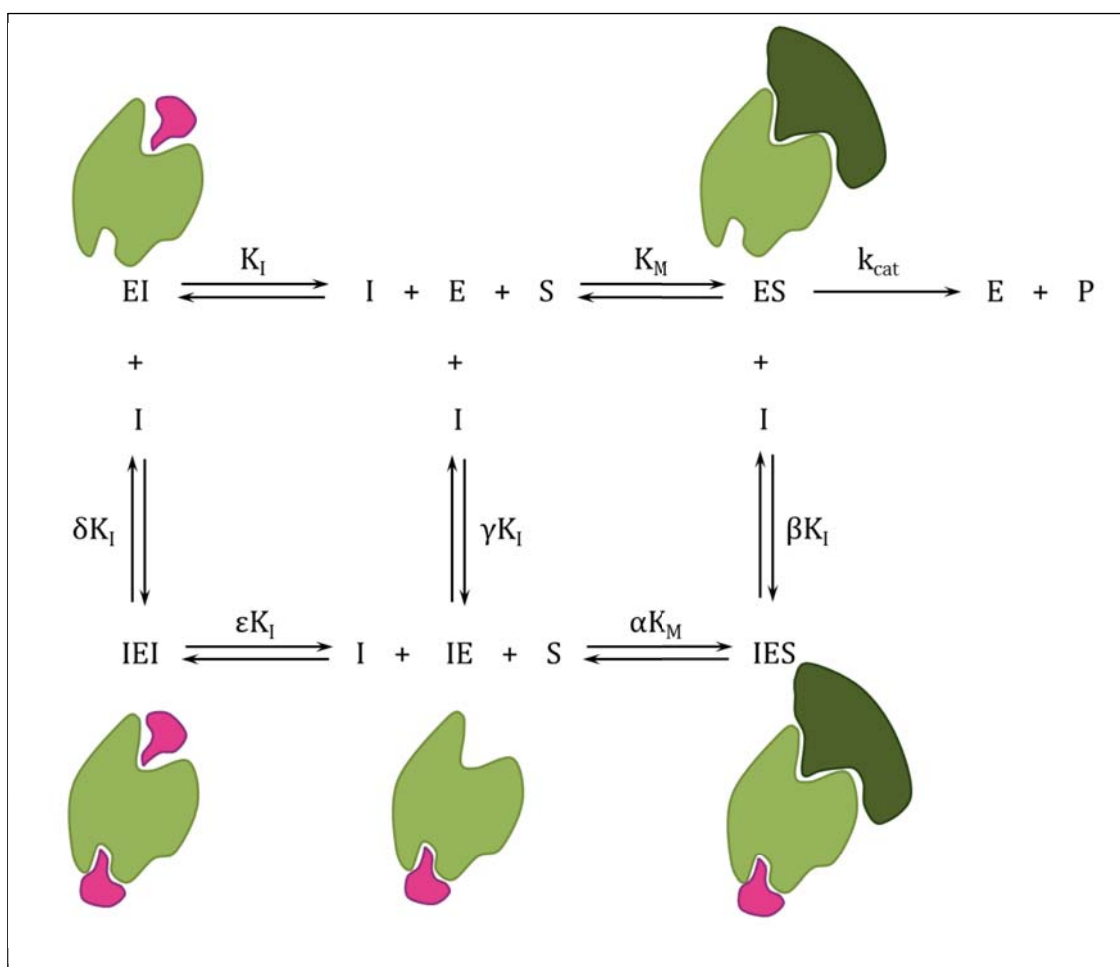


Figure 2.9.: Lineweaver-Burk plot of AP-3-a inhibition of DPP IV. In the square a zoom of the crossing area is depicted.

When $\alpha = 0$ and $\beta = 0$, then the molecule is performing a parabolic competitive inhibition. If $\alpha \neq 0$ and $\beta \neq 0$, the compound is inhibiting in a non-competitive parabolic manner.

The value of the constants can be calculated with the secondary plots. The slope-plot consists on a representation of the slopes from Lineweaver-Burk lines in front of the inhibitor concentration, while the y-intercept plot represents the intercept of each of the lines versus the compound concentration.

The slope-plot data was fitted into a parabolic curve, demonstrating the inhibition mechanism. However, the y-intercept secondary plot fitted into a linear mode. This behavior corresponded to a slope-parabolic intercept-linear noncompetitive parabolic inhibitor. Besides giving qualitative information, the secondary plots equations allowed the calculation of the enzymatic reaction parameters. Each equation was interpreted in terms of the enzymatic equilibrium, giving a relationship between the constants and the experimental data (figure 2.10.). The obtaining of the equations is detailed in the experimental section.

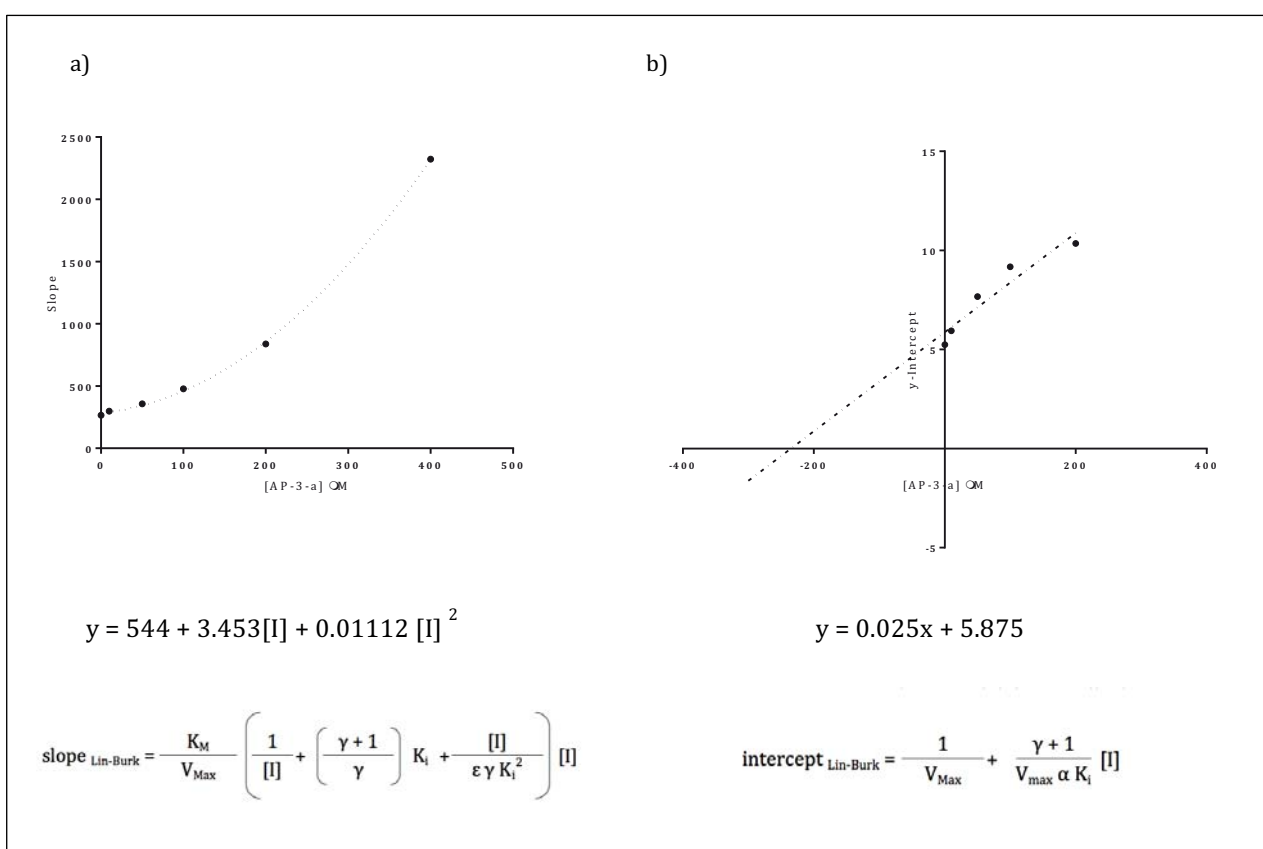


Figure 2.10.: Re-plots of Lineweaver-Burk. a) Slope re-plot; b) Intercept re-plot

The system was solved with the equations by an iterative process, since the number of equations was less than the number of unknown parameters (experimental section). Constants and factors are pointed out in figure 2.11.

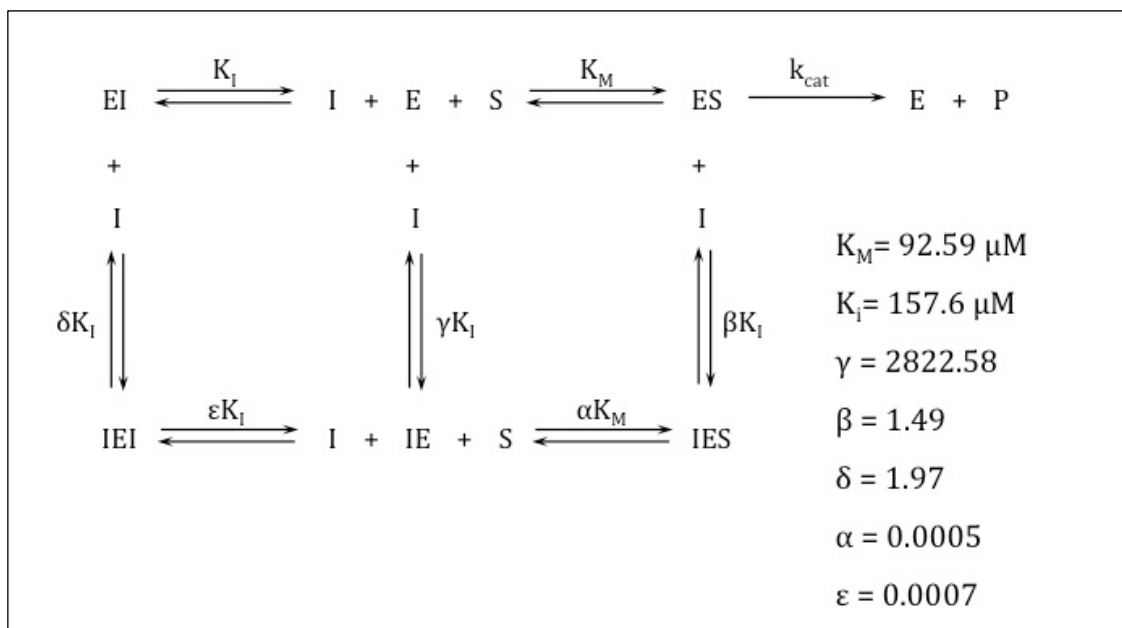


Figure 2.11.: Constants and factors for AP-3-a inhibition of DPP IV.

As deduced from the Lineweaver-Burk re-plots, AP-3-a was exerting a noncompetitive parabolic inhibition. Thus, the molecule can bind into two sites of the protein, however, affinity for them is different.

Once the enzyme is free, the substrate was preferentially bound. Affinity of AP-3-a for the competitive site was slightly greater than the K_M ($K_i = 157.6 \mu M$), but binding in the non-competitive site was clearly improbable ($\gamma = 2822.58$). Interestingly, this pocket was only occupied after binding of the substrate or the inhibitor at the competitive site. If the substrate was bound, AP-3-a had an affinity of $234.8 \mu M$ (βK_i) for the non-competitive site, while if there was an inhibitor molecule in the competitive site, affinity for the occupation of the second pocket was of $310.5 \mu M$ (δK_i). In both cases, the exit of the molecule on the competitive site was not likely to occur, as α and ε were close to 0.

Thus, the kinetic model could be simplified as indicated in figure 2.12.

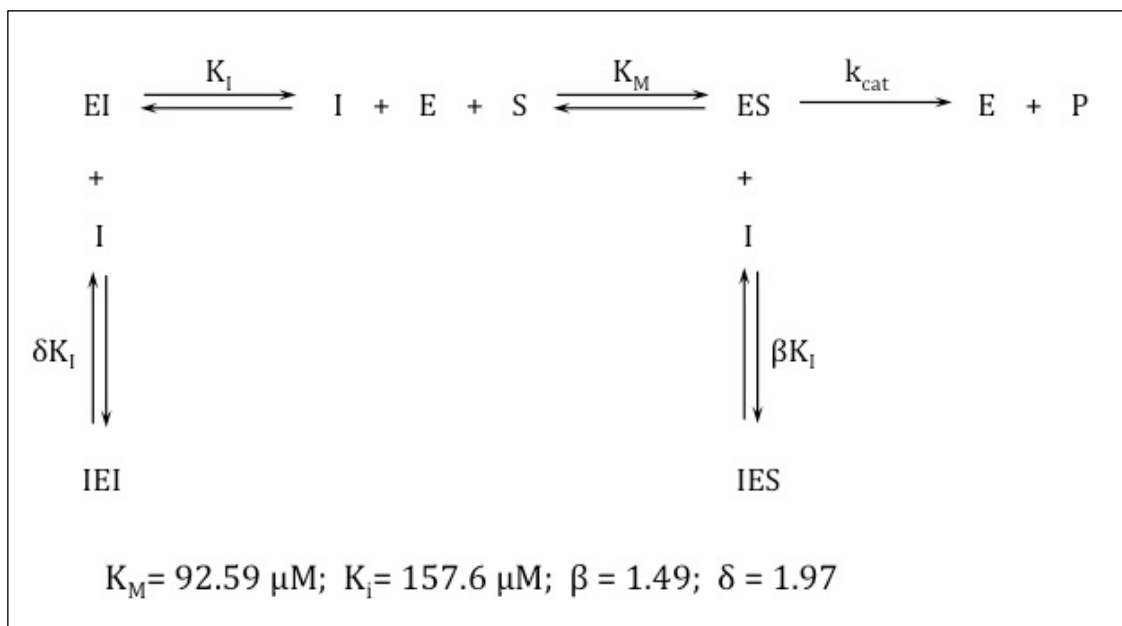


Figure 2.12.: DPP IV inhibition mechanism performed by AP-3-a-

In this scenario, the non-competitive site was only accessible after binding of a molecule (either inhibitor or substrate) into the competitive site. Thus, this interaction promoted a conformational change that opened a new cavity or unburied the non-competitive site. The substrate interaction is more efficient in unveiling this second binding position, as deduced from the factors. Interestingly, AP-3-a, despite being a small molecule, was able to produce a similar effect to the substrate, and performed the conformational change needed for the access to the non-competitive site.

2.2.3.3. NMR Analysis

The last step of AP-3-a characterization consisted in the study of the interaction with labeled DPP IV by NMR. It allowed the detection of possible changes in the chemical shift of the methionines after binding of the molecule to the protein. As seen in figure 2.13. no significant changes were observed between the free DPP IV compared to the sample of DPP IV with AP-3-a. The signal named as D in section 1.3.4 was not visible. This signal appeared in presence of DPP IV inhibitors, despite of their chemical natures. The fact that signal D was not present when incubating DPP IV with AP-3-a could be due to the low affinity of AP-3-a compared to the control molecules used in chapter 1 for assay optimization. However, in the spectra of DPP IV / Berberine, a molecule that has an IC_{50} of $140 \mu\text{M}$ for DPP IV, signal D was still visible, though its intensity was reduced compared with the one of P32/98. IC_{50} of AP-3-a for DPP IV is smaller ($115 \mu\text{M}$), then, the low affinity was not considered as a crucial factor. However, as observed by the kinetic assay, the molecule interacts with two sites of the protease. If the signal corresponding to D is visible when the inhibitor is binding to the competitive site, but it is not after binding of the second molecule in the non-competitive site, then, AP-3-a will give 2/3 of the expected signal, since the affinity for binding in the second position is twice the one to the

competitive site. This reduction in the competitive site-inhibitor/protein population will cause a loss in the D signal.

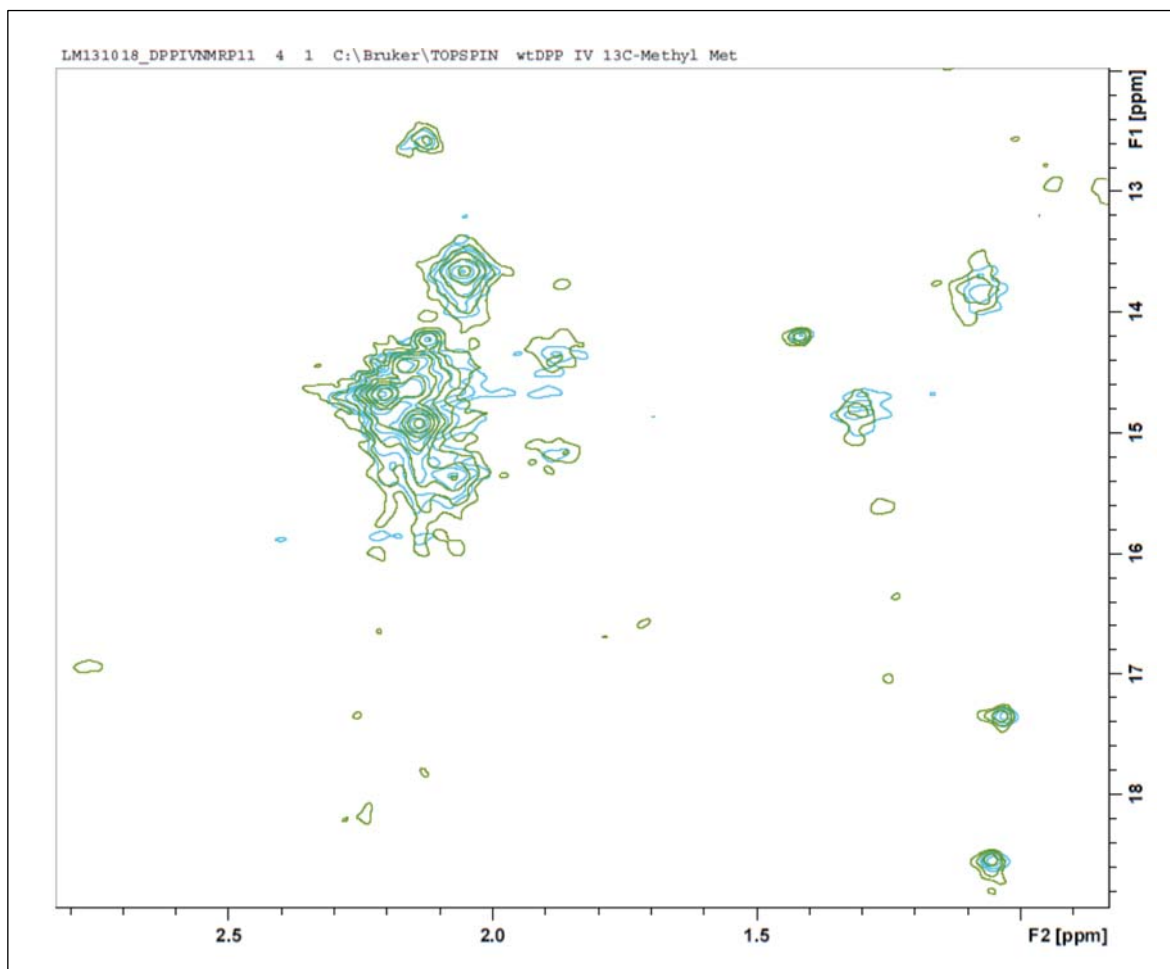


Figure 2.13.: Methyl-TROSY spectra of methyl- ^{13}C methionine labeled DPP IV (39-766) in complex with AP-3-a (green) versus free DPP IV (blue).

The latter hypothesis would imply that binding of an inhibitor into the competitive site causes a minor conformational change that does not occur if the inhibitor is binding to the non-competitive binding site. This does not imply that union to non-competitive site is not modifying the protein conformation, but if it occurs, it would be a different one from the one observed from competitive inhibition.

Chapter 2 overview

We planned to find DPP IV from botanical sources.

First, we selected plants that were already reported to have antidiabetic action.

Common antidiabetic plants were chosen, as well as Brazilian plants and others from the Traditional Chinese Medicine. Besides, a library of Mediterranean plants was also selected. After, extraction and testing of DPP IV inhibitory activity was done.

From our tailored collection, the plant AP-3 was selected for further analysis. After fractionation and purification, two molecules were found to be DPP IV inhibitors. The best one, AP-3-a was submitted to a thorough enzymatic analysis.

Kinetic experiments of AP-3-a demonstrated that it was inhibiting DPP IV in a parabolic manner. This behaviour is extremely rare in proteins with a single anchorage substrate and has only been reported once before for DPP IV.^[197]

Then, AP-3-a inhibition of DPP IV was analysed by NMR.

Despite the resolution of the spectra did not match the quality requirements, the extra signal that was observed with competitive inhibitors was not present.

We hypothesized that the lack of appearance of this signal is a result of the parabolic inhibition of AP-3-a.

Chapter 3:
Discovery of POP inhibitors by
High Throughput Screening

CHAPTER 3 CONTEXT**3.1. SET-UP FLUORESCENCE POLARIZATION EXPERIMENT**

3.1.1. Fluorescence Polarization main concepts

3.1.2. Probe design and synthesis

3.1.3. FP assay optimization

3.2. HIGH THROUGHPUT SCREENING

3.2.1. HTS experiment

3.2.2. Analysis of the results

3.2.2.1. Calculation

3.2.2.2. Evaluation

3.2.2.3. List

3.3. HIT VALIDATION

3.3.1. Clustering

3.3.2. Docking

3.3.3. Enzymatic assay

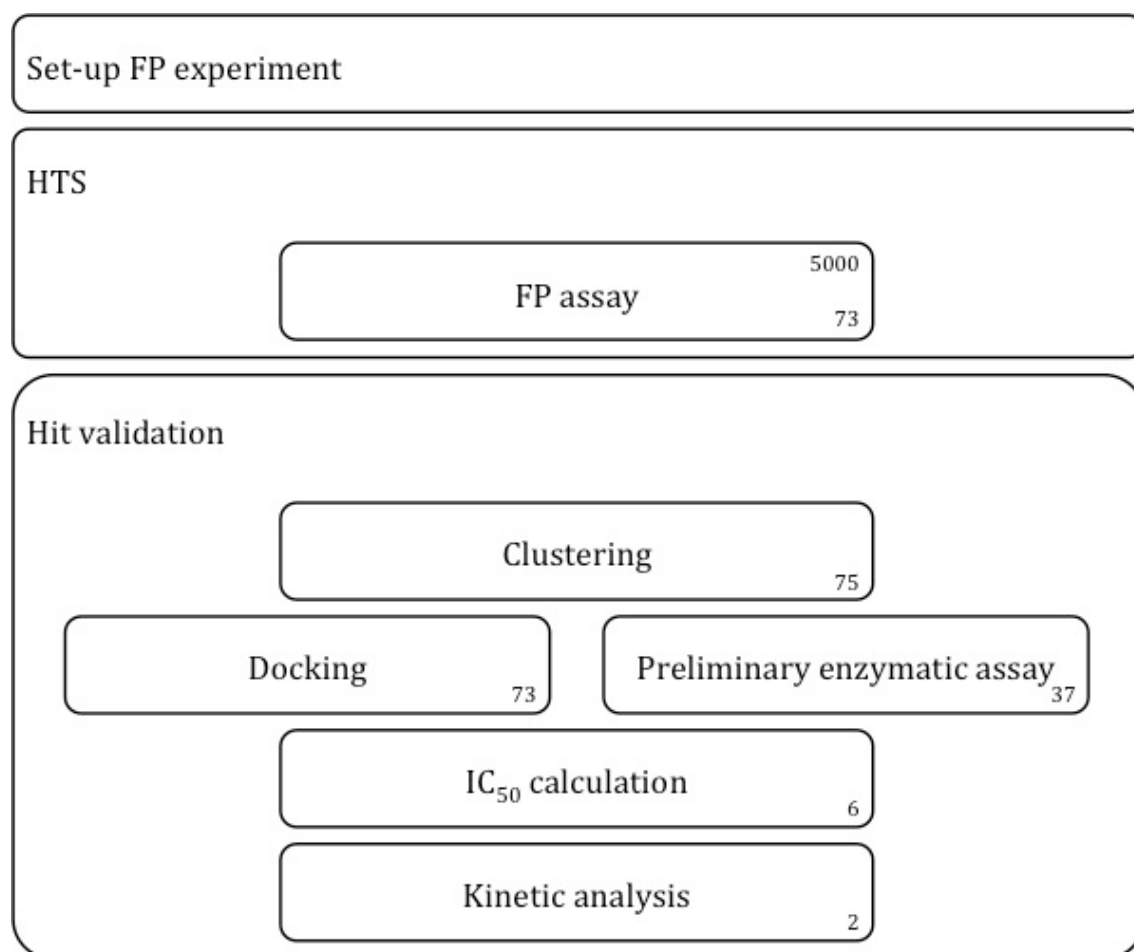
3.3.3.1. Preliminary screening

3.3.3.2. IC₅₀ curves**3.4. KINETIC STUDIES OF LEAD CANDIDATES**

3.4.1. Kinetic constants elucidation

CHAPTER 3 OVERVIEW

Experimental scheme



Numbers on lower corners refer to the number of molecules in each step.

Chapter 3 context

In the decade of the 90s, molecular biology allowed the discovery and identification of molecular targets. It represented the possibility to tackle diseases in a more efficient manner, and pharmaceutical industries rapidly focused their efforts in finding drugs addressing these new targets. In parallel, combinatorial chemistry methods were optimized affording the creation of large chemical libraries. In this scenario of large molecule series synthesized by combinatorial chemistry and new targets discovered by molecular biology, a demand for testing methodologies was originated. High-throughput screening (HTS) was then proposed as a testing method that combined microtitre plate-based assays and automated screening platforms.

Among the features of HTS, it assumes no *a priori* knowledge of the binding site of the molecule into the target, it allows the identification of new chemical entities as drugs and it is applicable to a broad class of targets.^[26] However, its main disadvantage is its cost, due to the high technology infrastructure required.^[198] Having evaluated all factors, HTS was envisaged as a promising technique and thus, pharmaceutical companies devoted a substantial effort in HTS application in drug discovery, expecting the finding of new and potent compounds, even the possibility of finding first-in-class drugs.

But despite the huge investment in HTS by pharmaceutical industries and governments and its worldwide application, results have not been as expected. In a study from 2004, where a group of laboratories reported their data on success with this technique, 43 laboratories reported 746 leads obtained from HTS in that year.^[199] However, the number of leads that actually entered into clinical phase as candidates was 104.^[199] The percentage of candidates that finally arrive to the market is expected to be considerably lower. From 58 drugs approved between 1991 and 2008, only 19 of them had HTS as their origin.^[198]

Comparison of HTS results with the expected ones when the tool started to be applied in laboratories of all over the world has arisen criticism about the efficacy of the technique in drug discovery. An intense debate has been originated around HTS attrition rates and the main reason underlying the failure of HTS hits to reach the market has been attributed to the library design. In the beginning of HTS implementation, large libraries grouping compounds from combinatorial chemistry were created. Quantity of compounds was considered as a fundamental factor for the success of the technique. The more compounds to test, the more probability to find a hit. Years of experience have demonstrated that quality is further more important than quantity. In order to improve the quality of HTS libraries, several strategies are being implemented. Here, three of them are further discussed.

1) Lead-like properties of the molecules.

HTS rate of hit success is relatively high, therefore, it could be assumed that the subset of compounds present in the library represent a promising pool. However, a hit is not likely to be the final drug. On the contrary, the molecule would be object of subsequent modifications in order to improve affinity, specificity and biophysical properties. While hits are identified in basis of potency, drugs should be not only potent, but also selective and able to reach their target in the organism, crossing barriers such as the intestinal or the blood-brain barriers. Then, molecules with lead-

like properties (or even better, drug-like properties) are favoured in the library design. By this, compounds in the HTS library would be useful as starting points for a successful hit-to-lead programme.^[200]

2) Chemical space

Traditional libraries contained compounds that overlapped in the chemical space. To increase the coverage of it, cheminformatics tools have originated in the last years a series of computational programmes oriented to increase the library diversity.^[198]

3) Focused libraries

The actual trend is to create small focused libraries rather than large and random ones. Smaller subsets can be created based on the target of study. With the help of computational programs, and previous information about target structure or existing or natural ligands, personalized libraries are tailored.^[200]

| | Combinatorial Chemistry | Actual libraries |
|-----------------------------|--------------------------------|-------------------------|
| Drug-like properties | Low | Medium-High |
| Diversity | Limited | Extense |
| Target-focused | - | Optional |
| Library size | Large | Medium |
| Attrition rate | High | Expected to be lower |

Table 3.1.: Combinatorial chemistry versus actual libraries

In general, HTS libraries should have equilibrium between molecule properties, diversity, target focus, intellectual property potential and price.^[198]

Furthermore, by increasing quality, libraries are not required to be large in compound number. This ultimately results in an economic saving and a more efficient process.

The improvement on the quality that the actual libraries offer is expected to lower the attrition rates related with HTS. However, changes in the manner that libraries were designed, were implemented in the beginning of the 2000's. Taking into account that between 10 and 15 years are required for a hit to reach the market, results of this "quality-driven libraries" will be noticeable in the next years.^[198]

One particular case of optimized libraries is the use of marketed drugs or known bioactive compounds. Drug repositioning, repurposing, reprofiling or redirecting, are terms used for the process of finding new uses for existing drugs, with the condition that these new indications have to be not related with the original one.^[201] This technique has suffered a boom in the last years, partially driven by the gap between the high economic pharma investment and the low number of new chemical entities/year.

The advantages of drug repositioning are consequence of the existence of a previous drug discovery and development. Pharmacokinetics is already performed, meaning that the biophysical properties of the molecule have been optimized. Phase I is also done, which implies less development time and less cost for the repurposing. In general, repositioning

candidates have a smaller associated risk, since the molecule has good pharmacokinetic profiles and is proven to be not toxic in healthy volunteers.^[201]

However, this process has three main disadvantages. First one is related to the explored chemical space. As reported by the FDA, 1593 new chemical entities have been approved since 1938 until 2013. Thus, the number of molecules is quite small.

The second disadvantage depends on changes in regulatory standards, which may force repetition of the phase I trials in order to be in agreement with the new regulation. This will cause a raise of the total cost of the drug development.

Finally, the third disadvantage is related with its legal protection and commercial exploitation. The method-of-use (MOU) patent is the document that allows the protection of a compound for its utilization in a certain disease treatment, independently of the product patent. Here, there are two scenarios. If the product patent is still active, then, in order to commercialize the molecule for a second use, the owners of the respective patents must reach an agreement. On the other hand, if the compound is off-patent, the principal controversy arises when the generic is being manufactured worldwide. Thus, prosecution of patent is difficult to follow.

Several examples of drug repositioning exist in the current drug panorama. Early in the 1987, the first drug for HIV treatment was approved, but the molecule, zidovudine, was previously developed in 1964 as an antineoplastic drug.^[201] Another example, and maybe one of the most famous, is sildenafil. Originally developed for angina treatment, several volunteers of phase I study reported that their erections were stronger and more persistent than usual. Time after this finding, Pfizer started a trial on impotent men. Viagra, the trade name for sildenafil in the use for erectile dysfunction, had in 2003, annual sales of \$1.88 billion.^[201] Far more surprising is the case of thalidomide. Originally used to treat morning sickness in pregnant women, it was banned after demonstrating that the drug caused severe skeletal birth defects. Nowadays, despite the devastating effect it had in thousands of children, it is the only treatment for erythema nodosum leprosum (ENL), and is also indicated in HIV and cancer.^[201]

These examples, and others, come from serendipity or novel investigations that led scientists to give the drug an extra application in other conditions. HTS took advantage of this knowledge, and libraries containing marketed drugs were designed. Thus, the whole panorama of marketed drugs is screened against a certain target.

Following the same reasoning, drugs that have reached preclinical and phase I stages and demonstrated to be non-toxic, were also included in HTS libraries.

Based on the possibilities that HTS platforms offered as a source for hits, a strategy for the finding of POP inhibitors was envisaged. As source for active molecules, a collection of known bioactives (in preclinical, phase I or FDA approved) was used.

As a screening technique, Fluorescence Polarization (FP) was selected, given the feasibility of its application to HTS. Since these experiments were performed for the first time, the assay was set-up. A labeled-probe with affinity for POP active-site was designed and synthesized. After optimization of the FP assay, HTS was applied for the finding of POP ligands.

The HTS was performed in the Harvard Medical School Screening Facility (ICCB Longwood), during the 4 months of stay at Prof. Gerhard Wagner laboratory.

Later, hit validation was carried out at IRB Barcelona.

3.1. Set-up Fluorescence Polarization Experiment

3.1.1. Fluorescence Polarization main concepts

Fluorescence Polarization (FP) assay takes profit of two concepts. First is related to the size of a molecule. In a solution, molecules rotate following the brownian model: small molecules tumble faster than large molecules. Second concept is related to the fluorophore nature. When a fluorophore is excited with polarized fluorescence, it emits polarized fluorescence. Taking together these two aspects, as the fluorophore turns, the amount of detected polarized fluorescence is high (if the molecule is tumbling slowly – case a figure 3.1.) or small (if the molecule is tumbling fast – case b figure 3.1.).

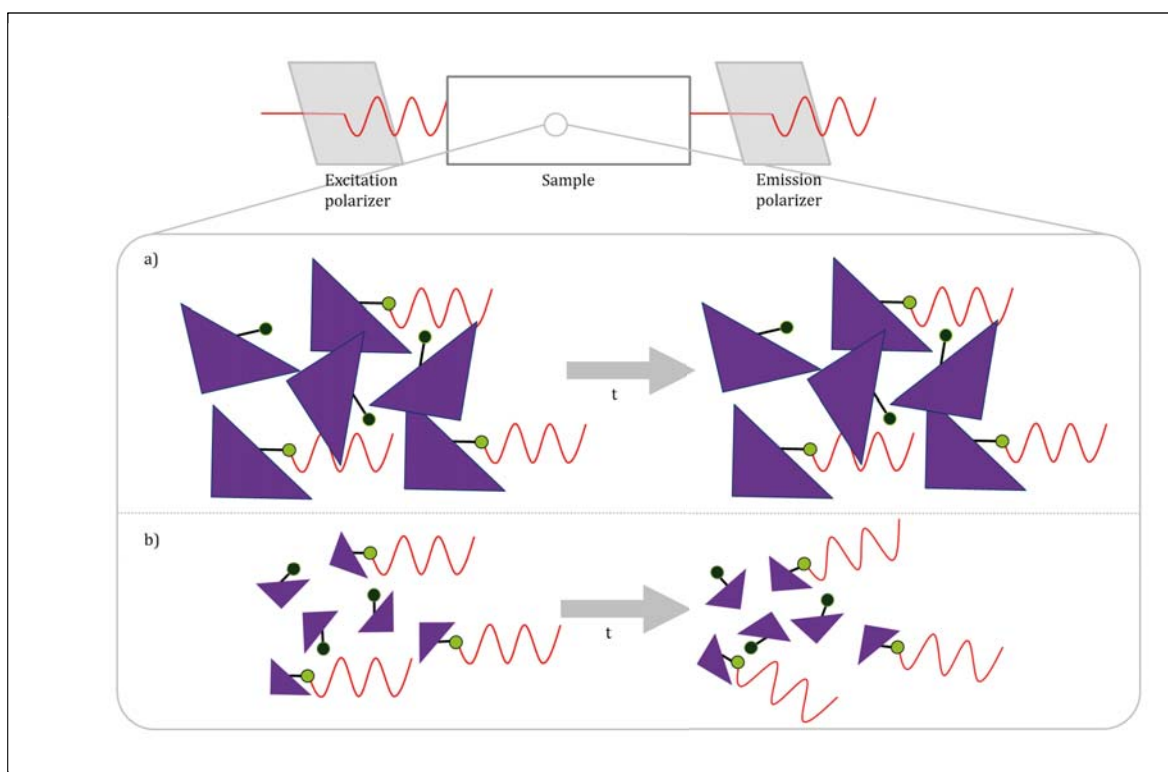


Figure 3.1: FP of a) large and b) small molecules. Large molecules tend to tumble slower than smaller ones. Once the fluorophore is excited with polarized fluorescence it emits polarized fluorescence. In the case that the molecule is tumbling faster, a percentage of the emitted fluorescence will change its plane and thus, will not be detected. On the contrary, larger molecules will have a slower movement, and a higher percentage of polarized fluorescence will be detected.

The FP assay applied to drug discovery is based on a competition between a fluorophore-labeled probe that has affinity for a certain protein hotspot and screened molecules. If the affinity of the molecule for the target is smaller than the one of the probe, the probe will remain bound to the target (case a figure 3.2.). On the contrary, if a screened molecule displays a strong affinity for the target, it will displace the probe, which will be free (case b figure 3.2.). The complex of target-probe compared to the free probe is larger in size and this difference alters the rotational correlation time. (The complex will tumble slower than

the free probe). After excitation of the fluorophore, the polarized fluorescence that will be detected will be higher for the complex than for the free probe. In summary, samples with low FP values will contain molecules that interact with the specific site of the target with high affinity.

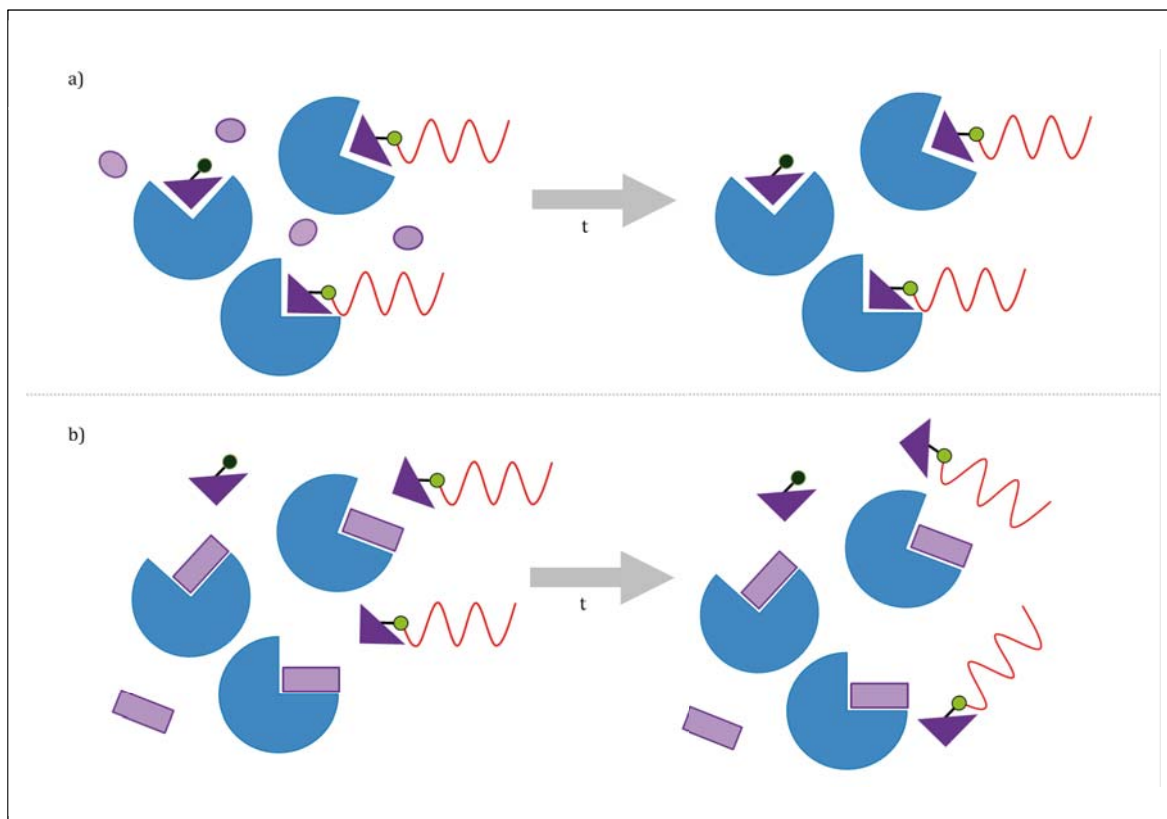


Figure 3.2.: FP applied in competition assays. A labeled probe (purple) interacts with protein target (blue). When a compound (light purple) is added to the system, it can either a) don't alter this binding, or b) displace the probe. In the case b), the labeled probe is free and its tumbling is faster than in the complex protein-probe. Thus, FP values of b) will be slower than in case a). Comparing control FP values with samples FP values allows the identification of molecules that interact with the target.

3.1.2. Probe design and synthesis

The obtaining of an optimum probe for its application in HTS is crucial.

For the design of the probe for POP, an already studied peptide was selected as scaffold. Previously in the laboratory, Eduard Sabidó designed in his thesis a series of activity-based probes for the identification of proteases.^[202] One of them (compound HTS-p) was used as a precursor for the ligand of the current FP assay (figure 3.3. a).

The peptide interacts with POP active site in a non-covalent manner. Its design was based in the structure of the canonical POP inhibitor Z-Proline Prolinal (ZPP) (figure 3.3. b). ZPP is a covalent inhibitor that interacts with the Ser of the active site through the aldehyde moiety.^[100] In the case of the probe design, a covalent bond would impossibility the displacement of the probe by competition with molecules. Thus, an acid terminus was selected in spite of an aldehyde moiety.

Besides, the IC_{50} value of the compound is 370 nM, thus it presents an optimum affinity versus POP to be used as a probe. In FP, probes with low affinity are not useful for drug discovery since they would be easily displaced even by weak protein binders. On the contrary, extremely high affinity-probes would impossibility the discovery of any ligand. Thus, equilibrium has to be reached. In this regard, the affinity of the peptide used here fulfilled the requirements for its application in HTS.

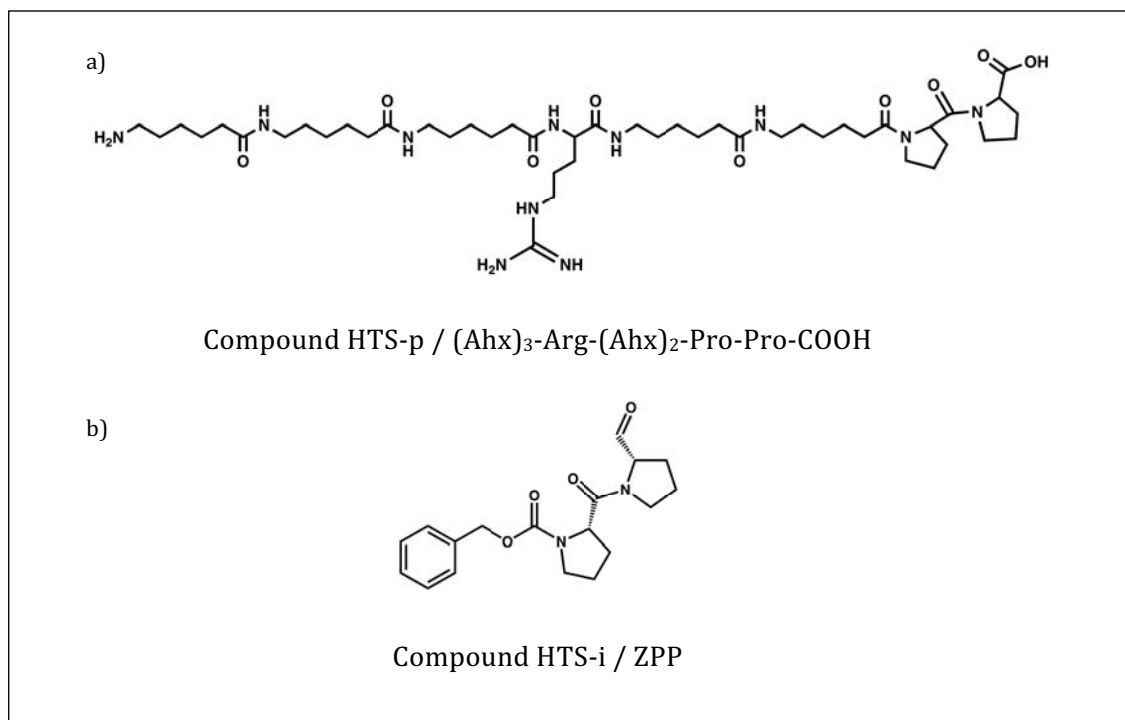


Figure 3.3.: a) Peptide used for probe design; b) ZPP, an active site POP inhibitor.

Regarding the fluorophores, in order to diminish the uncertainties of the experiment, two molecules were selected for the present study. By using two fluorophores that absorb and emit at different wavelengths, data analysis of hits can be stricter (explained in section 3.2.2. Analysis of the results)

The selected fluorophores were carboxyfluorescein (5-carboxyfluorescein) and TAMRA (5- and 6-carboxytetramethylrhodamine).

Carboxyfluorescein is one of the most common fluorophores for labeling peptides. Compared to FITC (fluorescein isothiocyanate – also broadly applied in FP), the conjugate of peptide and carboxyfluorescein is more stable and less prone to hydrolysis. Moreover, carboxyfluorescein has higher absorptivity and is water-soluble. However, it presents some photobleaching and is pH-stable only between 5 and 8. Its absorption wavelength is 494 nm and the emission is 519 nm.

In the case of TAMRA, it is as well, one of the most popular fluorophores in bioconjugations. The coupling with molecules can be easily performed by carbodiimide activation of the carboxylic acid, that can then, react with primary amines. Rhodamine-derived dyes are very photostable and are not altered by pH changes in between 4 and 10. Here, TAMRA, it is one of the most stable and fluorescent dyes of the rhodamine family.

However, it is more hydrophobic than carboxyfluorescein. Then, it is less water-soluble and aggregates at high concentrations. Its absorption wavelength is 541 nm while the emission is 565 nm.

Synthesis of the peptide was accomplished as previously described,^[202] using a Fmoc/*t*Bu strategy in solid phase peptide synthesis. In the case of the cold peptide (unlabeled probe; HTS-p), the probe was cleaved from the 2-clorotritil resin and then it was purified in a HPLC by phase-reverse (figure 3.4.).

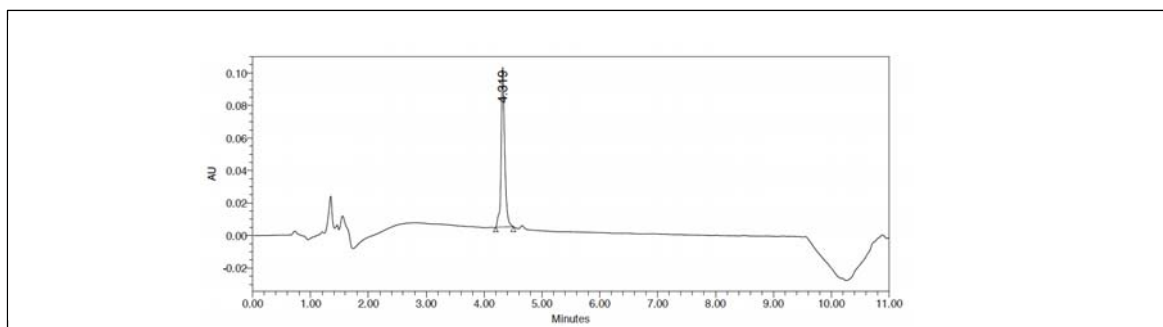


Figure 3.4.: Chromatogram corresponding to the cold peptide (HTS-p).

In the case of the probes, fluorophores were coupled to the peptide on-resin. Then the probes were cleaved and were finally purified in a HPLC by phase-reverse (figure 3.5.). For TAMRA-probe, two isomers were detected, corresponding to the 5- and 6- isomers of the fluorophore. The peaks were named TAM-I and TAM-II.

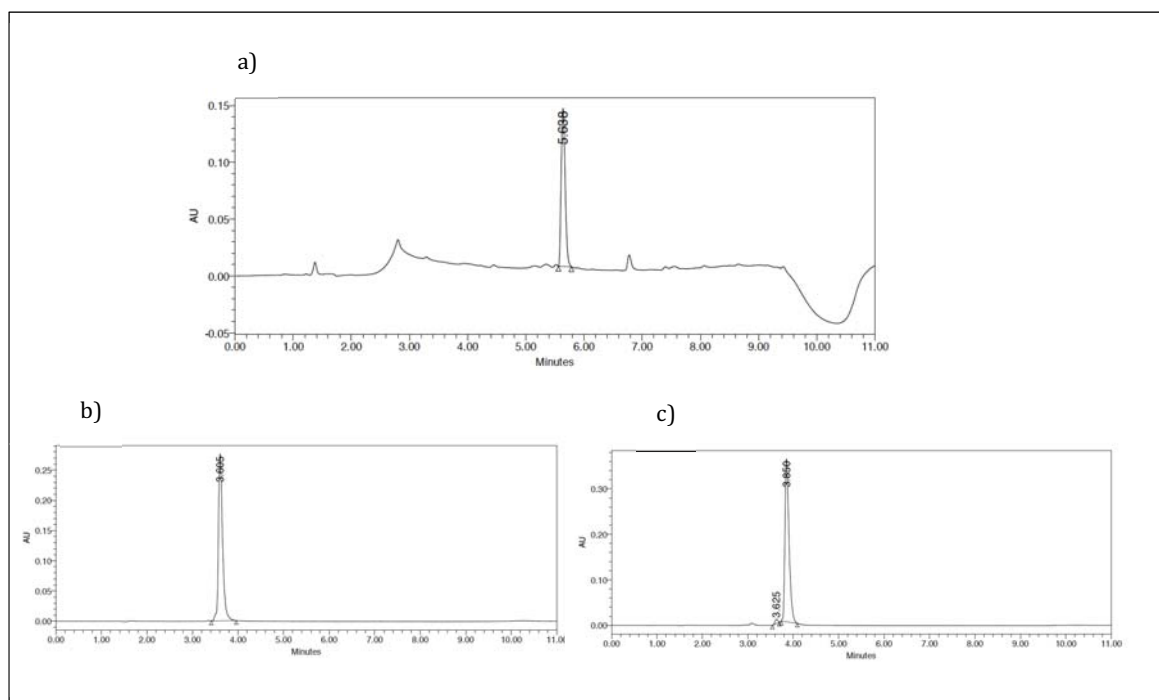


Figure 3.5.: Chromatograms corresponding to the FP probes. a) Carboxyfluorescein-probe; b) TAM-I probe; c) TAM-II probe.

3.1.3 FP assay optimization

In order to assay the possible application of the obtained probes in HTS, the labelled peptides were submitted to three evaluations.

First, the binding of the probes with POP was tested. FP values for different POP concentrations at a fixed concentration of probe were measured. The probe concentration was arbitrary chosen at 50 nM, based on the previous experience in other FP experiments with different proteins and probes performed previously at Harvard Medical School. In the case of the carboxyfluorescein probe the concentration was lowered to 30 nM, as the signal was intense, but for TAMRA probe it was not modified. For this probe, the isomer TAM-I was chosen, but both of them displayed the same behaviour.

The FP value is increased as the POP concentration increases (figure 3.6.) for carboxyfluorescein- and TAMRA- probes. This implies that the free probe is binding to the protein in a concentration dependent-manner, and as the proportion of free probe is lowering, the FP value is increasing. In addition to proving that the probes are binding to the target, this assay allowed the determination of protein concentration. POP concentration was set as the one affording around the 60% of the maximum signal (1 μ M). In this situation the system is not saturated. POP saturation with probe would avoid or difficult the identification of inhibitors.

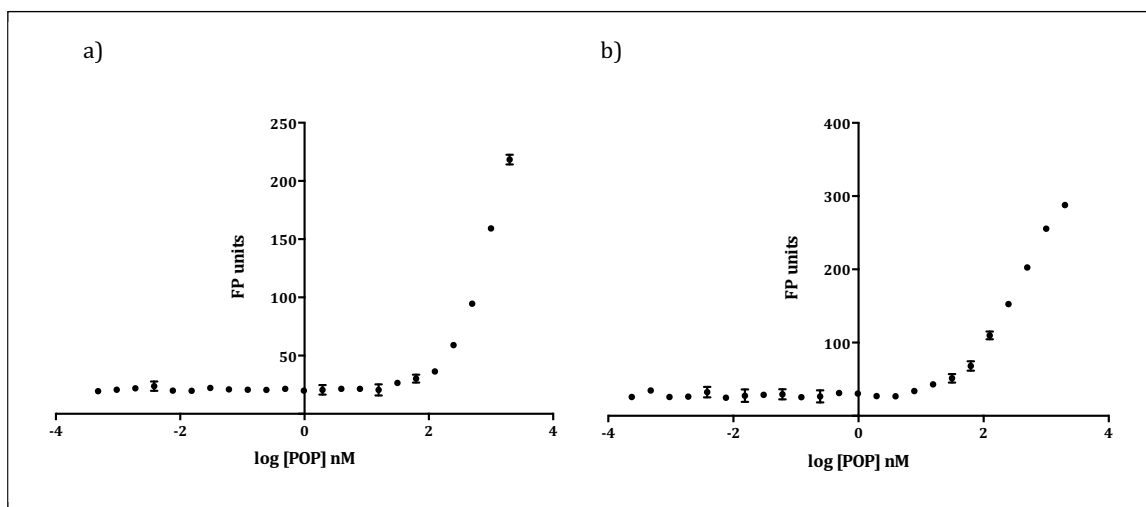


Figure 3.6.: FP assay optimization: binding of probes to POP. Determination of probe and POP concentrations. a) Carboxyfluorescein labeled probe (30 nM); b) TAMRA labeled probe (50 nM).

Secondly, an evaluation of the confidence of the assay was performed. The FP value of the probe was measured and compared to the FP value of the probe in complex with the selected POP concentration (figure 3.7.). The difference between the two values had to be high enough to diminish the appearance of false positives. If the difference between the FP of the probe/protein complex and the FP of the probe is too small, detection of hits would be almost impossible. The parameter used for the confidence evaluation was the z-score. It is a value between 0 and 1 that correlates the FP values and the standard deviation of the measurements. A z-score closer to 1 implies large difference between FP values and small SD. If the z-score is greater than 0.5 the probe is accepted to be used in the HTS.

The measured z-score for the carboxyfluorescein-probe was 0.80, demonstrating the robustness of the assay.

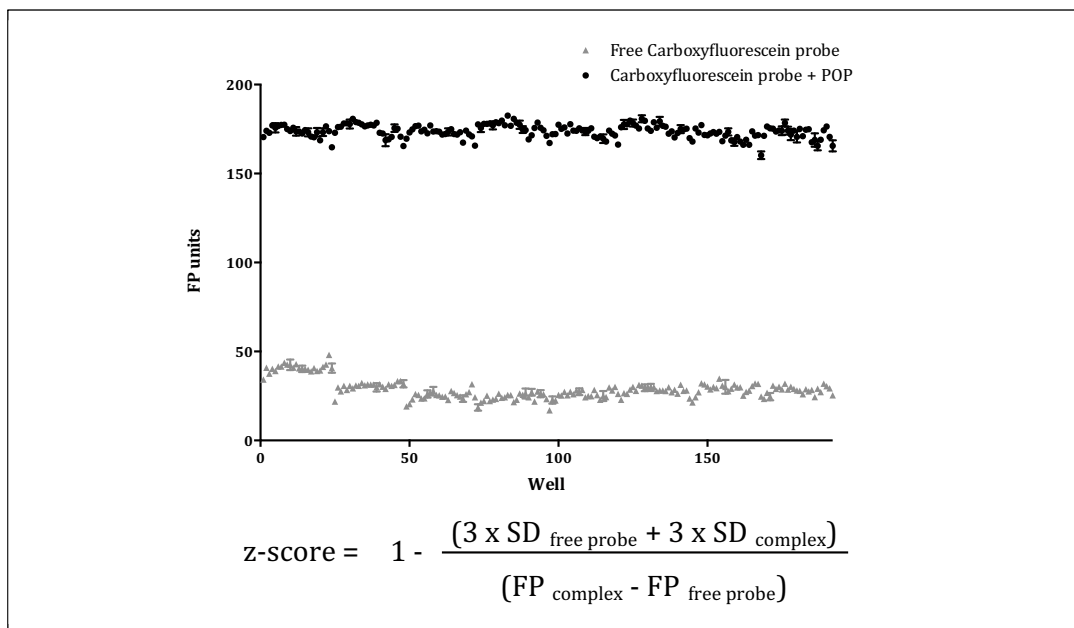


Figure 3.7.:FP assay optimization: confidence of the assay. Determination of z-score value for Carboxyfluorescein labeled probe.

Third evaluation consisted in the test of the displacement capacity that the probes had. In HTS molecules compete with the probes for the protein-binding site. Thus, probes should be able to be displaced. Competition of the probes with cold-probe (non-labeled; HTS-p) and an active site POP inhibitor (*Z*-Pro-Prolinal; ZPP) was evaluated. POP concentration was 1 μ M, carboxyfluorescein-probe was at 30 nM and TAMRA-probe at 50 nM. For the cold-probe and ZPP, a range of concentrations were tested.

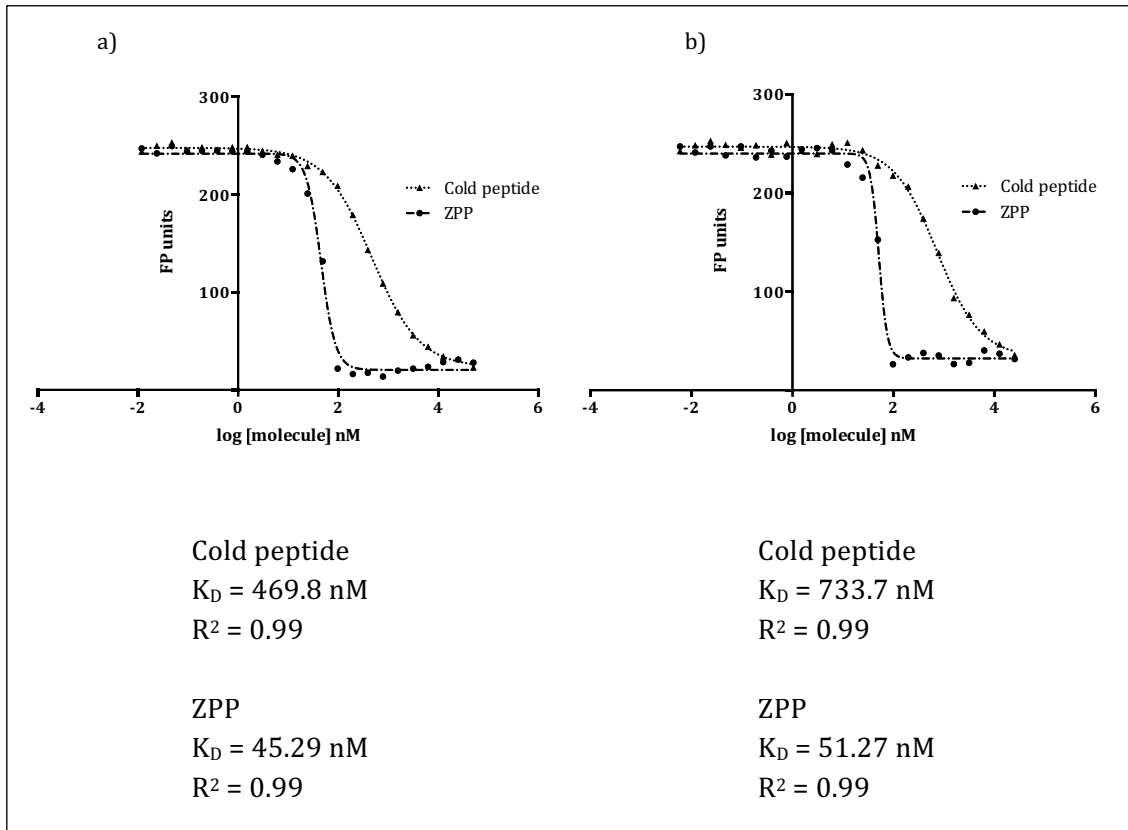


Figure 3.8.: Competition of cold peptide and ZPP with the labeled probe for the active site of POP. a) Carboxyfluorescein probe; b) TAMRA probe.

From this experiment it was concluded that the probes could be displaced by POP ligands.

The affinity values for the cold-peptide and ZPP were slightly different for the labeled-probes. This effect is a result from the distinct affinities of the probes, as consequence of the fluorophore nature. Carboxyfluorescein-probe has more affinity for POP than TAMRA-probe (though not very much). For that reason, concentrations of the first probe is smaller, but in both cases 60% of POP in the probe-bound state.

However, differences on concentrations has an effect in competition. Once the screened molecule is added to the system, it has to compete with the probe. Despite of the affinities, the more concentration of ligand, the less accessible is the active site for the inhibitor. By probability, an inhibitor molecule surrounded by 5 molecules of probe is less prone to reach POP active site than if only 3 molecules of probe were surrounding it. Thus, the equilibrium of the inhibitor-probe-protein system is reached at different time scales for each probe. In the case of TAMRA probe, it could be that the required time to reach the equilibrium was larger than the time between plate preparation and reading. Large incubation time would demonstrate this hypothesis.

3.2. High-Throughput Screening

3.2.1. HTS Experiment

As previously noted, the screened compounds belonged to a collection of known bioactive libraries (in preclinical, phase I or FDA approved). Six libraries were selected, which accounted for a total of 4811 compounds.

Biomol 4 is composed by FDA approved molecules, and thus, are marketed drugs.

NIH Clinical collections 1 and 2 and MicroSource 1-US libraries included compounds that reached clinical phases in the US and are drug-like with known safety profiles.

Sigma LOPAC (Library of pharmaceutically active compounds) includes marketed drugs and pharmaceutically relevant molecules with known biological activity.

Finally, Tocriscreen Mini library is composed by biologically active non-toxic compounds.

Table 3.2. contains the the number of compounds and stock concentration of each library.

| Library | Number of compounds | Stock concentration |
|------------------------------------|----------------------------|----------------------------|
| Biomol 4-FDA Approved Drug Library | 640 | 2 mg/mL |
| NIH Clinical collection 1 | 450 | 10 mM |
| NIH Clinical collection 2 | 281 | 10 mM |
| Microsource 1-US Drug Collection | 1040 | 2 mg/mL |
| Sigma LOPAC | 1280 | 10 mM |
| Tocriscreen Mini Library | 1120 | 10 mM |

Table 3.2.: Screened libraries in the HTS

The total of the six libraries accounted for 17 plates (384 well-plates). Thus, since two probes were being used, the experiment was doubled: 17 plates of POP and carboxyfluorescein-probe and other 17 of POP and TAMRA-probe.

First, a mixture of POP and probe was prepared and added to the plates.

After, the 100 nL pin transfer of compounds was done. Final concentrations of compounds depended upon stock concentration and were 33 μ M (if 10 mM stock) or 6.67 μ g/mL (if 2 mg/mL) stock. To have an approximate comparison, if 350 g/mol was considered as a good standard for small molecule weight, then 6.67 μ g/mL would equal to 19 μ M.

Finally, plates were read for FP and TF (total fluorescence) twice at the appropriate wavelengths.

As a result, 8 values were obtained for each compound:

| | |
|------------------------------------|-----------------------|
| FP carboxyfluorescein-probe read 1 | FP TAMRA-probe read 1 |
| FP carboxyfluorescein-probe read 2 | FP TAMRA-probe read 2 |
| TF carboxyfluorescein-probe read 1 | TF TAMRA-probe read 1 |
| TF carboxyfluorescein-probe read 2 | TF TAMRA-probe read 2 |

3.2.2. Analysis of the results

From the HTS, a list of FP and TF values were obtained. To evaluate the activity of the compounds as POP binders, the data has to be first transformed into meaningful parameters. Afterwards, the working information is filtered through three restrictions to finally obtain the real hits.

3.2.2.1. Calculations

In order to transform the data obtained in the plate reading into working data, z-scores of FP and TF were calculated. In this case, z-score formulas are different from the one used in the assay optimization. In here, z-score is a comparison of a given sample value respect a reference value and measured in terms of SD (Standard Deviation).

$$\text{z-score} = \left| \frac{\text{FP}_{\text{sample}} - \text{Average FP}_{\text{control}}}{\text{SD Average FP}_{\text{control}}} \right| \quad \text{z-score} = \left| \frac{\text{TF}_{\text{sample}} - \text{Average TF}_{\text{all}}}{\text{SD Average TF}_{\text{all}}} \right|$$

First, FP values of control wells (wells without compound) were selected and the average of FP as well as its standard deviation was calculated. Empty wells were distributed all over the plate to avoid equipment-reading alterations. Then, the z-score of FP for each well was calculated. Taking into account that each well was measured twice for each fluorophore, a total of 4 z-score values of FP per compound were obtained.

Second, TF data was transformed in a similar way as for FP. However, in this case, the reference value was not the empty wells but the total wells of the plate. In this way, an average fluorescence value of compounds was used. Analogous to FP, 4 z-score values were obtained for each compound.

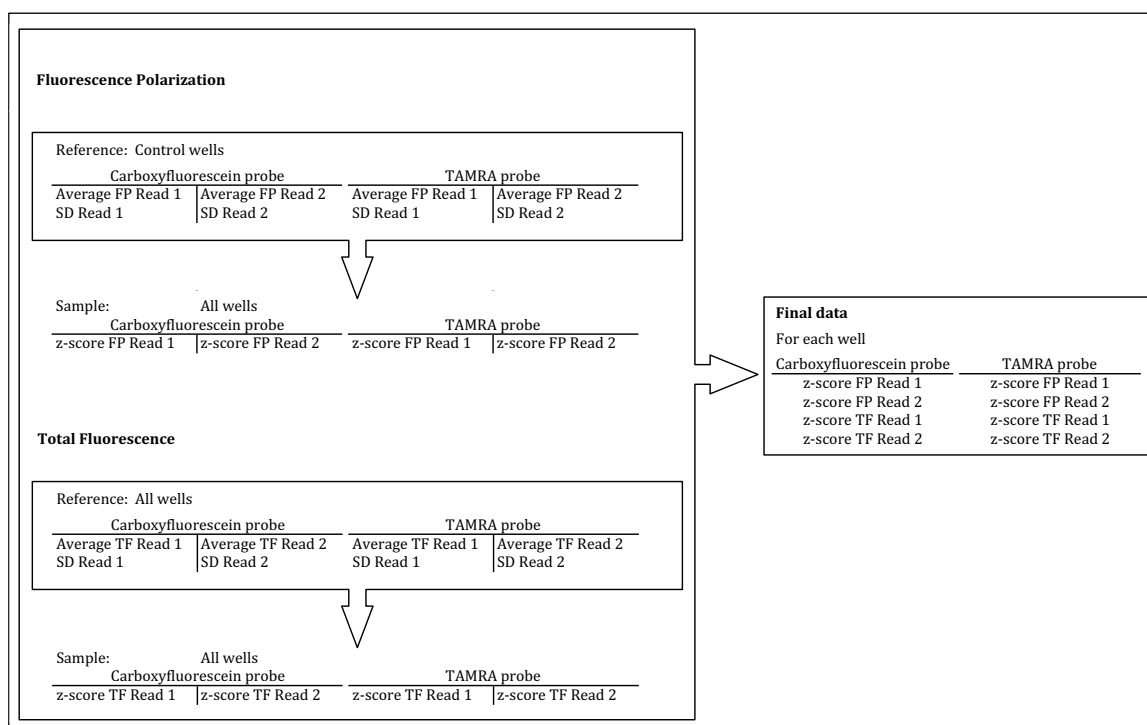


Figure 3.9.: Schematic representation of calculations in the HTS.

3.2.2.2. Evaluation

Next step was to evaluate the z-score data. This implied the application of three restrictions.

First restriction consisted in defining what was a positive result. Taking into account the z-score of FP, a well position was considered positive if the z-score of FP was higher than 3.99. This implied that the difference between the sample FP value and the average of the FP values was higher than 4 times the SD. Restriction had to be fulfilled for the 4 z-scores of FP (2 readings, 2 fluorophores).

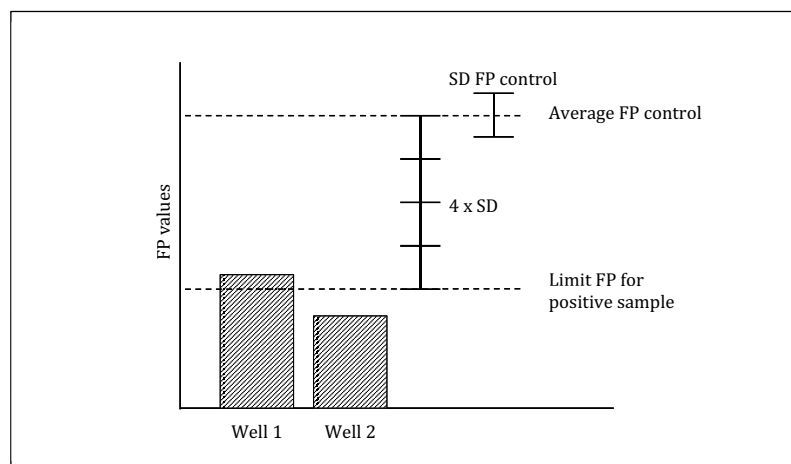


Figure 3.10.: First restriction in the evaluation of HTS results. Only samples with a difference 4 times or higher the control SD between the FP sample and the FP control were considered positives. Higher z-score implied more difference. In this figure, well 1 contains a compound considered as negative in the competition assay, despite its ability to lower the FP compared to the average of the FP from the controls. On the contrary, well 2 contains a compound that is considered as positive.

Second restriction removed samples with intrinsic fluorescence or autofluorescence.

Screened molecules could have auto-fluorescence that will interfere with the measurement of FP. By definition, FP is the difference of the emission fluorescence intensity between the parallel and the perpendicular excitation light plane, respect to the total fluorescence emission intensity. Thus, auto-fluorescence molecules increase the fluorescence, reducing the FP.

A well position was considered negative in autofluorescence if z-score of TF was lower than 2.99. By this, the fluorescence value was not far enough from the average. Higher differences may lead to false positives.

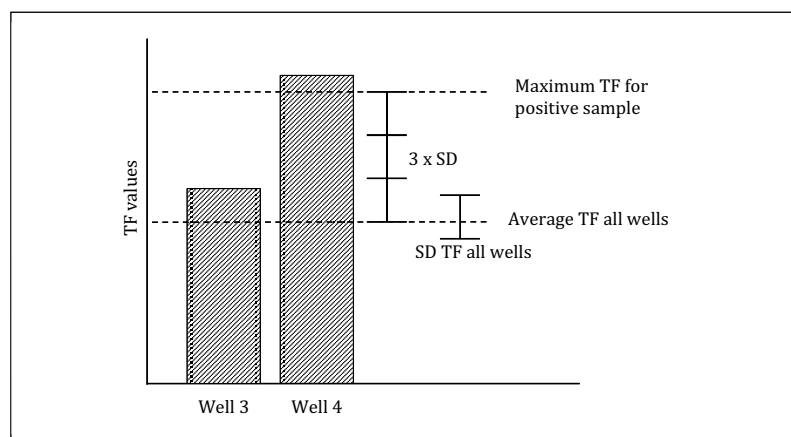


Figure 3.11.: Second restriction in the evaluation of HTS results. Only samples with a difference smaller than 3 times the control SD between the sample TF and the control TF were considered positives. Lower z-score implied smaller differences. In this figure, well 3 contains a compound considered as positive in the competition assay, because its TF value is close to the average. On the contrary, well 4 contains a compound that is considered as negative because the TF value is extremely different from the average. This implies intrinsic fluorescence of the compound, which interferes with the FP measurement.

Third and final restriction removed false positives due to aggregation. In first restriction, positive results were defined as those with a z-score 4 or higher. However, z-score is an absolute number, and thus, samples with 4 times lower FP and 4 times higher FP compared to the control were selected. If the FP value was higher than the control FP, it implied that the complex of protein/probe/molecule had a significantly lesser rotational time than the protein/probe complex, or in other words, it was a larger complex. Then, the compound was aggregating. In order to remove false positives, the value of FP was considered (instead of z-score). Only wells where FP value was lower than the average were defined as real positives.

In the present study, all molecules that passed filters 1 and 2 automatically passed filter 3, meaning that no aggregator has been detected as a possible hit.

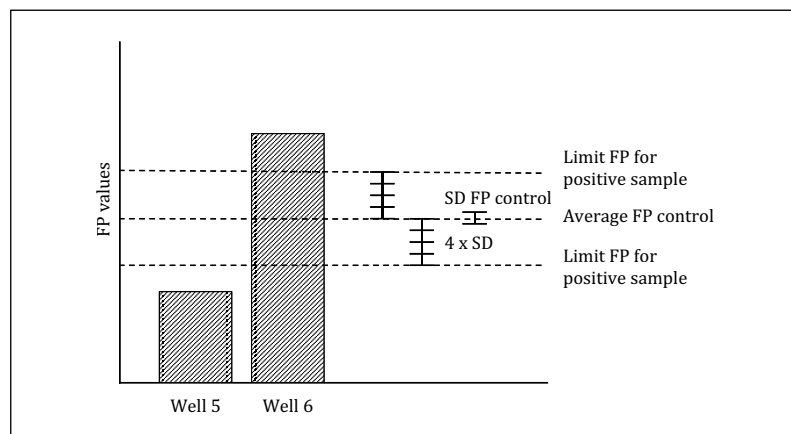


Figure 3.12.: Third restriction in the evaluation of HTS results. Only samples with an absolute value of FP smaller than the average FP of the control wells are accepted as positive samples. In this figure wells 5 and 6 have passed restriction 1, since their corresponding z-scores are higher than 4 times the SD of the average of FP from controls. However, since z-score is an absolute parameter, it does not take into account whether this difference is because the FP value of the sample is much more smaller (well 5) or higher (well 6). Higher FP values mean that the complex form is tumbling slowly than the control. This effect may be a consequence of an aggregation factor, which is undesirable. In this example, well 6 will be driven out from the positives list.

As a summary, in first restriction, positive samples were defined as those with a FP value $4xSD$ different from the control FP. Second filter removed false positives by autofluorescence. Only samples where the TF was less than $3xSD$ different from the control TF were accepted. Third restriction removed false positives by aggregation. Molecules with a real FP value inferior to the control FP were considered positives.

For a compound to be considered as positive in the HTS, it had to overcome the three restrictions, for the 2 fluorophore-labeled probes and the two readings separately. If one parameter was not in the range, the compound was discarded.

It means that if a molecule represented a hit for one probe, but not for the other, it was not accepted. Even if in the second probe one of the readings was in the range while the other was not, the molecule was rejected.

3.2.2.3. List of FP assay results

A total of 84 hits were obtained in the HTS FP assay. Careful examination of compounds led to the observation of molecule redundancy in different plates. Thus, the number of hits was finally reduced to 73. However, the fact that the same molecule was spotted as a POP interacting molecule in different plates reinforced the statement of the molecule as a true positive. Repeated hits were the following:

| Id | Equal to | Id | Equal to | Id | Equal to |
|--------|----------|--------|----------|--------|----------|
| HTS-8 | HTS-3 | HTS-52 | HTS-34 | HTS-62 | HTS-32 |
| HTS-10 | HTS-6 | HTS-54 | HTS-49 | HTS-64 | HTS-41 |
| HTS-14 | HTS-5 | HTS-56 | HTS-21 | HTS-67 | HTS-23 |
| HTS-51 | HTS-42 | HTS-58 | HTS-33 | | |

Table 3.3.: Repeated hits that were discarded and their corresponding equal hit Id.

The list of the 73 hits is listed in table 3.4. For single-repeated molecules, the final z-score is the average of the two readings, separately for each fluorophore-labeled probe. In the case of repeated molecules, the z-score was the average of the four readings, separately for each fluorophore-labeled probe. In those cases, differences between z-scores were on the same range. Data is represented in a bar diagram (figure 3.13.)

| Id | z-score | |
|--------|---------------------------|-------------|
| | Carboxy-fluorescein probe | TAMRA probe |
| HTS-1 | 19.06 | 8.38 |
| HTS-2 | 11.88 | 4.31 |
| HTS-3 | 11.14 | 6.57 |
| HTS-4 | 16.68 | 10.60 |
| HTS-5 | 17.81 | 6.32 |
| HTS-6 | 12.67 | 5.46 |
| HTS-7 | 16.95 | 9.17 |
| HTS-9 | 18.80 | 12.40 |
| HTS-11 | 5.74 | 7.23 |
| HTS-12 | 5.34 | 5.47 |
| HTS-13 | 14.73 | 4.35 |
| HTS-15 | 11.93 | 19.16 |
| HTS-16 | 12.78 | 6.61 |
| HTS-17 | 15.83 | 9.81 |
| HTS-18 | 8.74 | 5.75 |
| HTS-19 | 8.67 | 4.11 |
| HTS-20 | 6.38 | 4.92 |
| HTS-21 | 10.90 | 6.96 |
| HTS-22 | 27.30 | 17.93 |
| HTS-23 | 9.11 | 8.10 |
| HTS-24 | 10.38 | 7.69 |
| HTS-25 | 5.47 | 6.35 |
| HTS-26 | 5.70 | 4.71 |
| HTS-27 | 7.44 | 6.35 |
| HTS-28 | 14.81 | 9.89 |
| HTS-29 | 6.99 | 7.26 |
| HTS-30 | 7.66 | 7.95 |
| HTS-31 | 4.97 | 4.21 |
| HTS-32 | 5.23 | 7.85 |
| HTS-33 | 7.94 | 5.18 |
| HTS-34 | 11.13 | 8.12 |
| HTS-35 | 9.57 | 10.52 |
| HTS-36 | 4.67 | 4.24 |
| HTS-37 | 4.80 | 9.73 |

(continued in next page)

| Id | z-score | |
|--------|---------------------------|-------------|
| | Carboxy-fluorescein probe | TAMRA probe |
| HTS-38 | 23.43 | 13.52 |
| HTS-39 | 7.22 | 4.81 |
| HTS-40 | 20.42 | 13.18 |
| HTS-41 | 6.39 | 5.07 |
| HTS-42 | 8.99 | 5.30 |
| HTS-43 | 16.70 | 7.72 |
| HTS-44 | 6.81 | 4.64 |
| HTS-45 | 6.20 | 4.69 |
| HTS-46 | 12.86 | 6.34 |
| HTS-47 | 11.42 | 5.76 |
| HTS-48 | 10.33 | 5.35 |
| HTS-49 | 9.53 | 5.96 |
| HTS-50 | 4.89 | 9.09 |
| HTS-53 | 8.16 | 8.10 |
| HTS-55 | 7.64 | 5.09 |
| HTS-57 | 5.23 | 23.06 |
| HTS-59 | 11.95 | 15.21 |
| HTS-60 | 6.43 | 9.15 |
| HTS-61 | 4.71 | 4.40 |
| HTS-63 | 5.86 | 6.38 |
| HTS-65 | 20.77 | 11.29 |
| HTS-66 | 11.67 | 6.78 |
| HTS-68 | 21.94 | 15.60 |
| HTS-69 | 9.71 | 7.62 |
| HTS-70 | 10.18 | 6.66 |
| HTS-71 | 11.57 | 7.10 |
| HTS-72 | 8.41 | 7.20 |
| HTS-73 | 24.13 | 14.45 |
| HTS-74 | 15.81 | 5.82 |
| HTS-75 | 5.93 | 6.93 |
| HTS-76 | 6.58 | 4.08 |
| HTS-77 | 22.88 | 12.28 |
| HTS-78 | 8.57 | 4.35 |
| HTS-79 | 4.98 | 13.40 |
| HTS-80 | 9.31 | 5.06 |
| HTS-81 | 13.15 | 5.72 |
| HTS-82 | 12.05 | 5.41 |
| HTS-83 | 8.07 | 5.68 |
| HTS-84 | 10.47 | 6.99 |

Table 3.4.: HTS hits for POP active-site binding obtained by FP assay using Carboxyfluorescein and TAMRA-labeled probes. Higher z-scores relate to better displacement efficacy of the probe by the molecule.

Molecules with highest z-score when using carboxyfluorescein probe were HTS-22 and HTS-73. These compounds were also spotted in the higher positions of the z-score ranking when using TAMRA probe (HTS-22 had number 3 and HTS-73 was number 6).

Regarding the z-score when using TAMRA probe, HTS-57 and HTS-15 presented the highest z-score. However, these molecules did not have a high z-score for the other probe (HTS-57 was number 66 in the carboxyfluorescein probe ranking and HTS-15 was number 24). In general, a 57% of the molecules were present in both top 30 rankings, using carboxyfluorescein and TAMRA probe. These discrepancies pointed out that the assay of FP, with a single concentration of compound, may not be suitable in order to rank molecules in basis of their potency, but to identify compounds that interact with the target displacing the probe. Further experiments with complementary tools would allow the potency ranking.

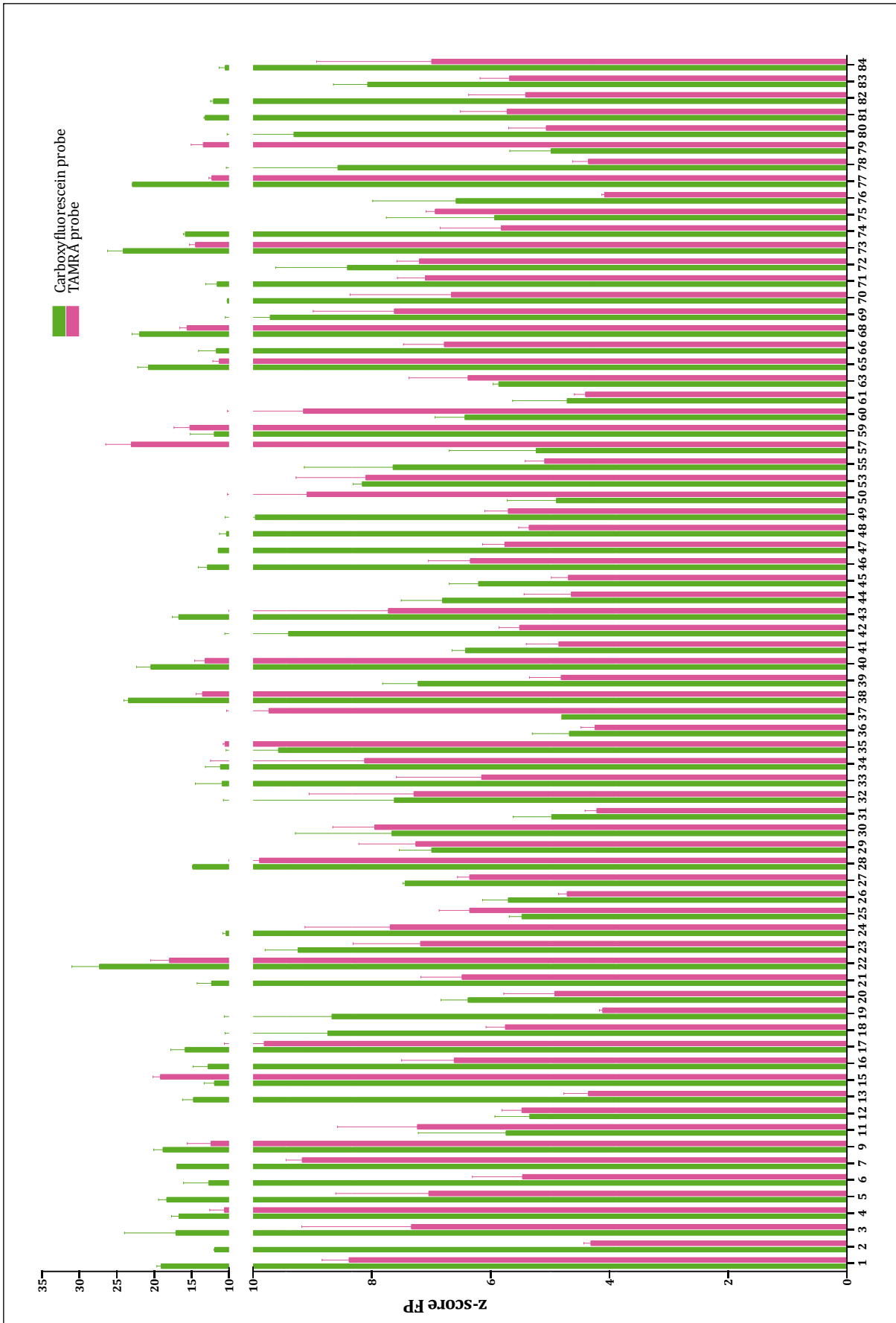


Figure 3.13.: z-score of HTS hits. X-axis shows Id number HTS-xx.

3.3. Hit validation

In order to fully consider a molecule as a POP inhibitor, hits in the FP assay were subjected to evaluation by a second technique. Due to the well-established enzymatic assay of POP, hits were planned to be tested against POP proteolytic activity.

From the almost 5,000 molecules screened against POP in the HTS, 73 were selected as hits after the FP competition assay. Despite that the number has been reduced by a factor around 70, the final list was still large to be validated by enzymatic assays. Thus, a strategy was planned in order to reduce the number of compounds. First, by an *in silico* clustering molecules were classified into groups that enclosed similar compounds in terms of chemical structure. Then, one to three representatives of each cluster were selected for testing against POP activity. Selection of molecules was based on sequence similarity inside a cluster and results from docking analysis. Best molecules in the preliminary POP activity test were characterized as inhibitors.

3.3.1. Clustering

Clustering was performed by Dr. Martin Ivanov. In the subset of 73 hits, two compounds were found to be used in *in silico* calculations due to the presence of metal atoms Pt and Au. Thus, the *in silico* library was reduced to 71 molecules. However, two salts were found to be repeated in several hits. Thus, they were included in the clustering, forming a total of 73 compounds. The Id numbers that were given to them were HTS-85 and HTS-86. The 75 chemical structures were analysed by Canvas, a cheminformatics package of Schrödinger that provides a range of applications like scaffold decomposition, maximum common substructure and R-group analysis.

The binary fingerprints of the molecules (a combination of 0 and 1) were obtained. After, they were compared using the Tanimoto coefficient calculations. As a result, molecules were represented in hierarchical clustering (family tree).

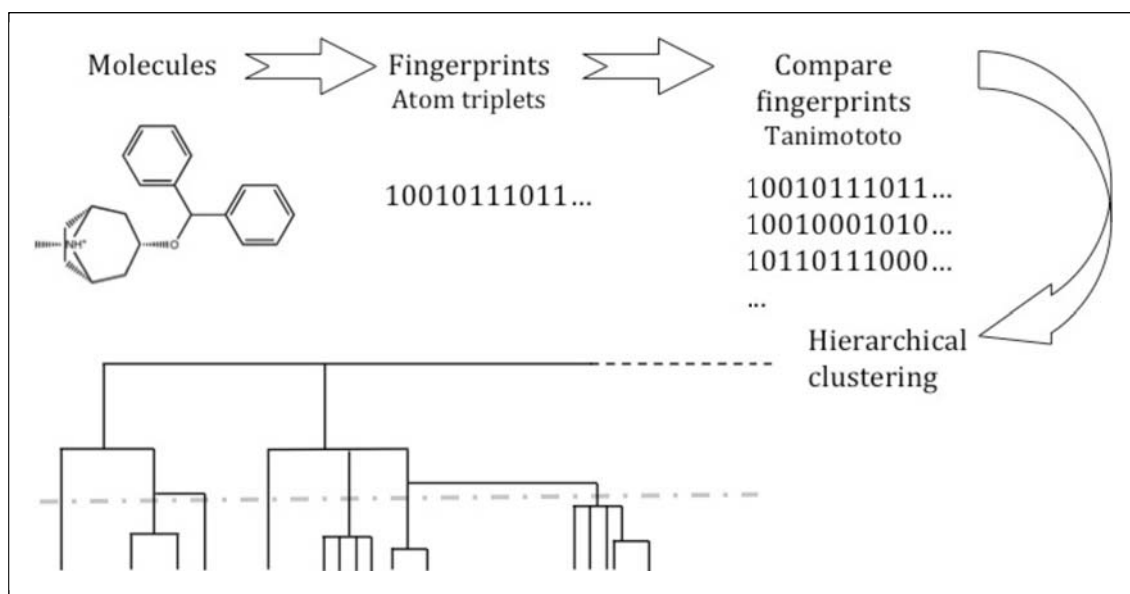


Figure 3.14.: Scheme of clustering performed using Canvas (Schrödinger).

From the clustering calculations, 15 clusters were obtained. Cluster 11 was the largest one with 18 molecules. Clusters 8 and 10 were large as well, with 10 and 9 compounds, respectively. The rest of the clusters were represented with 2-5 molecules (except cluster 7 with just one member).

To the list of 15 clusters, cluster 0, which included the two molecules that were unable to be used in calculations, was added. In total, 75 molecules (73 hits + 2 salts) were clustered into 16 families, as seen in table 3.5.

| Cluster | Id | Cluster | Id | Cluster | Id |
|---------|----|---------|----|---------|----|
| 0 | 26 | 8 | 9 | 11 | 12 |
| 0 | 38 | 8 | 18 | 11 | 17 |
| 1 | 11 | 8 | 23 | 11 | 19 |
| 1 | 27 | 8 | 37 | 11 | 21 |
| 1 | 55 | 8 | 39 | 11 | 24 |
| 1 | 60 | 8 | 59 | 11 | 30 |
| 2 | 33 | 8 | 75 | 11 | 35 |
| 2 | 34 | 8 | 78 | 11 | 36 |
| 3 | 47 | 8 | 85 | 11 | 41 |
| 3 | 2 | 8 | 86 | 11 | 45 |
| 3 | 31 | 9 | 3 | 11 | 48 |
| 4 | 6 | 9 | 5 | 11 | 50 |
| 4 | 7 | 9 | 13 | 11 | 66 |
| 4 | 16 | 9 | 15 | 11 | 68 |
| 5 | 40 | 9 | 63 | 11 | 72 |
| 5 | 42 | 10 | 20 | 11 | 73 |
| 5 | 43 | 10 | 22 | 11 | 77 |
| 5 | 46 | 10 | 29 | 11 | 80 |
| 5 | 49 | 10 | 44 | 12 | 53 |
| 6 | 25 | 10 | 65 | 13 | 28 |
| 6 | 69 | 10 | 70 | 13 | 82 |
| 6 | 84 | 10 | 71 | 13 | 83 |
| 7 | 79 | 10 | 74 | 14 | 32 |
| | | 10 | 76 | 14 | 57 |
| | | | | 15 | 1 |
| | | | | 15 | 4 |
| | | | | 15 | 61 |
| | | | | 15 | 81 |

Table 3.5.: 75 molecules clustered into 16 families.

Taking into account the FP results, molecules were ranked in basis of z-score and their corresponding cluster. The top 30 compounds for each probe is described in tables.

| Cluster | Id | z-score Carboxy-fluorescein peptide | Present in TAMRA top 30? |
|---------|----|---|-----------------------------|
| 10 | 22 | 27.30 | |
| 11 | 73 | 24.13 | |
| 0 | 38 | 23.43 | |
| 11 | 77 | 22.88 | |
| 11 | 68 | 21.94 | |
| 10 | 65 | 20.77 | |
| 5 | 40 | 20.42 | |
| 15 | 1 | 19.06 | |
| 8 | 9 | 18.80 | |
| 9 | 5 | 17.81 | No |
| 4 | 7 | 16.95 | |
| 5 | 43 | 16.70 | |
| 15 | 4 | 16.68 | |
| 11 | 17 | 15.83 | |
| 10 | 74 | 15.81 | No |
| 13 | 28 | 14.81 | No |
| 9 | 13 | 14.73 | No |
| 15 | 81 | 13.15 | No |
| 5 | 46 | 12.86 | No |
| 4 | 16 | 12.78 | |
| 4 | 6 | 12.67 | No |
| 13 | 82 | 12.05 | No |
| 8 | 59 | 11.95 | |
| 9 | 15 | 11.93 | |
| 3 | 2 | 11.88 | No |
| 11 | 66 | 11.67 | No |
| 10 | 71 | 11.57 | No |
| 3 | 47 | 11.42 | No |
| 9 | 3 | 11.14 | No |
| 2 | 34 | 11.13 | |

Table 3.6.: HTS hits ordered by higher z-score with the carboxyfluorescein probe. Only the 30 best molecules are depicted.

| Cluster | Id | z-score TAMRA peptide | Present in Carboxy top 30? |
|---------|----|-----------------------------|-------------------------------|
| 14 | 57 | 23.06 | No |
| 9 | 15 | 19.16 | |
| 10 | 22 | 17.93 | |
| 11 | 68 | 15.60 | |
| 8 | 59 | 15.21 | |
| 11 | 73 | 14.45 | |
| 0 | 38 | 13.52 | |
| 7 | 79 | 13.40 | No |
| 5 | 40 | 13.18 | |
| 8 | 9 | 12.40 | |
| 11 | 77 | 12.28 | |
| 10 | 65 | 11.29 | |
| 15 | 4 | 10.60 | |
| 11 | 35 | 10.52 | No |
| 13 | 28 | 9.89 | |
| 11 | 17 | 9.81 | |
| 8 | 37 | 9.73 | No |
| 4 | 7 | 9.17 | |
| 1 | 60 | 9.15 | |
| 11 | 50 | 9.09 | No |
| 15 | 1 | 8.38 | |
| 2 | 34 | 8.12 | |
| 12 | 53 | 8.10 | No |
| 8 | 23 | 8.10 | No |
| 11 | 30 | 7.95 | No |
| 14 | 32 | 7.85 | No |
| 5 | 43 | 7.72 | |
| 11 | 24 | 7.69 | No |
| 6 | 69 | 7.62 | No |
| 10 | 29 | 7.26 | No |

Table 3.7.: Thirty HTS hits with the highest z-score using the TAMRA probe.

Focusing on the tables 3.6. and 3.7., the proportion of molecules in the top 30 z-score for each cluster was analysed. The higher the percentage, it means that the cluster is a “hot-group” for POP inhibitors.

Regarding carboxyfluorescein probe, clusters with higher percentage of molecules in the top 30 z-score were 10 and 11, while cluster 11 was the one with more molecules in the top 30 when TAMRA probe was used.

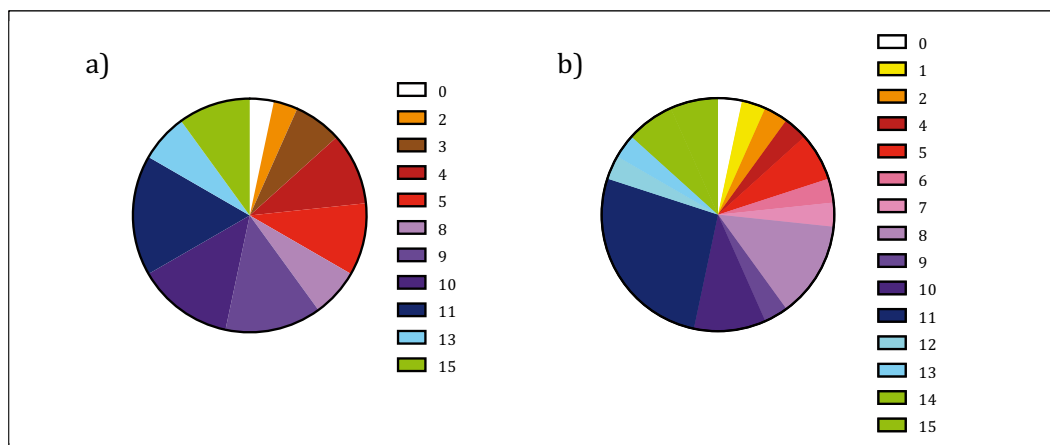


Figure 3.15.: Relative proportion of clusters depending on the number of molecules they had in the top 30 ranking of docking analysis. Percentages refer to the proportion of molecules from each cluster respect the total top 30. a) Carboxyfluorescein probe; b) TAMRA probe.

Another analysis that was done was the enrichment of each cluster in top 30 molecules. Depending on the absolute number of compounds/cluster and the proportion of its molecules that reached the top 30, a cluster was considered, or not, as “hot-cluster” for POP inhibitors.

For carboxyfluorescein probe, clusters 4 and 15 were the most enriched in positive molecules in FP. On the contrary, for TAMRA probe, cluster 14 showed the highest proportion of POP interactors. (Clusters 7 and 12, with only one molecule were fully represented as well).

| Cluster | Molecules | Number of hits/probe in the top 30 ranking | | | |
|---------|-----------|--|------------|--------------|------------|
| | | Absolute numbers | | Percentages | |
| | | FP - Carboxy | FP - TAMRA | FP - Carboxy | FP - TAMRA |
| 0 | 2 | 1 | 1 | 50 | 50 |
| 1 | 4 | 0 | 1 | 0 | 25 |
| 2 | 2 | 1 | 1 | 50 | 50 |
| 3 | 3 | 2 | 0 | 67 | 0 |
| 4 | 3 | 3 | 1 | 100 | 33 |
| 5 | 5 | 3 | 2 | 60 | 40 |
| 6 | 3 | 0 | 1 | 0 | 33 |
| 7 | 1 | 0 | 1 | 0 | 100 |
| 8 | 10 | 2 | 4 | 20 | 40 |
| 9 | 5 | 4 | 1 | 80 | 20 |
| 10 | 9 | 4 | 3 | 44 | 33 |
| 11 | 18 | 5 | 8 | 28 | 44 |
| 12 | 1 | 0 | 1 | 0 | 100 |
| 13 | 3 | 2 | 1 | 67 | 33 |
| 14 | 2 | 0 | 2 | 0 | 100 |
| 15 | 4 | 3 | 2 | 75 | 50 |

Table 3.8.: Number of molecules ranked in the best 30 positions for FP assay using carboxyfluorescein or TAMRA probe versus the cluster to which they belong. Percentages refer to the proportion of positives/cluster.

3.3.2. Docking

A subset of 73 molecules (71 hits + 2 salts) were docked into POP active site.

Respect to the protein, four crystallographic structures were used: 1QFS, 2XDW, 1QFM and 3DDU. The first two correspond to porcine POP that was co-crystallized with covalent inhibitor, the third one was porcine as well but without inhibitor, and the last one belonged to the human POP in complex with a non-covalent inhibitor.

For ligands, their conformers were generated with ConfGen (Schrödinger), obtaining a total of 1238 conformations in total. Each one of them was docked to the four POP structures with Glide XP precision (Schrödinger), obtaining best result when the human protein was used.

For each compound, the score result for the best conformation is depicted in table 3.9.

| Cluster | Id | GlideXP docking score | Cluster | Id | GlideXP docking score |
|---------|----|-----------------------|---------|----|-----------------------|
| 15 | 1 | -11.96 | 9 | 15 | -5.64 |
| 10 | 65 | -10.28 | 15 | 4 | -5.62 |
| 14 | 57 | -10.24 | 3 | 47 | -5.57 |
| 5 | 43 | -9.54 | 11 | 72 | -5.51 |
| 15 | 81 | -8.93 | 11 | 45 | -5.41 |
| 1 | 11 | -8.40 | 15 | 61 | -5.39 |
| 5 | 49 | -8.14 | 3 | 2 | -5.37 |
| 5 | 46 | -8.04 | 11 | 77 | -5.37 |
| 10 | 74 | -7.91 | 8 | 18 | -5.36 |
| 11 | 17 | -7.70 | 11 | 19 | -5.35 |
| 7 | 79 | -7.47 | 11 | 36 | -5.28 |
| 9 | 5 | -7.37 | 9 | 13 | -5.26 |
| 5 | 42 | -7.35 | 11 | 41 | -5.25 |
| 11 | 24 | -7.26 | 13 | 83 | -5.17 |
| 10 | 22 | -7.20 | 3 | 31 | -5.14 |
| 12 | 53 | -7.19 | 11 | 50 | -5.11 |
| 10 | 29 | -7.16 | 8 | 23 | -5.03 |
| 6 | 69 | -7.03 | 11 | 73 | -4.99 |
| 10 | 20 | -7.02 | 8 | 85 | -4.93 |
| 10 | 44 | -7.01 | 5 | 40 | -4.81 |
| 13 | 82 | -7.00 | 6 | 25 | -4.71 |
| 10 | 71 | -6.90 | 11 | 21 | -4.58 |
| 11 | 48 | -6.84 | 8 | 37 | -4.56 |
| 10 | 70 | -6.69 | 8 | 86 | -4.56 |
| 13 | 28 | -6.64 | 11 | 66 | -4.51 |
| 4 | 6 | -6.59 | 8 | 59 | -4.49 |
| 11 | 12 | -6.34 | 8 | 9 | -4.47 |
| 10 | 76 | -6.31 | 11 | 30 | -4.41 |
| 4 | 7 | -6.26 | 9 | 3 | -4.37 |
| 14 | 32 | -6.07 | 9 | 63 | -4.24 |
| 11 | 68 | -5.95 | 11 | 80 | -3.99 |
| 1 | 27 | -5.93 | 8 | 75 | -3.87 |
| 1 | 55 | -5.89 | 11 | 35 | -3.68 |
| 6 | 84 | -5.89 | 8 | 78 | -3.62 |
| 4 | 16 | -5.87 | 1 | 60 | -3.36 |
| 2 | 34 | -5.76 | 8 | 39 | -2.86 |
| 2 | 33 | -5.72 | | | |

Table 3.7.: HTS 73 hits ranked by Glide XP Docking score.

According to docking results, best compound was HTS-1. However, since this molecule was a peptide with high molecular weight, docking prediction found several interaction points that accounted for a lower binding energy. It has to be pointed out that docking of peptidomimetic molecules returns an inaccurate binding energy. This *in silico* software is optimized for small organic molecules and thus, this result has to be carefully discussed.

A parallel analysis to the one done with the top 30 molecules with higher z-scores in the FP assay, was performed with docking data.

Cluster 10 was the one with a higher percentage of molecules in the top 30 ranking (9 compounds).

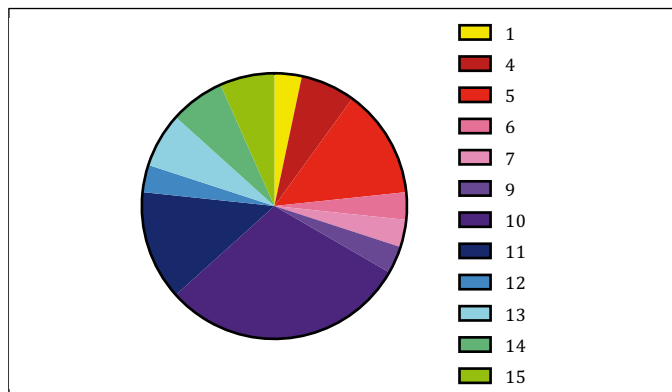


Figure 3.16.: Relative proportion of clusters depending on the number of molecules they had in the top 30 ranking of docking analysis. Percentages refer to the proportion of molecules from each cluster respect the total top 30.

A closer view to the top 30 molecules revealed that clusters 10 and 14 were enriched in molecules with low binding energy to POP. Thus, clusters 10 and 14 were confirmed as “hot-clusters” for POP inhibitors, regarding docking scores.

| Cluster | Molecules | Number of hits/tool | |
|---------|-----------|---------------------|-------------|
| | | Abs. Numbers | Percentages |
| 0 | 2 | 0 | 0 |
| 1 | 4 | 1 | 25 |
| 2 | 2 | 0 | 0 |
| 3 | 3 | 0 | 0 |
| 4 | 3 | 2 | 67 |
| 5 | 5 | 4 | 80 |
| 6 | 3 | 1 | 33 |
| 7 | 1 | 1 | 100 |
| 8 | 10 | 0 | 0 |
| 9 | 5 | 1 | 20 |
| 10 | 9 | 9 | 100 |
| 11 | 18 | 4 | 22 |
| 12 | 1 | 1 | 100 |
| 13 | 3 | 2 | 67 |
| 14 | 2 | 2 | 100 |
| 15 | 4 | 2 | 50 |

Table 3.8.: Number of molecules ranked in the best 30 positions for docking versus the cluster to which they belong. Percentages refer to the proportion of positives/cluster.

3.3.3. Enzymatic assay

3.3.3.1. Preliminary screening

For each cluster, representative molecules were chosen. The selection criteria were the following: 1) Actual availability in the lab, 2) Chemical representation of the molecule inside the cluster, 3) z-score for FP assays and docking results and 4) Price. The list of representative molecules is listed in the table 3.9.

| Cluster | Id | Cluster | Id |
|---------|----|---------|----|
| 0 | 38 | 10 | 20 |
| 1 | 55 | 10 | 70 |
| 2 | 34 | 10 | 71 |
| 3 | 31 | 10 | 76 |
| 4 | 6 | 11 | 12 |
| 4 | 7 | 11 | 21 |
| 4 | 16 | 11 | 30 |
| 5 | 43 | 11 | 35 |
| 6 | 25 | 11 | 41 |
| 6 | 84 | 11 | 45 |
| 7 | 79 | 11 | 73 |
| 8 | 9 | 12 | 53 |
| 8 | 39 | 13 | 28 |
| 8 | 75 | 13 | 82 |
| 9 | 3 | 14 | 32 |
| 9 | 5 | 15 | 1 |
| 9 | 13 | 15 | 4 |
| 9 | 15 | 15 | 61 |
| | | 15 | 81 |

Table 3.9.: Thirty-seven cluster representative molecules, ordered by their cluster.

POP activity was checked in the presence of the cluster representative molecules at a concentration of 200 μM . Then, percentage of POP inhibition was calculated and represented in a bar diagram (figure 3.17.).

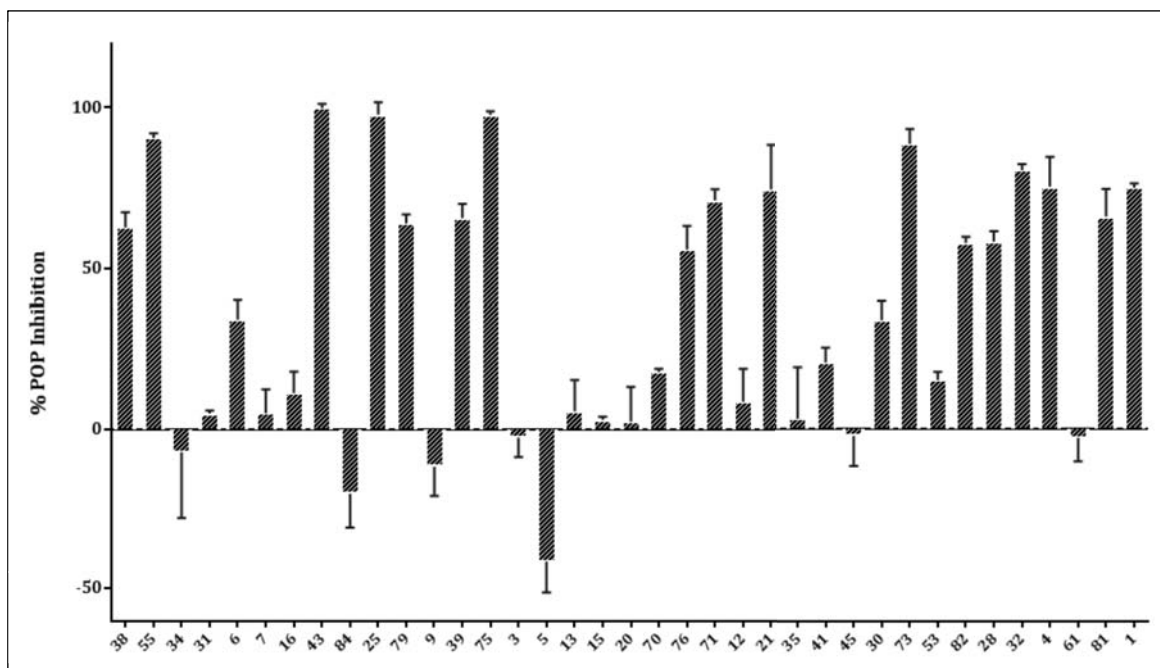


Figure 3.17. Percentage of POP inhibition of the 37 cluster representative molecules. Numbers on the x-axis refer to the HTS-Id code.

3.3.3.2. IC_{50} curves

From the 37 selected molecules for the initial enzymatic screening, compounds with a POP inhibition rate higher than 80% at 200 μM were selected: HTS-43 (cluster 5), HTS-25 (cluster 6), HTS-75 (cluster 8), HTS-73 (cluster 11) and HTS-32 (cluster 14). Besides these 5 molecules, HTS-1, which depicted the lowest energy in the docking calculation, was also selected. It has to be remarked that, HTS-1 was a peptide molecule and docking result should be analysed carefully. However, the high POP inhibition in the preliminary screening (75%) implied that HTS-1 was interacting with the active-site of the protease.

IC_{50} values were calculated for each compound. Results are shown in table 3.10.

| IC_{50} (μM) | | | | | | IC_{50} (nM) |
|------------------------------------|--------|--------|--------|--------|--------|-----------------------|
| HTS-1 | HTS-25 | HTS-32 | HTS-43 | HTS-73 | HTS-75 | ZPP |
| 59.9 | 27.0 | 37.3 | 5.9 | 33.8 | 3.1 | 1.8 |

Table 3.10.: IC_{50} calculation for selected. HTS hits. IC_{50} of ZPP was calculated as a control value.

HTS-75, from cluster 8 was the most potent POP inhibitor, followed by HTS-43 from cluster 5.

A thorough evaluation of the IC_{50} curves showed an unexpected behaviour. As it can be seen in figure 3.18., slopes of compounds HTS-25, 32 and 43 (in pink-purple colours; group A) were more pronounced than those of compounds HTS-1, 73 and 75 (in green colours; group B). A closer look to the curves evidenced that the slope of ZPP is similar to the one of the compounds in the group A.

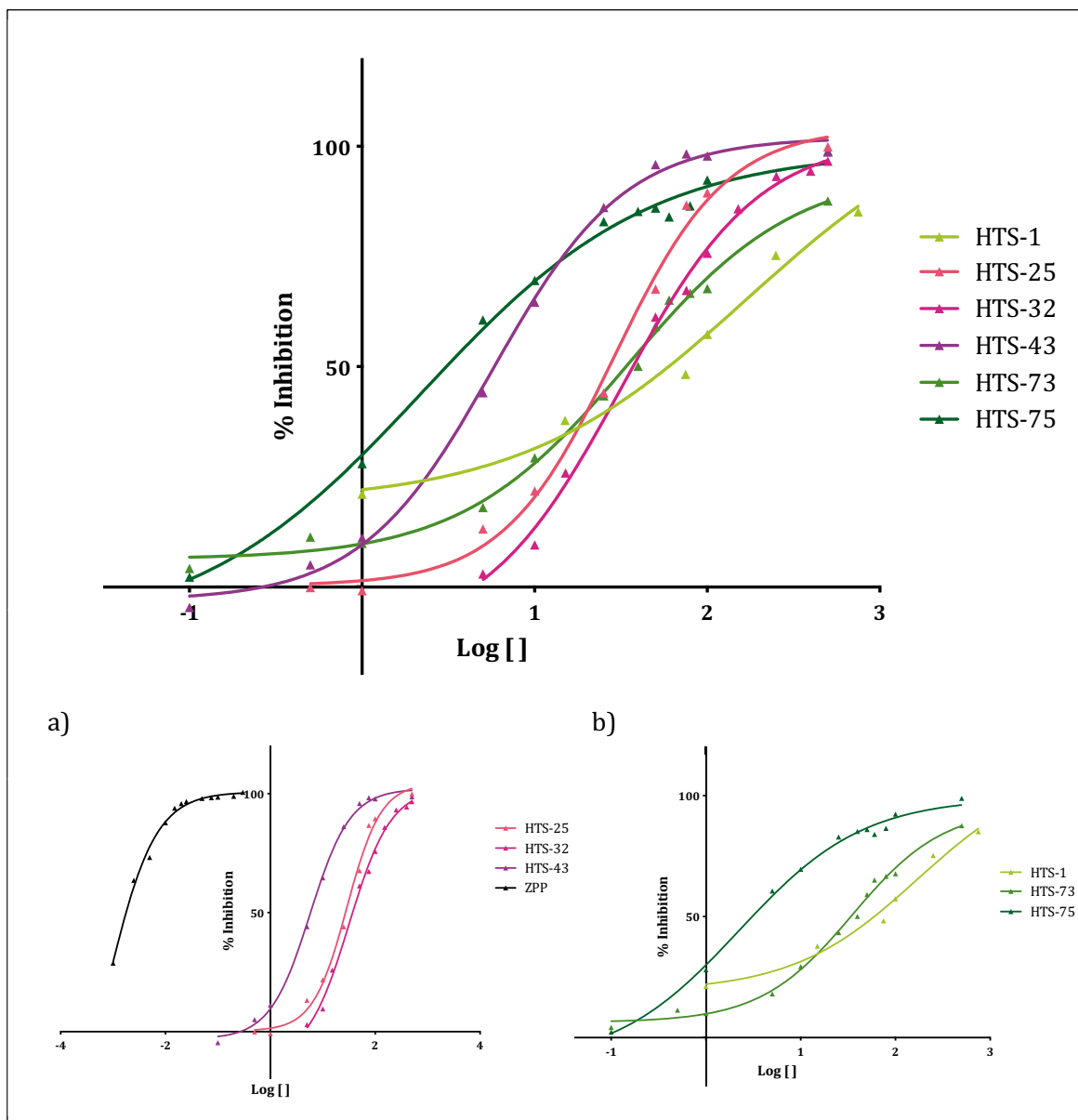


Figure 3.18.: IC_{50} curves of selected HTS hits and ZPP. a) Only group A and ZPP curves are represented. b) Curves belonging to group B inhibitors.

This result suggested the presence of two inhibition behaviours: a ZPP-like mode of action and an unknown type of inhibition. In order to fully understand the POP inhibition mechanism, kinetic analysis of HTS-43 (group b in slope) and HTS-75 (group A in slope).

3.4. Kinetic studies of lead candidates

3.4.1. Kinetic constants elucidation

A subset of concentrations of HTS-43 and HTS-75 were assayed for POP activity at a range of substrate concentrations. First, progress-curves (Fluorescence vs time) for each molecule concentration were done and the linearity time was defined as 30 minutes. Once the time window was set, initial velocities for each inhibitor and substrate concentration were calculated.

Then, the linear-transform of Lineweaver-Burk for each compound was done (figure 3.19.).

The Lineweaver-Burk linear-transform of HTS-43 data afforded a set of straight lines that intersected on the origin. This evidenced the competitive behaviour of HTS-43. Then, Dixon plot was represented, and the K_i value was obtained from the line intersection. (figure 3.20.).

On the contrary, the lines of Lineweaver-Burk plot of HTS-75 were not intersecting in a single point. This was suggesting a non-linear behaviour. In order to characterize the mode of action, the secondary plots were represented (figure 3.21.). The slope-plot, that is the representation of the Lineweaver-Burk slopes in front of the inhibitor concentration, demonstrated a parabolic behaviour. The intercept-plot (Lineweaver-Burk y-intercept vs compound concentration) depicted a linear mode of action. Thus, HTS-75 was characterized as a slope-parabolic intercept-linear noncompetitive parabolic POP inhibitor. Interestingly, HTS-75 was the first POP parabolic inhibitor to be reported.

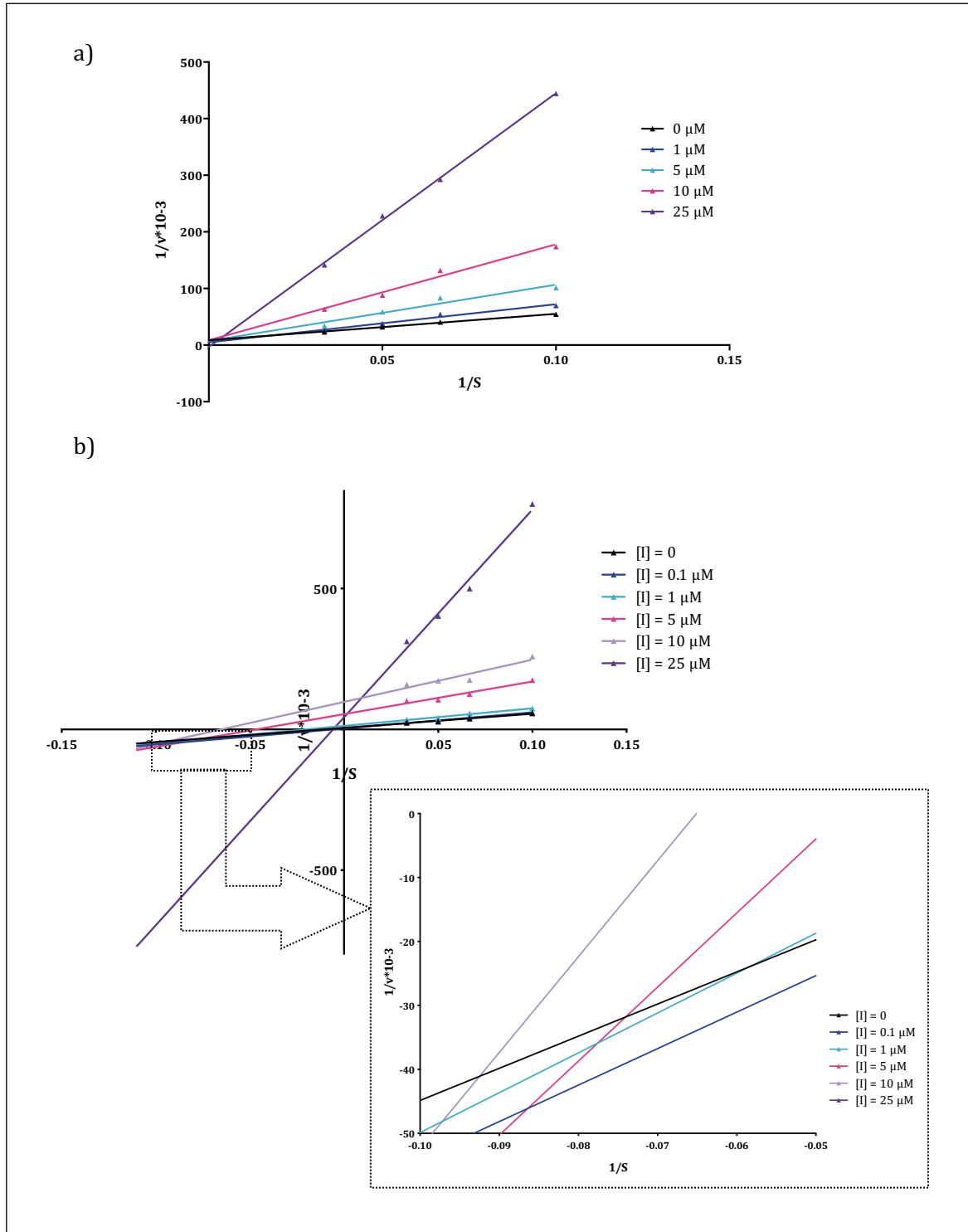


Figure 3.19.: Lineweaver-Burk representations of a) HTS-43 and b) HTS-75. In b) a zoom view is depicted.

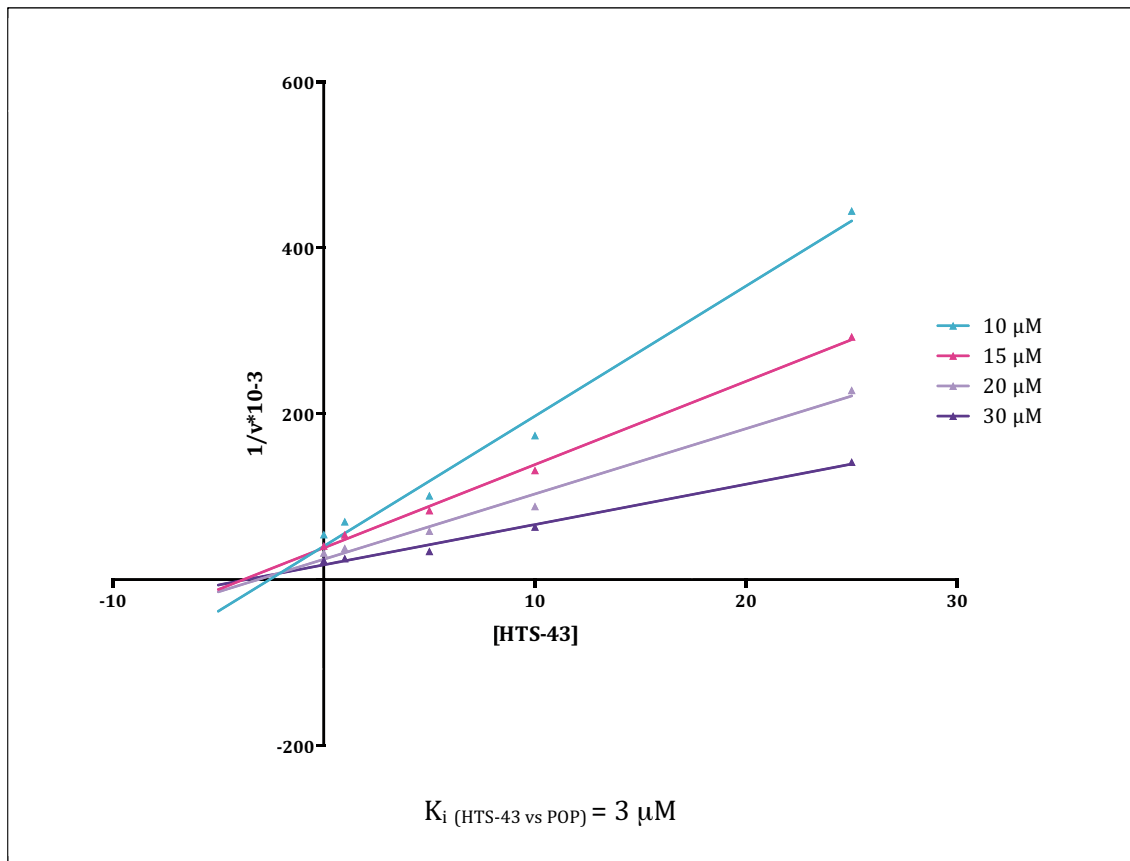


Figure 3.20.: Dixon-plot of molecule HTS-43.

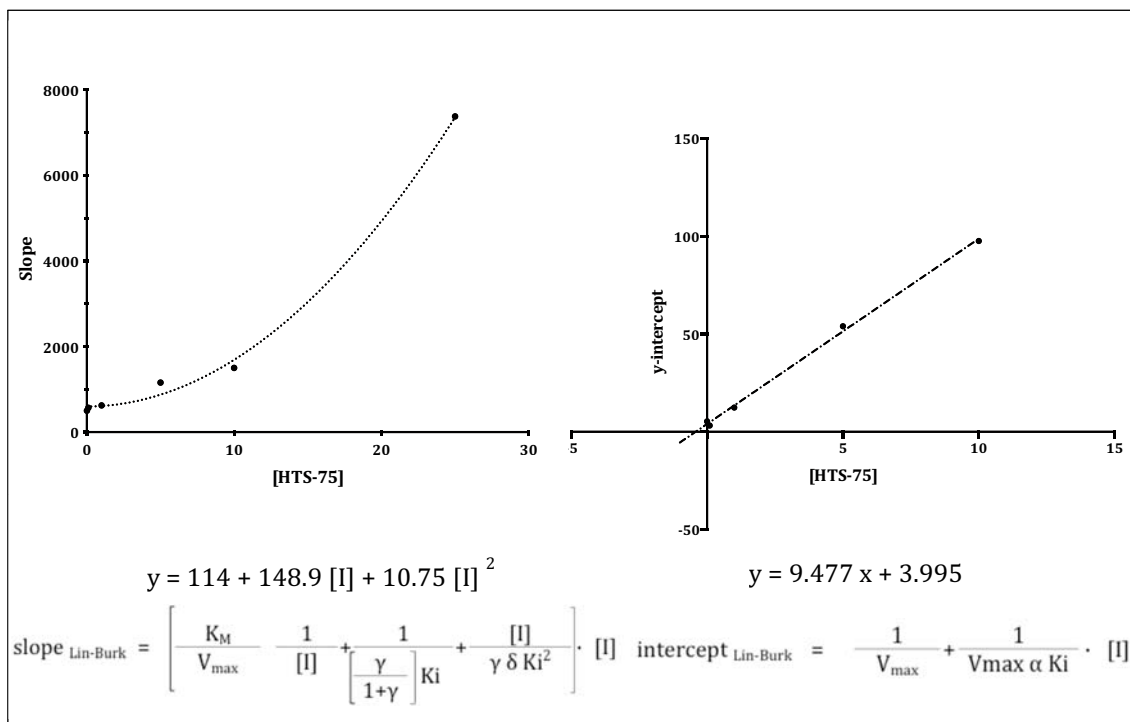


Figure 3.21 Secondary plots of Lineweaver-Burk representation of compound HTS-75.

Elucidation of the inhibition constants and parameters was possible by iteration. (figure 3.22.).

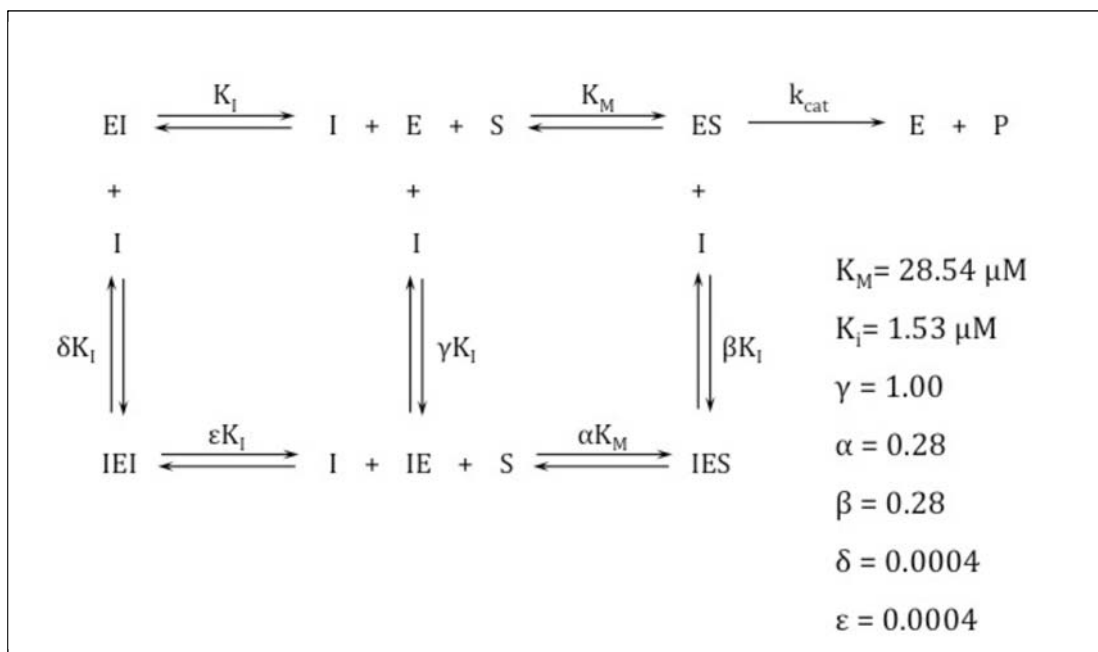


Figure 3.22.: Enzymatic scheme and inhibition constants of HTS-75 versus POP.

HTS-75 was described as low micromolar POP inhibitor. The affinity of HTS-75 to the competitive was equal to the one of the non-competitive binding site. However, binding of a second inhibitor molecule was almost impossible, independently of the first site of POP in being occupied.

Nevertheless, binding of the substrate to POP allowed the binding of HTS-75 in the non-competitive binding site. Thus, the effect POP suffers upon substrate binding differs from the one that exerts the binding of the compound in the competitive site.

Based on this, the scheme could be simplified as in figure 3.23.

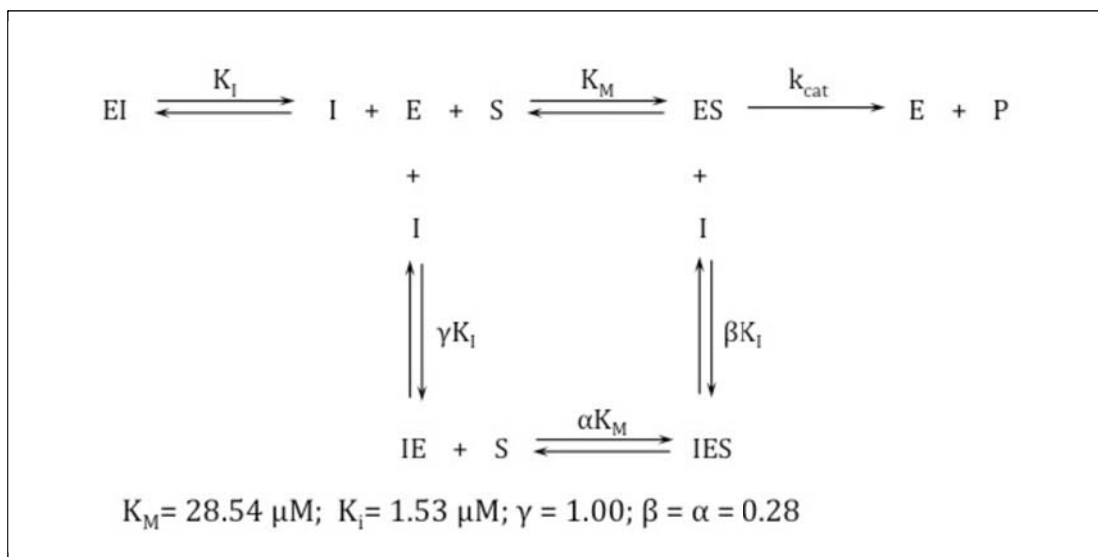


Figure 3.23.: Enzymatic scheme and inhibition constants of HTS-75 versus POP.

As mentioned in the introduction of this manuscript, POP is a highly dynamic protein, and samples two states: opened and closed forms.

A possible explanation for the HTS-75 unexpected behaviour is that binding of the inhibitor displaces POP to the closed state, independently of the binding site. Thus, occupancy of the non-competitive site would close POP, avoiding the entrance of either substrate or inhibitor, and union of HTS-75 to the competitive site would close POP as well, hiding the non-competitive site.

However the effect of the substrate would be different. The substrate may change POP conformation but would not be the closed form that resulted from inhibition. In spite, it could be a medium-closed form, where the competitive site is occupied by the substrate while the non-competitive one is free to be occupied by HTS-75. If the inhibitor is bound, the complex is inactive and transformation of substrate to product does not take place.

Nevertheless, this is a hypothesis that should be proved by combination with other techniques, such as high-field NMR.

Chapter 3 overview

In this chapter we planned the identification of POP inhibitors by HTS. Our strategy was based in the use of libraries containing non-toxic compounds and lead-like properties. The assay we selected was FP, which allowed the identification of protein binders by competition with a fluorophore-labelled probe.

First, the peptide probes were validated as useful probes for the FP assay. Then, the HTS was carried out. Over the 4,500 tested compounds, 73 hits were found to be POP binders. Later, 37 hits were selected by means of clustering, docking data and FP results in order to be validated as POP inhibitors. The subset of molecules was evaluated by enzymatic assays. As a result, six compounds presenting the highest POP inhibition ration were selected for further study. During the IC_{50} calculation of this small group, two inhibition behaviours were identified. A representative of each one was selected and kinetic assays were done for the two POP inhibitor candidates.

Finally, two POP inhibitors have been described. HTS-43 is a competitive POP inhibitor and HTS-75 displays a parabolic behaviour.

It is the first time that parabolic inhibition is reported for POP. We believe that the existence of a non-competitive site would help in the understanding of the relationship between POP and mental diseases.

Chapter 4:
Characterization of peptidyl
aryl vinyl sulfones as Cathepsins
L and B inhibitors

Chapter index

CHAPTER 4 CONTEXT

4.1. LIBRARY SCREENING

4.1.1. IC₅₀ calculation and structure-activity analysis

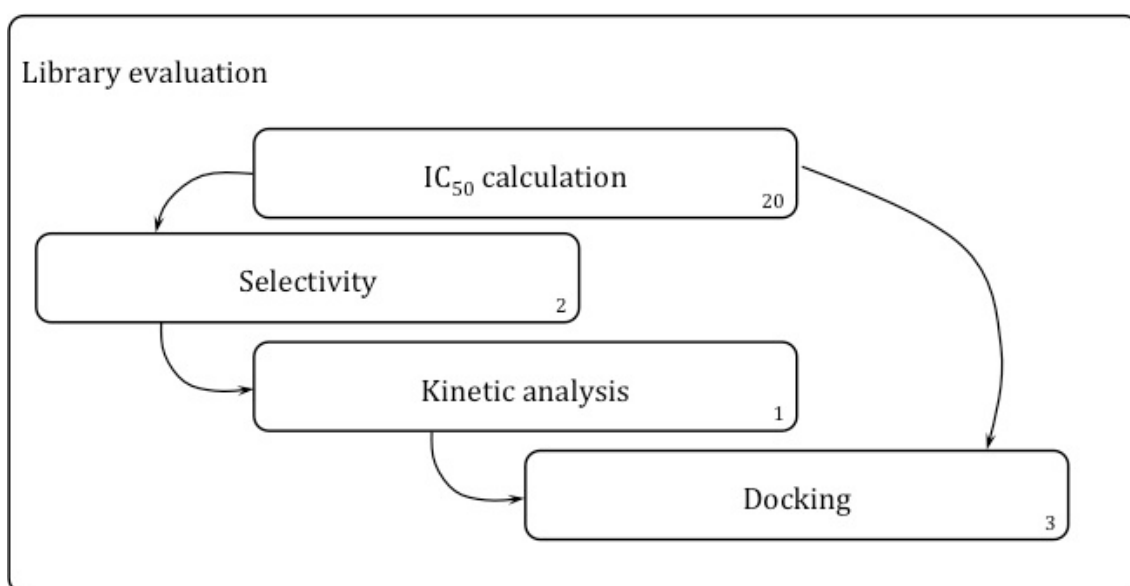
4.1.2. Selectivity of compounds

4.2. ELUCIDATION OF THE INHIBITION MECHANISM

4.3. DOCKING

CHAPTER 4 OVERVIEW

Experimental scheme



Numbers in corners refer to the number of compounds.

Chapter 4 context

Cathepsins L and B, two cysteine proteases, are relevant targets in cancer. These enzymes are involved in the proteolysis of tumor-surrounding tissues, allowing cancer cell invasion, tumor growth, angiogenesis and metastasis [128, 129]. Thereby, inhibition of cathepsins L and B is a key strategy for the development of new anticancer drugs.

An ensemble of cathepsins L and B inhibitors had been described on the literature, including reversible and irreversible inactivators. Among reversible inhibitors, cathepsin L inhibitors include molecules with aldehyde [203], aldehyde-hydroxamic acid combination [204], nonpeptidic cyanamide [205] and β -lactam penam and oxapenam [125] moieties as reactive group. On the contrary, compounds including dipeptidyl nitrile as a reactive group are selective reversible cathepsin B inhibitors [206]. Respect to irreversible inhibitors, two main groups are included: E-64 analogues [125] and alkyl vinyl sulfones. In the latter subgroup, peptidyl vinyl sulfones comprise a group of potent, selective and covalent inhibitors of cysteine proteases. Examples include inactivators of falcipains (*Plasmodium falciparum*) (Figure 4.1. a) [207], cruzipain (*Trypanosoma cruzi*) [125, 208, 209] (Figure 4.1. b), cathepsin S [210], cathepsin K [211], cathepsin F [212] and cathepsin L [207].

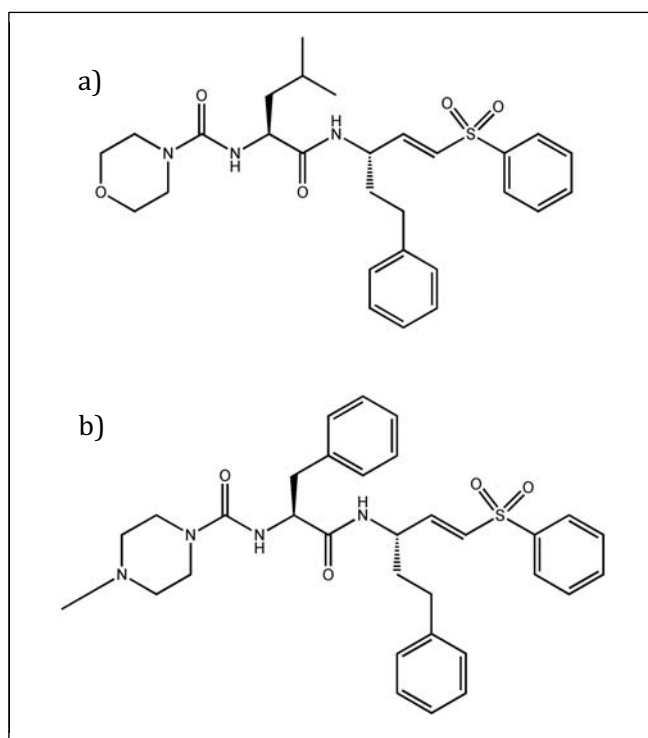


Figure 4.1.: Peptidyl vinyl sulfones reported in the literature as cysteine protease inhibitors. a) Antimalarial compound, discontinued to selectivity issues. b) Molecule for the African trypanosomiasis treatment, under active development.

All peptidyl vinyl sulfone molecules described in the literature are peptidyl alkyl vinyl sulfones. However, no peptidyl aryl vinyl sulfone has been characterized as cathepsins inhibitors. This is mainly due to the synthesis hurdles, such as racemization of the

intermediates. In order to solve this obstacle, our collaborators in this project (Dr. Albert Moyano, University of Barcelona; and Lluís Rafecas, Enantia) developed a route [213] that allowed the synthesis of peptidyl 3-aryl vinyl sulfones [214]. Compounds were prepared by Fmoc/*t*Bu solid-phase synthesis strategy, followed by in-solution coupling to the 3-amino-3-aryl vinyl sulfones. A library of 20 molecules was obtained, where building blocks were combined in 3 positions (figure 4.2.). Molecules were named as PAVS-xx, where xx was the number, from 1 to 20.

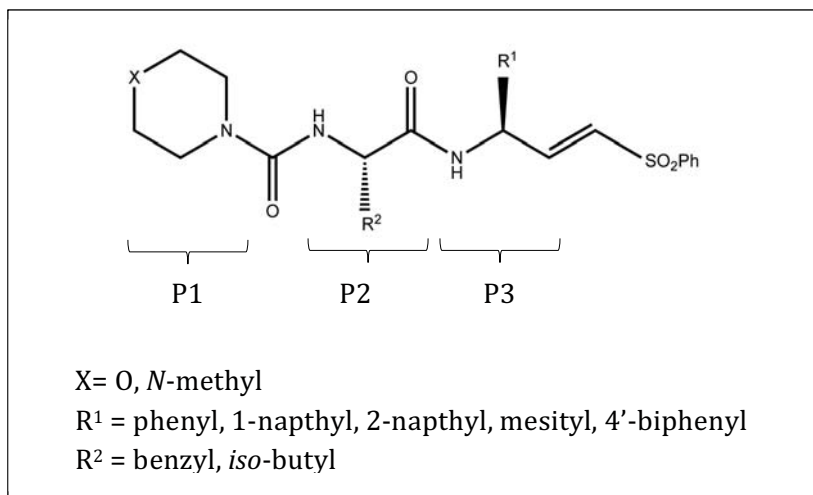


Figure 4.2.: Scaffold of the peptidyl 3-aryl vinyl sulfone library.

Position 1 (P1) could be either a morpholinyl group (if $X = \text{Oxygen}$), or a 4-methylpiperazinyl group (if $X = N\text{-methyl}$). Position 2 (P2) could be a Leucine (Leu) ($R^2 = \text{iso-butyl}$) or a Phenylalanine (Phe) ($R^2 = \text{benzyl}$). And the highest variability was introduced in position 3 (P3), which could be represented by 5 chemical blocks as R^1 substituents.

This library allowed, for the first time, the evaluation of structure-activity relationships with respect to aryl chain substituents in position 1. Furthermore, its chemical diversity could be increased by changes in the sulfone substituent, which was restricted for this study to a phenyl substituent. Thus, with a simple scaffold, a versatile library with biological application was obtained.

4.1. Library screening

4.1.1. IC₅₀ calculation and structure-activity analysis

The library was screened against human cathepsins L and B activity. The enzymatic assay consisted in a fluorimetric test with aminomethyl coumarin labeled substrates (Z-Phe-Arg-AMC for cathepsin L, and Z-Arg-Arg-AMC for cathepsin B). The concentrations for half-maximal inhibition were calculated as an initial screening assay and used for the selection of the most promising compounds ^[215] (table 4.1.). As a control, IC₅₀ values of E-64 for cathepsins L and B were calculated (IC₅₀ CatL = 8.3 nM; IC₅₀ CatB = 2.3 nM). Results were in agreement with those of bibliography, proving the reliability of the assay.

In basis of the results from the preliminary screening, a structure-activity study was done for each of the enzymes.

One general observation was that all compounds inhibited cathepsin L with higher potency than cathepsin B. Besides, 4-methylpiperazinyl group in position 1 afforded more active inhibitors. Regarding positions 2 and 3, differences in the two cysteine proteases preferences were observed. While cathepsin L did not discriminate between Leu or Phe in position 2, cathepsin B was inhibited in a more potency by compounds with Phe. This behavior may be caused by a favorable π interaction of Phe of the molecules with a specific residue of cathepsin B, located in the active site. Finally, in position 3, molecules with phenyl group exhibited higher cathepsin L inhibition, and compounds with 2-naphtyl group were preferred for cathepsin B. A possible explanation could be a reduced space in the active-site of cathepsin L, which would difficult the binding of compounds with a bulkier substituent.

Compound PAVS-20 was not only the strongest cathepsin L inhibitor (IC₅₀ = 2.6 nM), but also displayed a high selectivity versus cathepsin B (IC₅₀ CatB/IC₅₀ CatL = 404). Regarding cathepsin B, the more potent inhibitor was PAVS-13 (IC₅₀ = 304 nM), despite selectivity was not accomplished (IC₅₀ CatB/IC₅₀ CatL = 17).

| Compound | Substituents | | | IC ₅₀ [nM] | | | |
|----------|----------------|----------------|----------------|-----------------------|------------------|-------------------|-------------------------------------|
| | R ₁ | R ₂ | R ₃ | Cathepsin L | Cathepsin B | | |
| | | | | | | | |
| PAVS-1 | | | | 32 | 865 | | |
| PAVS-2 | | | | >10 ⁵ | >10 ⁵ | | |
| PAVS-3 | | | | 1000 - 1500 | 860 | | |
| PAVS-4 | | | | >10 ⁵ | >10 ⁵ | | |
| PAVS-5 | | | | 100 - 1000 | 415 | | |
| PAVS-6 | | | | | | 860 | 1700 |
| PAVS-7 | | | | | | 5x10 ⁶ | >10 ⁵ |
| PAVS-8 | | | | | | 10 | 703 |
| PAVS-9 | | | | | | >10 ⁵ | >10 ⁵ |
| PAVS-10 | | | | | | 930 | 660 |
| PAVS-11 | | | | | | 100 | 520 |
| PAVS-12 | | | | | | 5x10 ⁴ | 5x10 ⁴ - 10 ⁵ |
| PAVS-13 | | | | | | 17.5 | 305 |
| PAVS-14 | | | | | | >10 ⁵ | >10 ⁵ |
| PAVS-15 | | | | | | 8.5 | 300 |
| PAVS-16 | | | | | | 82 | 886 |
| PAVS-17 | | | | | | 536 | >10 ⁵ |
| PAVS-18 | | | | | | 5 | 690 |
| PAVS-19 | | | | | | >10 ³ | >10 ⁵ |
| PAVS-20 | | | | | | 2.6 | 1050 |

Table 4.1. Library of peptidyl 3-aryl vinyl sulfone compounds.

4.1.2. Selectivity of compounds

Selectivity of the peptidyl aryl vinyl sulfone library for cathepsins was evaluated. As a control, the serine proteases POP and DPP IV were selected (table 4.2.). No significant inhibition was observed for DPP IV for compounds PAVS-13 (best cathepsin B inhibitor) and PAVS-20 (strongest cathepsin L inactivator). Regarding to POP, it was slightly inhibited by these compounds, but the IC_{50} values were around 100 μ M, which affords enough selectivity.

| Enzyme | % Inhibition | | | |
|--------|---------------|---------------|---------------|---------------|
| | PAVS-13 | | PAVS-20 | |
| | [100 μ M] | [500 μ M] | [100 μ M] | [500 μ M] |
| DPP IV | 6 | 13 | 1.8 | 6.7 |
| POP | 44.5 | 70 | 48.6 | 81.2 |

Table 4.2.: Inhibition of the serine proteases POP and DPP IV by the library compounds PAVS-13 and -20.

4.2. Elucidation of the inhibition mechanism

Compound PAVS-20 was the most potent and selective cathepsin L inhibitor. Thus, it was selected for the enzymatic study.

First, reversibility was tested. A range of PAVS-20 concentrations were pre-incubated with cathepsin L. Followed that time, substrate was added and fluorescence was read after incubation time. The percentage of fluorescence respect to the sample without inhibitor (enzyme activity) was calculated and represented versus the pre-incubation time (figure 4.3.).

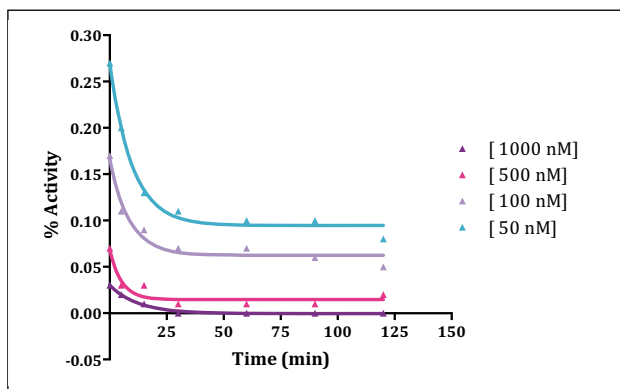


Figure 4.3.: Progress-curve of PAVS-20 versus cathepsin L respect the pre-incubation time.

The progress-curve graphic confirmed that PAVS-20 was a time-dependent inhibitor. Based on previous bibliography, sulfone-bearing compounds are well established covalent inhibitors of cysteine proteases. As such, they undergo a Michael addition with the thiol group of the active site cysteine to form an irreversible thioether bond [207, 216-219]. Then, it was hypothesized this mechanism for the peptidyl aryl vinyl sulfone library.

Once established the irreversibility of PAVS-20, the type of inhibition and the constants were determined. Regarding the type of inhibition, in the case of irreversible inhibitors the representation of k_{obs} versus substrate gives information about the mode of interaction. However, the traditional method used for reversible inhibitors is valid as well. Then, kinetic experiments for PAVS-20 against cathepsin L were performed. A set of compound concentrations was assayed for the enzyme activity at a range of substrate concentrations. Time linearity was set to 40 min based on the progress-curve representations (Fluorescence vs time). Then, the Lineweaver-Burk linear-transformation was done (figure 4.4.). From this plot, PAVS-20 was considered a competitive inhibitor.

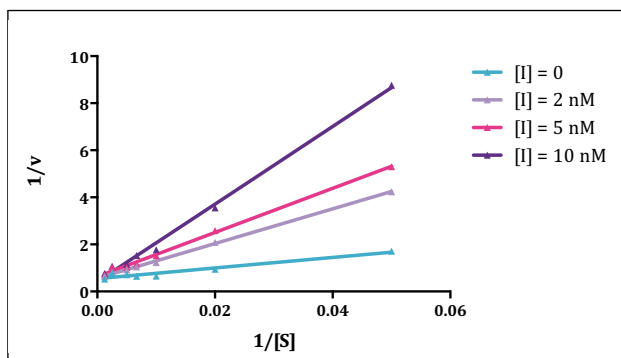


Figure 4.4.: Lineweaver-Burk representation of PAVS-20 versus cathepsin L.

Then, constants were determined. In the case of irreversible inhibitors, the value of K_i , derived from the Dixon-plot is not an appropriate approximation. Instead, constants should be calculated from the progress-curve of the pre-incubation time representation (figure 4.3.).

For each inhibitor concentration, data was fitted to a one-phase exponential decay, which decreases with a rate constant of k_{obs} , the apparent first-order rate constant of inactivation.

| [PAVS-20] nM | 1000 | 500 | 100 | 50 |
|--------------------------------|------|------|------|------|
| k_{obs} (min ⁻¹) | 0.08 | 0.19 | 0.12 | 0.10 |

Table 4.3.: k_{obs} values respect the inhibitor concentration.

Representation of k_{obs} versus PAVS-20 concentration afforded a hyperbola representation. (figure 4.5.). From this graphic, k_{inac} was determined as the maximum value, and K_{iapp} as the inhibitor concentration that afforded half of the maximum k_{inac} . k_{inac} is a first-order rate constant of inhibition related to the inactivation rate, while K_{iapp} is the apparent K_i .

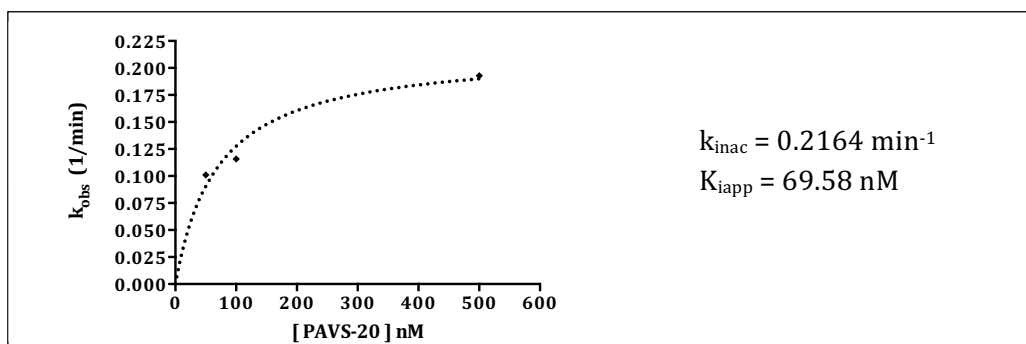


Figure 4.5. Representation of the first-order rate constant k_{obs} versus PAVS-20.

Finally, K_{inac} (dissociation constant, [I] at which the inactivation rate is half of the maximum) and k_{2nd} (second-order rate constant) could be obtained. These two parameters are the ones admitted in the literature for analysis of irreversible inhibitors. The constants were obtained by the formulas of figure 4.6. K_{inac} is the K_{iapp} value corrected by the substrate concentration, and the second-order rate constant. For its calculation the value of K_M was a premise. From the kinetic experiments previously performed for the Lineweaver-Burk representation, K_M was determined as the substrate concentration that yielded half of the maximum velocity at $[I] = 0$ ($K_M = 40 \mu\text{M}$). Finally, k_{2nd} is the relationship between the first-order rate constant and the affinity constant K_{inac} .

| | |
|---|---|
| $K_{inac} = \frac{K_{iapp}}{(1 + ([S]/K_M))}$ | $k_{2nd} = \frac{k_{inac}}{K_{inac}}$ |
| $K_{inac} = 19.9 \text{ nM}$ | $k_{2nd} = 181.420 \text{ s}^{-1} \text{ M}^{-1}$ |

Figure 4.6.: Formulas for the constant calculation of irreversible inhibitors and their values for PAVS-20 respect to cathepsin L

4.3. Docking

In order to provide a structural basis to peptidyl aryl vinyl sulfone library as cathepsins inhibitors, computational docking was performed. It has to be noted that no covalent binding was considered in the docking studies, and protein was set as rigid. Thus, since protein is flexible, reported distances may be smaller (after protein-inhibitor approach, potential hydrogen bonds may be favored).

Compound PAVS-20, the most potent cathepsin L inhibitor was docked into human cathepsin L (PDB accession number 1cs8). The molecule was extended from S2 to S2' (figure 4.7.).

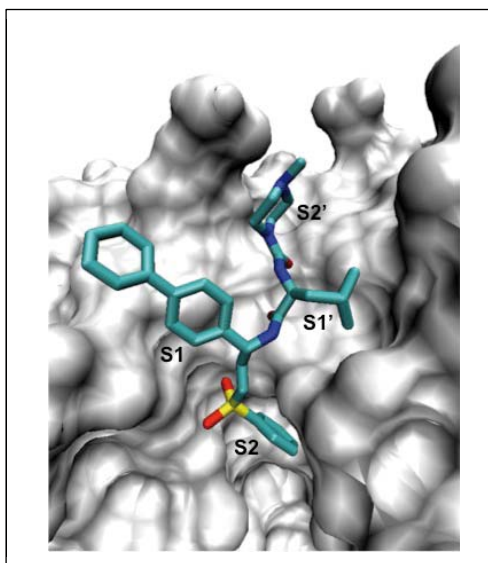


Figure 4.7.: PAVS-20 docked into cathepsin L active site. Subsites are showed.

The catalytic residues of cathepsin L are Cys 25, His 163 and Asn 187. Cys 25 interacted with the aryl vinyl sulfone through two potential H bonds, performed by the oxygen of the position 2 and the oxygen of the sulfone group (figure 4.8. a). Cys 25 was also close to the β -vinyl carbon. This proximity may favor the nucleophilic attack of the sulfur atom and subsequent formation of the covalent bond (Figure 4.8. b). His 163 also interacted with PAVS-20 through one H bond (Figure 4.8. c). Other two H bonds might occur after inhibitor approximation to protein.

In a closer view, interactions of each PAVS-20 moiety could be analyzed. Sulfonyl moiety, located at the S2, is one of the groups that showed most interaction with the cathepsin, through 2 H bonds and 3 other potential ones, involving Cys 25, Trp 26 and Gly 68. The 4'-biphenyl substituent was perfectly surrounded by Asn 66 and Gln 21 (figure 4.8. d). Other possible H bonds were present in the interaction between PAVS-20 and cathepsin L, involving residues Asp 162, Gln 19 and Trp 189, which might favor the contact between the two molecules. Finally, the 4-methylpiperazinyl group, placed at the S2', may form three H bonds with residues Gln 19 and Cys 22.

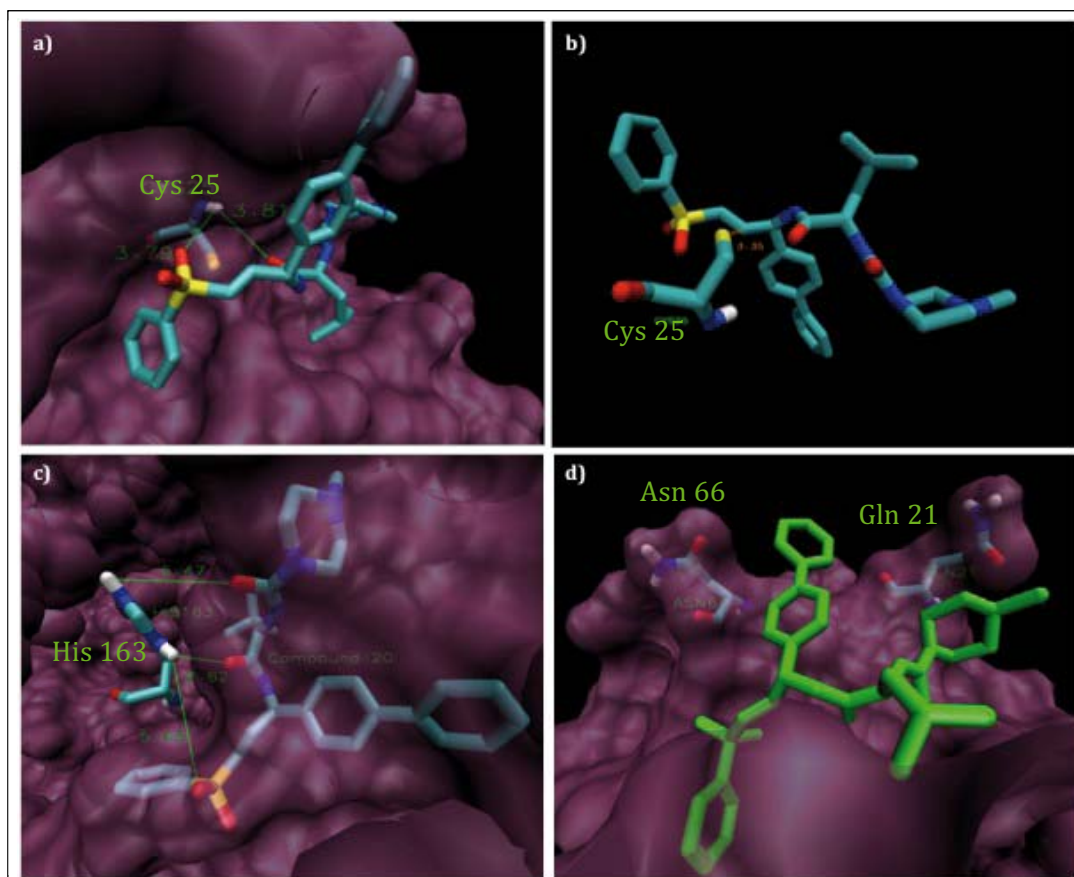


Figure 4.8.: PAVS-20 docked into cathepsin L active site. a) Interaction between Cys 25 and PAVS-20; b) Interaction between the sulfur atom of the active site cysteine and PAVS-20; c) Interaction between the active site His 163 and PAVS-20; d) Cleft defined by Asn 66 and Gln 21 into which the 4'-biphenyl substituent projected.

To further understand the differences in inhibition rates, compound PAVS-19 was docked as well against cathepsin L. PAVS-19 had a mesityl group while PAVS-20 had a 4'-biphenyl in position 2. In spite of this single change, the effectiveness of the inhibitor decreased 500-fold. As observed in the docking images (figure 4.9.), the difference in potency was attributed to the large volume that the mesityl group required (figure 4.9. b) in comparison with the 4'-biphenyl group (figure 4.9. a). Since mesityl is bulky, it did not fit into the area delimited by Asn 66 and Gln 21 (subsite S1), whereas PAVS-20 showed a perfect fit (figure 4.9. c). As a result, the backbone of PAVS-19 could not accommodate near to the cathepsin surface, thereby increasing the distances of the H bonds and decreasing the binding energy (figure 4.9. d).

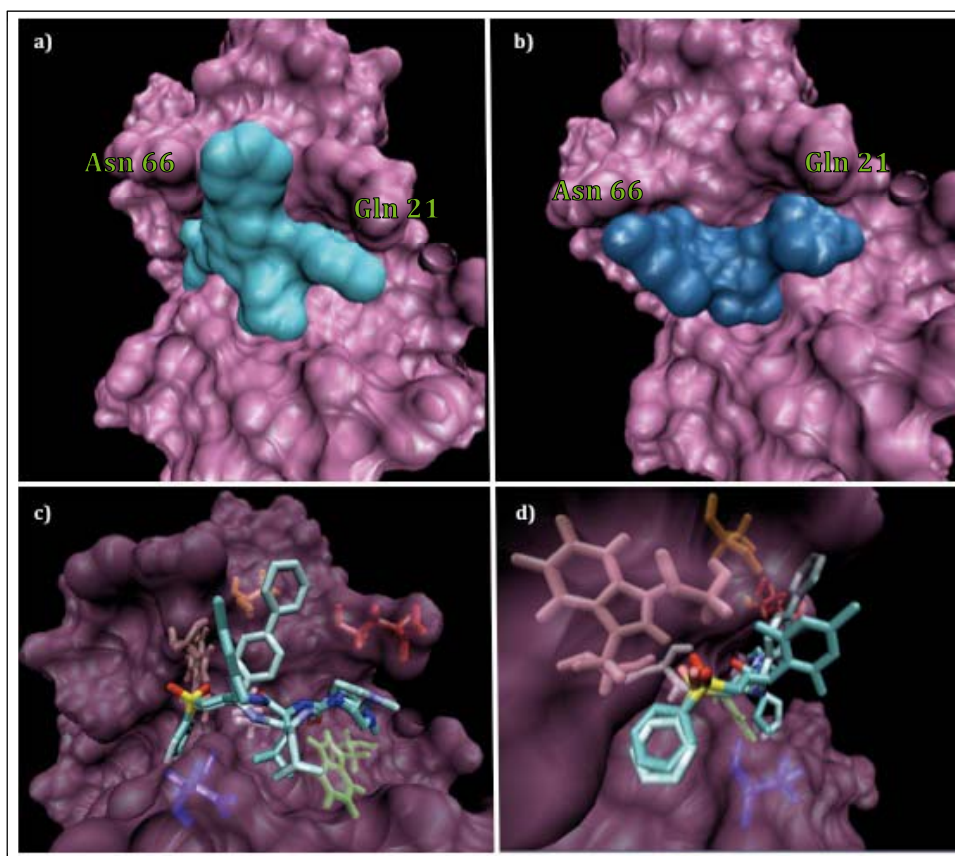


Figure 4.9.: a) PAVS-20 and docked into cathepsin L active site; b) PAVS-19 and docked into cathepsin L active site; c) PAVS-19 and 20 and docked into cathepsin L active site; d) PAVS-19 and 20 and docked into cathepsin L active site.

Finally, PAVS-13, the best cathepsin B inhibitor of the library, was docked into human cathepsin B (PDB accession number 1huc) [220].

The catalytic residues of cathepsin B are Cys 29, His 199 and Asn219. Both, the His and Cys residues interacted with PAVS-13 (figure 4.10. a) through hydrogen bonds. Besides, the sulfur atom of Cys 29 was also close to the β -vinyl carbon (as well as for PAVS-20 vs cathepsin L).

By analyzing the moieties of PAVS-13, it was noticed that this compound was interacting with the target through several hydrogen bonds. The sulfone group was likely to form a total of five hydrogen bonds with Gly 27, Cys 29, His 199 and Ala 200. The 2-naphthyl group was surrounded by His 111, Val 176, Leu 181 and Met 196 (Figure 4.10. c). Finally, the 4-methylpiperazinyl group formed 1 hydrogen bond and other 3 potential ones with residues His 110, His 111 and Trp 221.

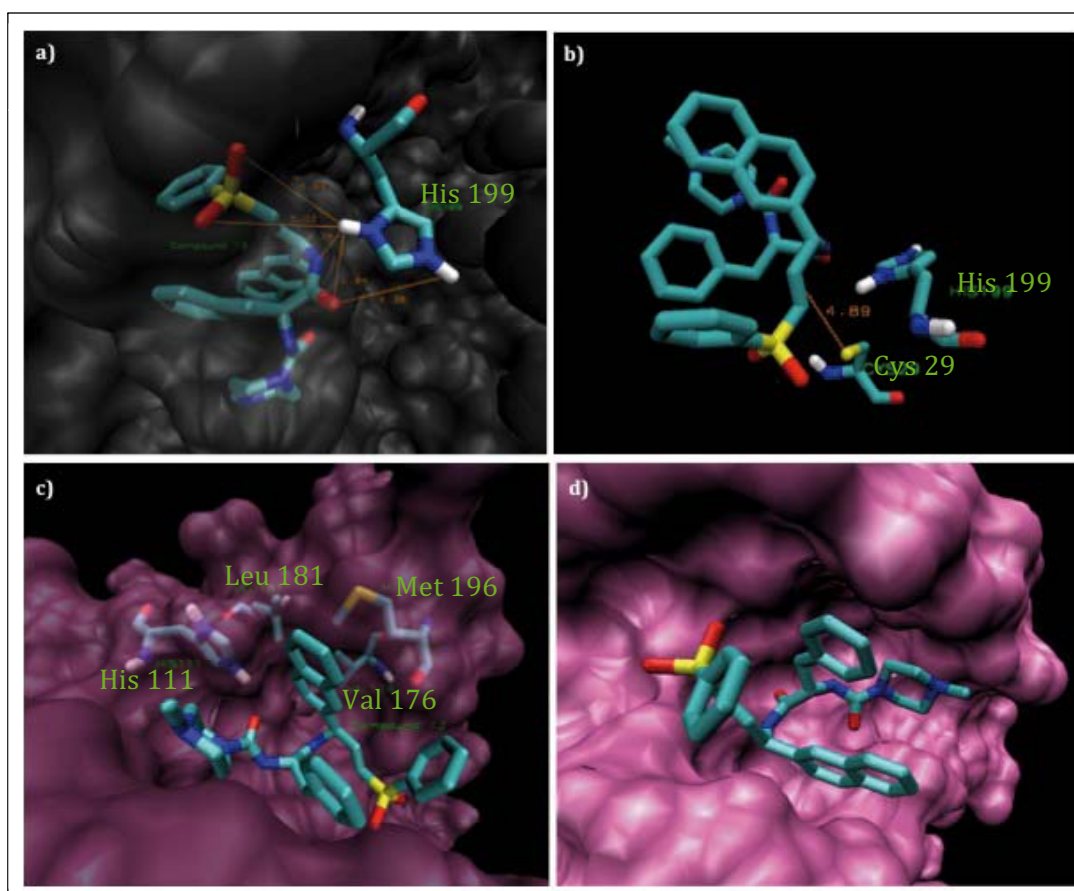


Figure 4.10. : PAVS-13 docked into cathepsin B active site. a) Interaction between His 199 and PAVS-13; b) Interaction between the sulfur atom of the active site cysteine and PAVS-13; c) 2-naphthyl group of PAVS-13 sitting into the cleft; d) Surface image of PAVS-13 docked into cathepsin B.

Chapter 4 overview

A novel peptidyl aryl vinyl sulfone library was tested for its inhibitory activity against cathepsins L and B. The scaffold was modified in three positions, namely P1, P2 and P3. For P1 and P2 there were two substituent groups while for P3 there were five moieties. Combinations of them gave rise to 20 molecules that were tested for cathepsins L and B proteolytic activity.

Interestingly, large differences in protease inhibition were detected. In general, compounds inhibited cathepsin L preferentially.

Among all the library molecules, a potent covalent irreversible cathepsin L inhibitor has been found, PAVS-20. The progress-curve of the pre-incubation time representation allowed the calculation of its inhibition constants.

Furthermore, evaluation of subsite preferences was done by docking analysis. This allowed understanding the experimental differences in inhibition constants obtained for similar compounds.

Since cathepsins L and B are targets for cancer, molecules of the PAVS library are promising candidates for the development of new anticancer drugs.

DISCUSSION

At the beginning of the present thesis, the main objective was the finding of new protease inhibitors. For all the proteolytic enzymes cited in here, DPP IV, POP and Cathepsins L and B, novel inactivators were found. However, the strategies followed in each case were different.

In the basis of the acquired knowledge, the present discussion will deal with the drug discovery steps in the protease inhibition field.

The goal

In drug discovery the final objective is clearly the finding of new active compounds. However, the goal, a more general concept, can vary depending on the application of the novel molecule.

Starting from the most basic research scenario, the inhibitor may be useful in order to obtain insights on the mechanism of action of a certain target. If nothing or little is known about a certain protease, phenotypic observation after inactivation can lead to hints about its function and/or roles. However, caution must be taken, since not selective inhibitors may provoke a misinterpretation.

Another use of the inhibitor may be target validation. In this case, functions and roles have not a consensus among the scientific community. Again, selectivity of the molecules is crucial for a good prognosis. Furthermore, animal models should be appropriate for the disease where the protease is believed to play a role.

Finally, the third goal is to take a drug onto the market. It does not mean only to report a new inhibitor, but also to improve the molecule properties in order to successfully achieve its commercialization as a therapeutic agent.

Following discussion is centred in this applied goal.

Key players

In the discovery of a new drug, the following aspects have to be taken into account.

First one is the market where to focus. It is a decision that is linked to the laboratory strategy. Secondly, once the therapeutic area is selected, the target of interest has to be identified. In a certain disease, several proteases may be involved. Then selection of the appropriate one is based on previous information (structure, expression, activity, anomalous expression and/or activity in the disease) and its druggability.

Afterwards, drug discovery process can be started. Since there is an ensemble of strategies, the choice is based in the previous knowledge of the target, basically structural information. However, special care must be taken in this step, because selection of the strategy is crucial for the success in the obtaining of a potential drug.

Once a hit is found, it still has to pass filters to become a lead, and additional ones to reach the market as a drug.

Thus, in drug discovery there are four main players: the market, the target, the strategy and the inhibitor. Each one of them is further detailed, focusing in the protease field.

The market: disease

The disease where to focus is a strategic criterion of the group of investigation or the company portfolio.

Decision based on market-size: large or small?

Regarding the size of the market, conditions whose market is large, or it is expected to be in the coming years, are more attractive to pharmaceutical companies, since revenues will be higher. Commonly, the objective is to generate a first-in-class drug or a blockbuster, a molecule which sales account for \$1 billion/year. The problem is that trends change along time. For example, first blockbuster was cimetidine (Tagamet), prescribed for the treatment of peptic ulcers. In 2011, atorvastatin (Lipitor), a cholesterol lowering agent, was the best sold, but only two years later is aripipazole (Abilify), for the treatment of depression, bipolar disorder and schizophrenia, the drug with the highest revenues. Therefore, taking into account that the time-to-market is between 15 and 20 years, I would not select a disease because its drug treatments are affording large sales in the present. An exception to this would be those diseases where the patient number is increasing and there is a possible new treatment (discussed later as mode of action).

On the contrary, rare diseases were neglected by pharmaceutical companies because of the reduced market. Thus, investigation in this field was carried out by small companies and public institutions, which pursued the discovery of orphan drugs (drugs that target rare diseases).

However, big pharmaceutical companies are now attracted by orphan drugs. This change is driven by several reasons. First, because of the process itself. Orphan drug discovery is shorter and thus, has less time-associated risks. Moreover it has tax advantages in some countries and more time of protection in front of generics. Second, once in the market, selling prices are higher than in other drugs. This is because the market is small, and competition is fairly inexistent. Thus, owing the monopoly, price can be fixed. And third,

potential market could be small, but the majority of the rare diseases are life threatening, which assures a chronic treatment for patients.

Thus, I would bet for rare diseases, since they suppose a niche with legal benefits and fast and safe economic return.

Decision based on chronic vs acute diseases

Chronic diseases, such as hypercholesterolemia, diabetes or high-blood pressure, suppose a daily-treatment. This assures constant income for the owner of the drug commercial rights. Thus, treatment of chronic diseases is a business where big pharmaceutical companies compete to be the first to reach the market and increase their revenues. Despite bring a drug to the market is risky, investing in pharma is a business opportunity.

Some acute diseases are also interesting in terms of economic return. The main example is cancer, a life-threatening disease with an expensive treatment.

But there are several acute diseases that seem to have been forgotten by big companies. Bacterial infections is one of them. In 2011 Pfizer, the leader on this field, closed the investigation on antibiotics and moved to vaccines, and nowadays few pharma industries devote efforts to the search of new antibiotics. Apart from originating low revenues, there are few targets and the FDA regulatory standards are high. Nonetheless, antibiotic resistance is promoting the appearance of harmful and extremely malignant new bacterial species. Thus, I expect that in the becoming years, new dangerous infections will appear and we will not be prepared to confront them in a fast and potent way. Antibiotics will be first-need drugs and their sales will generate considerable revenues.

Given the poor competition in the field, I think that **antibiotic drug discovery** is a good opportunity for starting a project and make a difference respect to the rest biotechnology companies. However, its main disadvantage is that it requires substantial basic research and it will not proportionate incomes in the close future, meaning that the cash flow will be negative. Furthermore, revenues are expected to be low. Then, what makes to me so interesting to invest on antibiotic research? Maybe the actual harmful bacteria are not widespread, but this situation is likely to change. The high-speed with which bacteria have acquired antibiotic resistance and are able to mutate, indicates that there will be even more life-threatening organisms all over the world in the future. Knowledge acquired in the search of antibiotics for the dangerous bacteria of today will be useful for the ones of tomorrow and will favour in reaching the market before than the competitors.

For that reason, a company solely based on antibiotic drug discovery would have difficulties to survive. Combination of a **moderate-risk line of business**, such as drugs for chronic and common diseases, and a more risky project for the discovery of new antibiotics would be a solution for the economic issue. Nevertheless, such a heterogenic company would need considerable human capital right from its starting point, enough to cover the different areas of expertise. Besides, financial entities may be discouraged to invest in a company which mission is not focused on a single objective.

Another strategy is to collaborate with public organizations. Universities and institutes actively work on the search of new antibiotics, but support of companies is imposed by the drug discovery process itself, in order to proceed with the clinical assays. Thus, whatever strategy is chosen, we need pharmaceutical industries investing on it.

The target

In a condition, which is the best target for drug discovery? Though a perfect target seems impossible to exist, there are preferred ones in basis of validation, druggability and competition.

Validated targets

Regarding the first aspect, it is desirable to have a validated target. It means that it has been previously established that its modulation affords beneficial effects for a certain disease.

Thus, validated targets, typically afford less risks in drug development process, since the mechanism of action has been previously elucidated.

On the contrary, the main inconvenient of these targets is that the competitor number increases with the time since validation.

Non-validated targets

If there are not validated targets in the condition of study, selection of the protease is based on previous knowledge of it. Principal aspects to be taken into account are: expression, structure, dynamism, substrate/s, protein-protein interactions, mechanism of action and regulation.

Respect to the expression, intracellular targets will require cell-penetrating properties of the drug, while molecules against proteins located in the CNS will have to cross the BBB. Despite that some molecules are able to overcome these obstacles, some others require the presence of a shuttle. Therefore, the obtaining of a drug for such targets is supposed to be more difficult, and thus, extracellular targets are preferred.

Structural elucidation is essential for structure-based drug design. Besides, with the help of computational methods, hotspots can be identified. Thus, alternative anchoring points for inhibition, such as allosteric sites, can be discovered. Allosteric inhibitors are, typically, more selective. For that reason their presence is an advantage in target selection.

Related to this, though crystallographic structures give a picture of the protein scaffold, dynamism has to be taken into account. In some cases, several conformations, of large structural differences, co-exist in equilibrium. Then, drug discovery will have to be directed towards a specific conformation and not to a rigid protein structure.

Knowledge of the function, signalling pathways and regulation of proteases has also to be taken into account in target selection. For example, endogenous substrates and inhibitors can give rise to mimetic inhibitors, which may simplify the structure-based drug design.

Finally, in order to select a target, previous reported bibliography on its inhibitors can be used as a direction for the design of new inactivators. Though the more information means the more competition, it also means better knowledge of subsite preferences for drugs.

If there are drugs in development or in the market, side effects of each one of them should be analysed. It is crucial to discern if the secondary effects are caused by the proteolytic

inhibition of the protease or by the eventual blocking of protein-protein interactions. For example, an inhibitor may not only ablate the protease activity, but by doing so, the overall protein structure may be modulated avoiding the interaction with partners. Then, side effects arise as a lack of protein cross-talking. In this case, problem can be solved by the design of molecules that do not modulate the PPI, opening a new niche for inhibitors.

Validated vs non-validated targets

To the best of my knowledge, **validated targets are far better than non-validated ones** in terms of drug discovery. The basic research is already done, which means less time and less money to invest. Besides, its relationship to the disease is established and it is more likely that there were animal models to test future drugs. Based on the fact that a great number of drugs fail in the demonstration of a positive effect in animal models, having the appropriate model is crucial for the success of the drug discovery.

I would select a non-validated target if there were not any treatment on the market for a certain condition. Though the process of validation would increase the time-to-market, being the first reaching it will provide high revenues, enough to cover the long and expensive drug discovery project.

Druggability

Does it exist a target whose drug discovery is easy and fast?

The druggability of a target depends upon which is the objective of the drug. In the case of proteases, the compound may inhibit the proteolytic function or disrupt PPI.

In the case of a desired proteolytic inhibition effect, molecules can be active-site directed or alternatively, compounds may bind to allosteric sites or exosites (if present). Presence of the latter in a protein makes the protease to be more plausible to be inhibited in a specific manner.

In the second case, the ligand-binding site is a large and flat surface. Small organic molecules fail in the disruption of PPI, while peptide moieties are recommended. However, problems such as low stability and biophysical properties are usually related to peptide derivatives.

While the first one is easiest due to the structural characteristics of binding pockets, the second one is a challenge. After several decades following the same drug discovery rules, it is believed that druggable targets are already covered and it is time to move to the non-canonical ones. In this regard, **molecules blocking PPI** are expected as the future drugs in a variety of diseases and I think that in the following years there will be a breakthrough in this area.

Competition

Competition and business opportunity are opposed terms, but not always.

Caution must be taken regarding competition. If someone has pursued inhibitors for the selected protease and has abandoned the project, it may be a sign that the target is undruggable or there are safety issues. If the investigation is still active, the competition may be closer to reach market, thereby our project would probably fell down and only if

the new drug is more potent or has less side effects, it will succeed. Besides, if competitors fail in any of the stages, an opportunity appears.

That is why **technologic watch** is crucial for any group in the drug discovery field. It will give competitiveness and will help in strategic decisions, such as target selection.

The positive aspect of competition is that there is more knowledge about potent and non-potent inhibitors. Especially in the case where the structure has not been elucidated, information SAR studies would help in drug design.

Nonetheless, I would select a **target without competition** that will afford a new mode-of-action (MoA). If competition is present but in a low degree, I would move for a new inhibition strategy. The good point of competition is that if the own compound succeeds in the pipeline early stages, there will be large pharmaceutical companies interested on it, and its licensing will be a profitable business. But competing with large pharmaceuticals is a crusade. This is the case of our DPP IV inhibitor search. We planned to find new inactivators from plant extracts. Despite we found an active molecule, the hit was substantially less potent than the actual drugs in the market. Then, which is our main advantage in front of the existing treatments? We hypothesized that compounds with a natural origin may have fewer side effects. However, the hit we found was poorly active and we realized that finding a DPP IV inhibitor in the nanomolar range would be nearly impossible. Thus, we discontinued the project.

In summary, I think it is the right time to move to more **challenging areas**, where few groups are in. It may represent an advantage in terms of being the firsts to reach the market, affording high revenues.

For example, I would go into detail about molecules disrupting PPI of POP and its partners. It has been suggested that these interactions are the effectors of conditions such as Parkinson's disease, schizophrenia and bipolar disease, but there are not any molecule of this kind in the market. Despite POP would have to be validated as a target, this field is a market niche with interesting future perspectives.

The strategy

Technique, which is related to the tool to assess compounds, and the source of molecules are the two main parameters of the drug discovery strategy. In here, knowledge of the protease structure is a great advantage.

First, a screening technique has to be selected. In the case of proteases, enzymatic assays are easy and standard tools. However, the techniques applied in protein for the assessing of ligand binding, are also suitable for proteolytic enzymes. The most typical include fluorescence polarization, NMR, crystallography, SPR and even MS.

Based on the experience of these years, I think that **enzymatic assays** are basic yet powerful experiments that not only give information on the potency of a molecule, but also describe the inhibition mechanism. Of course, **combination with other techniques** is needed in order to identify the exact residues involved in the interaction. I prefer

crystallography, since it is the most accurate technique for this objective, but in the case that protein crystallization is not possible, I would move to NMR (if the protein is smaller than 50 KDa) or informatics simulations. Using NMR for identifying hotspots in large proteins is, in my opinion, a hard and sometimes impossible task. Regarding informatics simulations, such as docking, I expect that there will be a great progress on these tools in a close future. This will simplify experimental work and reduce the time for testing a molecule. However, we have not reached the point where we can totally rely on computational data.

Regarding to the source for molecules, there are several options, but the three used in the present work are highlighted.

Medicinal extracts have been used since centuries ago and represent a powerful source, containing thousands of molecules. Besides, chemical variety is large, with new scaffolds to be discovered. However, the identification of the active molecule is a difficult and tedious task and there might be issues on protecting the API (active pharmaceutical ingredient) by means of a patent. This impossibilities making business of it. For example, the molecule found in chapter 2 to be a DPP IV inhibitor, was not possible to be patented due to “lack of serendipity” issues: one of the attributes for a compound to be protected is that its discovery has to be unexpected, mainly for two reasons.

First, as specified by the patent experts agency, if a molecule is discovered to act on a certain target, but the compound has already been related to the disease of the target, it cannot be patented. Thus, one may deduce that the action of the molecule is through this target. In our case, the molecule we discovered was already reported to have sugar-blood lowering properties, and DPP IV is a validated target of diabetes.

Secondly, the molecule may participate in other processes of the disease, and modulation of the given target is not the only action that leads to the beneficial effect. In fact, the DPP IV inhibitor we identified has been previously reported to be an α -glucosidase inhibitor, meaning that DPP IV inactivation was not the only activity of the compound.

Combinatorial chemistry is useful when detailed knowledge of subsite preferences is known. In this case, specific groups are selected depending on the properties of the protease residues located in the active site area. The main disadvantage is the limited chemical space that is explored. In the case of cathepsins, chapter 4, the library assayed afforded a potent compound, yet it resembled to previous cysteine protease inhibitors. Doing the same over and over during the past decades has give us powerful drugs, but the end of this era is arriving, and new and challenging strategies should be used.

HTS can be applied to a broad spectrum of targets in combination with techniques such as FP. Nowadays libraries are optimized, with compounds of better biophysical properties. Thus, once a hit is obtained, it is more likely to become a lead. The main disadvantage is related to the library nature. While in the past, effort was made in increase the quantity, now it is devoted to ameliorating the quality. However, molecules added to libraries are already described and limited to the known chemical space. Thus, an effort in basic chemistry has to be done for the input of “fresh compounds”. New scaffolds obtained from novel reactions may lead to the drugs of the future.

Based on that, if a new project was about to be started, I would select **HTS in combination with FP** for the identification of protease hits. Despite it supposes an elevated budget, HTS is a powerful technique in hit identification. I would use a mutant of the protease, where the residue of the active site has been substituted by another one, without interfering the protease structure. Thus, false positive that reacted with the active site but had no affinity for this pocket would not be identified. Besides, I would use enzymatic assays in order to confirm the hits as protease inhibitors, as done in chapter 3. Following this strategy a new and potent POP inhibitor has been found. Furthermore, by means of enzymatic assays, a novel inhibition mechanism has been identified. It is the first example of a parabolic inhibition on POP and suggests the presence of an, until now unknown, allosteric site. Taking into account that relationship of POP with mental conditions is believed to be through PPI, this site may be involved in those interactions.

Investigation on this direction will surely lead to potent and selective drugs for the treatment of Parkinson's disease, bipolar disease or schizophrenia. For that reason, I would continue with this project.

The inhibitor: hit to lead

The hit, in its way until becoming a drug, may fall due to failure in potency, selectivity, biophysical properties, and *in vivo* potency and safety.

In the hit to lead, modifications in the molecule are intended to diminish these problems.

Then, is there a guide to follow in this process? Obviously, each target is a completely different ecosystem and thus, inhibitor features will have to be discussed for each case.

These features refer to size, reversibility, formation or not of a covalent bond, and mode of inhibition (active-site or not).

In the particular case of a protease target for the treatment of a chronic disease, I would go for **small organic molecules** acting as **reversible inhibitors**. Regarding its binding site, if **allosteric sites** are present, they represent specific hotspots with respect to related proteases, which implies selectivity.

Once a compound has been established as potent and selective, pharmacokinetics (PK) is assayed: the drug has to achieve the appropriate biophysical properties (ADMET).

First, regarding administration, in my opinion, **oral drugs** are better accepted by the society than intravenous treatments. Thus, it implies that the drug must be stable during digestion, and be permeable to the intestinal wall.

For the description of the drug metabolism, I would perform P450 cytochrome assays since they are simple yet powerful *in vitro* assays. The results from the collection of studies performed in this step are input for the SAR studies, which can be done in parallel. As a result, **selective** and **bioavailable** drugs are obtained. However in this process, potency is sometimes lost. Nevertheless, there is equilibrium between potency, selectivity and biophysical properties that assures the success of a lead.

The inhibitor: lead to drug

In order to become a drug, the *in vivo* activity of a lead has to be confirmed, basically by testing the pharmacodynamics (PD). Activity, side effects, therapeutic window and time of

drug activity are key in order to define a lead as a future drug candidate. The molecule has to be in a defined window of potency and safety, which is very restrictive. Due to these constraints, several drugs fall in this step.

However, *in vivo* experiments in animal models are controversial in some diseases, such as schizophrenia or Alzheimer's disease, since models are not accurate or may not represent all the possibilities of a condition.

If a drug is safe in animals, it is more likely that it will be safe in humans as well. However, it does not discard unexpected events.

For that reasons, there is criticism around the high requirements in preclinical and its utility in the overall drug development process.

Nevertheless, it is a mandatory step that can also give feedback information about the drug behaviour, such as molecule organ distribution, metabolism and secretion.

Thus, changes in the structure of the lead can be proposed in order to obtain a definitive potent, selective and safe drug against a certain target.

Discussion overview

The ideas highlighted before are further detailed in here.

DPP IV, POP or cathepsins?

Based on the experience acquired during the present study, and as explained before, from the three drug discovery projects reported in here I would continue the project of POP inhibitors (chapter 3).

First of all, little competition exists nowadays in the field of POP inhibitors. Despite there is some, none of the large pharmaceutical companies is devoting efforts in this protease. Besides, there is no POP inhibitor registered in clinical phases to be under active investigation. Therefore, in the eventual finding of a potent, selective, safe and stable POP inhibitor, it will be a new MoA molecule. Thus, high incomes are expected.

Secondly, the compound that has been identified as a POP inactivator is performing a new type of inhibition, through an allosteric site. This area may be involved in PPI, which makes the project even more interesting, given the fact that POP is believed to be related to mental conditions through PPI. Then, the compound may be a potential therapeutic for the treatment of bipolar disorder, schizophrenia and Parkinson's disease.

Thirdly, mentioned conditions, despite not being life-threatening, affect life quality of both, patient and family. Besides, in the case of Parkinson's disease, there is an increase of subjects as the age increments. With the rise of maximum age in developed countries, interest for age-related conditions is gaining interest from the pharmaceutical industries, as they are business opportunities. Thus, market for POP inhibitors is assured.

However, there are disadvantages.

First, POP is yet a non-validated target. One may suggest that more time and research are needed, but there has been an intense effort in order to pursue this objective, without positive results. Then, it could be that in fact, POP is not a plausible target. However, I think that there could be another option. All studies performed until now have used competitive POP inhibitors, and it has been demonstrated that this inhibition mechanism has a beneficial effect *in vitro*. I hypothesized that the new mechanism of inhibition that is depicting the inhibitor we found would cause a different effect when tested *in vivo*. Of course, this is a merely speculation, but it is not ludicrous to think that allosteric sites are involved in PPI. Modifications of our hit may lead to potent PPI disruptors, that will, in turn, have a beneficial effect in the mental conditions.

It is, indeed, a challenging project, but not impossible. As I mentioned before, maybe all easy targets are already studied and it is time to move forward and go for more difficult targets. In this regard, POP is not the worst case scenario.

Second disadvantage is related to intellectual property. As known molecules were used in the HTS, they are already reported in the bibliography, and protection is for its use (reprofiling). Some are still under patent, which may force us to deal with the patent owners in order to obtain commercial rights for the repurposing in mental conditions. If the patent is expired, though no agreement should be reached, production of the molecule

all over the world would be legal. Protection would only extend to the drug use for the conditions we have included in the patent. Then, how to prove that external companies are commercializing the compound for other uses or for the one we have protected? In other words, patent prosecution is nearly impossible.

Final disadvantage is related to toxicity issues. The hit has passed these tests, but modification of it would be necessary in order to increase potency and lead with selectivity issues. Then, would be the modified hit non-toxic as well? The fact that toxicity should be tested increases the cost and includes uncertainty in the process.

However, after evaluation of pros and cons of the search of POP inhibitors, I think it is a promising area and business opportunity.

New project of protease drug discovery

If I had to start a new project, I would select bacterial infections as an area where to focus. I know it is risky business line with financial difficulties, but as mentioned before, combination with a safe project would solve this issue.

Investigation in dangerous bacteria of nowadays will afford drugs in the close future for a small market but will give the experience to be prepared for being the firsts when more dangerous bacteria appear. As commented before, I personally believe that we will experience an increase in bacterial eruptions, with higher potency and danger for the human life. Not only the threat would be large, but also the would easily wide spread, incrementing the number of patients. We need to be prepared for that, and once it arrives, we will be the firsts in reaching the market with new and potent antibiotic. Drugs would be, though single-treatment, a first-need for a large market.

Then, a bacterial protease should be identified as a target. Proteolytic enzymes are expressed in many bacterial organism, and are involved in processes such as infection, growth and proliferation. Thus, inhibitors of these proteases are promising antibiotics. Furthermore, there are few bacteria protease inactivators reported in the bibliography and none of them has reached the market. Taking into account that all commercialized antibiotics target proteins in the cell wall synthesis or in the ribosome synthesis, protease inhibitors will represent a new MoA, and a business opportunity.

Moreover, the fact that this type of drugs have not been used to control bacterial infection, there is no resistance to protease inhibitors.

In the drug discovery process itself, I would go for HTS in order to find new molecules, based on the potency of the technique. As seen with POP, a screening of nearly 5000 molecules afforded 73 hits, 2 of them with high activity. What if the library size is increased? Maybe even more potent molecules were found. In the particular case of antibiotics, molecules could be screened directly against the organism, but this is out of the scope of this discussion, since relationship of the drug with the protease would have to be demonstrated afterwards.

For the subset of molecules, I would go for small organic molecules, but I would not reject peptides. In the actual panorama, small molecules are more competitive than peptides in terms of drugs. Nevertheless, intense research has been devoted to this peptide drugs and

I expect to see the benefits in the future years. Thus, since drug discovery and development are long processes, it may be that during the time of antibiotic search, peptides start to be accepted by the totality of the science community as drugs for the treatment of a variety of diseases.

In the increase of potency and selectivity of the hit, computational tools are my choice. I expect that the new super computers together with the novel algorithms used in calculations will allow a better *in silico* representation of the protein/ligand binding event.

If computational tools prove to be comparable to experimental methods in the becoming years, I would use *in silico* methods not only for hit to lead process but also for the previous step, the hit discovery. Virtual screening would be the homologous of HTS without the need of bench work, reducing time and cost. But meanwhile, experimental methods are superior to computational in terms of confidence.

During the PK and PD studies, I would direct investigation into the obtaining of oral drugs with high standards for safety and selectivity in front human proteases and proteolytic enzymes of benign bacteria strains.

I think that for every disease there is a long pathway to reach the cure or the chronic treatment. Our decisions and strategy in the drug discovery process will help to shorten it, and is experience what assures the best choices. So, our duty is to keep on trying.

But drug discovery is a business as well, that generates high revenues in the most successful examples, but requires huge investment in all cases. It is a risky business, but certain scenarios are more promising than others.

Thus, we have to select the most attractive case to attract investors, but at the same time, a project with social repercussion.

Proposed scheme for protease's inhibitor drug discovery

Strategies followed in drug discovery can be tuned in different points. Next, there is a schematic plan of the process, with the questions that have to be answered and main aspects of each step. It has to be pointed out that if a step fails, the process can be iterated, adding changes based on the acquired information.

GOAL

Basic research or get to the market?

DISEASE

Related to business strategy, market size, chronic vs. acute diseases.

TARGET SELECTION

1) Compile information.

- Target:
 - Expression
 - Structure
 - Dynamism
 - Substrate/s
 - PPI
 - Mechanism of action (Pathway)
 - Regulation
- Inhibition:
 - Previous bibliography
- Objective of the drug in the target:
 - Inhibit proteolytic activity
 - Disrupt PPI

2) Analyse opportunities of drug targets

Compare druggability of the target vs. relevance of potential inhibitors

3) Select target depending on the goal, business strategy and drug opportunities

DISCOVERY OF A POTENT HIT

1) Select a technique:

- Enzymatic assays
- Fluorescence Polarization: requirement of a labelled probe.
- Nuclear Magnetic Resonance: labelling techniques are needed.
- Surface Plasmon Resonance: fixation of the protein to the support is required.
- Crystallography: known optimal crystallographic conditions have to be established. It is a long process that affords precise structural information.
- Mass spectrometry: irreversible and some reversible inhibitors.

2) Select the pool of inhibitors and tool to be used.

Decision is taken on basis of pre-existing target information.

- Required structure elucidation:
 - Combinatorial chemistry
 - Fragment-based screening
 - Virtual screening
- No required structural knowledge:
 - Natural sources
 - HTS
 - Fragment-based screening
 - Ligand-based virtual screening

3) Perform the experiment and fix criteria for positive hits

A hit may be considered positive in comparison with existing competitors.

4) Study inhibition properties of the hits:

- Reversibility
- Formation of covalent bonds
- Mode of inhibition
- Define kinetic constants as a potency parameter

OBTAINING OF A LEAD

Balance between potency, selectivity and biophysical properties must be found.

An optimal strategy is to perform a SAR study in parallel with the ADMET properties assessment.

| Potency and selectivity | | ADMET properties |
|-------------------------------------|------------------------|----------------------------|
| Computational methods | Synthetic methods | <i>In vitro</i> assays |
| Docking of derivatives | Synthesize derivatives | Administration: |
| Rank compounds (potency) | Test for potency | Stability assays |
| Propose modifications | Test for selectivity | Distribution: |
| Iterative docking | Propose modifications | Caco-2 assays |
| | Iterative synthesis | BBB <i>in vitro</i> models |
| For selectivity: | | Metabolism: |
| Dock derivatives in related enzymes | | P450 cytochrome |
| | | Toxicity: |
| | | Cell viability assays |

OBTAINING OF A DRUG

Confirmation of potency and low toxicity is performed in *in vivo* models.

CONCLUSIONS

Objective 1: To express DPP IV and perform an NMR study

1. Human DPP IV has been successfully **expressed in insect cells**, using the baculovirus technique. Several parameters including insect cell line, DPP IV construct, infection ratios and incubation times have been optimized.
DPP IV purification has been set-up in order to obtain pure protein. Purity has been assessed by means of FPLC and acrylamide gels.
Finally, DPP IV has been characterized as dimeric and active protease.
2. The protein has been used as a model for the **glycosylation pattern** study in the Mass Spectrometry facility of IRB Barcelona. Unfortunately, experiments have been unsuccessful.
3. A **selective labeling methodology** has been established for DPP IV, consisting on ¹³C-labelling of Methionine methyl groups.
Labeled DPP IV has been studied by Nuclear Magnetic Resonance using the **TROSY-HSQC** experiment. As a result, eleven of the 14 methionine signals have been detected in the NMR spectra.
Assignment of key residues by site-directed mutagenesis has been attempted, but all conditions failed.
4. NMR study of DPP IV in the **presence of inhibitors** has been performed. An extra signal appeared after incubation of the protease with the inhibitors.

Objective 2: To discover DPP IV inhibitors from herbal origin

1. A collection of plants extracts has been tested. This group includes plants of different origin: common antidiabetic plants, plants from the traditional Chinese medicine, Brazilian plants and Mediterranean plants: From this collection, the most promising extract corresponds to the ethanolic extract of the plant *Coutarea Latiflora* (AP-3). For that reason, it has been selected for further analysis. Extract fractionation and compound purification has afforded two DPP IV inhibitors (AP-3-a and AP-3-b) with IC₅₀ values against DPP IV of 115 μM and 340 μM, respectively.
2. The DPP IV inhibition mechanism of compound AP-3-a has been characterized by enzymatic analysis. It is described as a **reversible inhibitor with a parabolic behaviour** with a K_i value of 158 μM (for the competitive site).
3. An NMR study of DPP IV inhibited by AP-3-a has been done. The extra signal visualized in the control experiments from chapter 1 has not been observed in this case. An hypothesis for this phenomena is that the extra signal is characteristic of competitive inhibition, but disappears in complex inhibition type, such as the parabolic inhibitory mechanism of AP-3-a.

Objective 3: To discover POP inhibitors by High Throughput Screening

1. A Fluorescence Polarization assay has been set-up for the identification of POP active-site binders. Probes have been designed in basis of a competitive POP inhibitor (ZPP) and two fluorophore molecules have been used: carboxyfluorescein and TAMRA. Both probes have been used in the screening of POP binders by HTS.
A total of six libraries with 4,500 non-toxic molecules have been tested. This includes FDA-approved drugs and molecules that reached clinical and preclinical stages.
Data from HTS has been transformed into meaningful information (z-score). **73 hits** have been able to accomplish the predefined z-score limits for POP binders.
2. Hits have been clustered into 16 families, and representatives of each one of them have been selected in basis of docking analysis and FP results. 37 representative molecules have been screened against POP activity. **Five molecules** have inhibited POP with a percentage greater than 80%.
3. These five molecules, together with the one with best docking score, have been selected for the IC_{50} value calculation. Compounds were classified into **two groups**, A and B, depending on the slope of the IC_{50} curve. From each group, the most potent compound has been selected for a thorough kinetic analysis. Molecule HTS-43 (group A representative, $IC_{50} = 6\mu M$) has been defined as a **competitive POP inhibitor** with a K_i value of $3\mu M$. On the contrary, molecule HTS-75 (group B representative, $IC_{50} = 3\mu M$) has been characterized as a **parabolic POP inhibitor**, with a K_i value of $1.5\mu M$ (for the competitive site).
HTS-75 is the first reported POP parabolic inhibitor.

Objective 4: To identify cathepsins L and B inhibitors from a combinatorial chemistry library

1. A novel combinatorial library of peptidyl aryl vinyl sulfone molecules has been tested against cathepsins L and B activity. As a result, a potent cathepsin L inhibitor (PAVS-20) with an $IC_{50} = 2.6$ nM has been discovered.
2. The inhibition mechanism of PAVS-20 has been characterized by means of enzymatic assays. It has been defined as a covalent, irreversible inhibitor with $K_{inac} = 20$ nM and $k_{2nd} = 181.420$ s⁻¹ M⁻¹.
3. Finally, differences in protease inhibition of the library molecules have been explained by docking analysis. It has been observed that cathepsin L has a narrow **subsite S1, delimited by residues Asn 66 and Gln 21**. Thus, molecules with bulky groups in P2, such as PAVS-19 ($IC_{50}/catL > 10^3$ nM) do not fit in this groove. On the contrary, PAVS-20 ($IC_{50}/catL = 2.6$ nM), with a less bulky substituent in P2, fits perfectly. It has also been detected that **cathepsin B subsites are smaller** than the ones of cathepsin L, forcing the molecules to adopt a packed conformation, explaining the poor inhibition of the library compounds.

MATERIALS AND METHODS

Chapter 1. DPP IV Labeling methodology and the interaction with its inhibitors

1.1. Cloning

1.1.1. Gene and plasmid information

Homo sapiens DPP IV gene was bought at ATCC clones, Catalog No. 10437691.

GenBank ID: BC065265, BQ427580.

DPP IV gene was supplied into the carrier plasmid pCMV-SPORT6.

pFastBac EGT-N and pFastBac EGT-C were kindly provided by Dr. Arie Geerlof, at EMBL.

1.1.2. DPP IV cloning into pGEM-T Easy

1.1.2.1. DPP IV gene amplification and purification

The pCMV-SPORT6 plasmid was amplified in pre-culture tubes containing 3 mL of LB media supplemented with 50 µg/mL carbenicillin concentration. Cells were grown at T = 37°C during t = 1 hour at 250 rpm. Plasmid was then purified with the Kit "Illustra™ plasmid Prep Mini Spin Kit" from GE Healthcare. Amplification of the DPP IV DNA sequence coding from residue 39 to residue 766 was accomplished using the following primers:

FP Primer for the Nterm His-tag construct (≡FP3):

5'-CGCGGATCCCAGTCGCAAACTTACTACTCTA-3'

RP Primer for the Nterm His-tag construct (≡RP3):

5'-GGACTAGTTTAAAGGTAAAGAGAAACATTGTTT-3'

FP Primer for the Cterm His-tag construct (≡FP4):

5'-CTGACGTCAGTCGCAAACTTACTACTCTA-3'

RP Primer for the Cterm His-tag construct (≡RP4):

5'-CCGCTCGAGAGGTAAAGAGAAACATTGTTTTATGA-3'

PCR conditions were as follow:

- Primers (10 µM) : 2.5 µL each
 - dNTP's (10 mM): 2 µL
 - PCR Buffer 10x : 10 µL
 - H₂O: 77 µL
 - Template: 1 µL
 - Taq Polymerase (1U/ µL): 5 µL
- (Total Volume = 100 µL)

PCR reaction was carried out in a Thermocycler (MiniCycler, MJ Research), with the conditions:

- 94°C 5 min
 - 94 °C 30 s
 - 50°C 60 s
 - 72°C 3 min
 - 72°C 10 min
- } 30 cycles

In order to verify the amplification, an agarose gel was done.

Agarose gel: 50 mL of 1% agarose in TAE Buffer + 4 µL Ethidium Bromide.

- Markers (DNA Molecular Weight Marker III (0.12-21.2 kbp) and DNA Molecular Weight Marker VIII (19-1114 bp) from Roche): 4 µL of each marker + 1 µL 6xLoading Buffer.
- Samples: 5 µL of each one + 1 µL of 6xLoading Buffer
- V = 80 V t ≈ 40 min

PCR amplified genes were purified using the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” (GE Healthcare). For each PCR reaction (3/construct) 150 µL were eluted.

1.1.2.2. DPP IV gene ligation into pGEM-T Easy

Ligation was performed using the following conditions:

- 2xRapid Ligation Buffer (Promega): 5 µL
 - pGEM-T Easy Vector (Promega): 1 µL (50 ng)
 - PCR product: 3 µL
 - T4 DNA Ligase (3U/ µL) (Promega): 1 µL
- (Total Volume = 10 µL)

This conditions corresponded to 1:3 ratio (vector:insert).

Incubate at T = 4°C for t = O/N.

Then, XL-Gold cells (Agilent Technologies) were transformed with the ligated products.

For each ligation reaction, a 14 mL pre-culture tube was prepared with the following:

- XL-Gold cells : 100 µL
- β-mercaptoethanol: 4 µL

Tubes were swirled and incubated on ice for t = 10 min, swirling every 2 min by finger-flicking.

Then, to each tube:

- Ligated product: 10 µL

Tubes were swirled and incubated on ice for t = 30 min. After that, cells were heat-shocked at T = 42°C for t = 30 s, and then, incubated on ice for t = 2 min.

Right after, 900 µL of warm LB were added to each tube, and cells were incubated at T = 37°C during t = 1 hour. Next, 200 µL of each transformant were plated out in LB Agar plates supplemented with 50 µg/mL carbenicillin, 80 µg/mL X-Gal and 20 mM IPTG. Plates were incubated at T = 37°C O/N.

Next day, plates were checked for white colonies. For each construct, 12 pre-culture tubes of 15 mL were prepared with 3 mL LB cultures supplemented with 50 µg/mL carbenicillin. Then, media-containing tubes were inoculated with picked single white colonies. Tubes were grown at T = 37°C during t = O/N and constant agitation of 250 rpm. On the

following day, the plasmid was purified using the “Illustra™ plasmid Prep Mini Spin Kit” (GE Healthcare). For each colony, 50 µL of product were obtained.

In order to confirm the presence of the DPP IV gene into the pGEM-T Easy plasmid, sequencing reactions were performed. For each product:

- Premix (Applied Biosystems): 1 µL
 - 5xBuffer (Applied Biosystems): 2 µL
 - T7 Forward primer (10 mM): 1 µL
 - Plasmid: 3 µL
 - H₂O: 3 µL
- (Total Volume = 10 µL)

PCR reaction was carried out in a Thermocycler (MiniCycler, MJ Research), with the conditions:

- 96°C 1 min
 - 96 °C 10 s
 - 50°C 5 s
 - 60°C 4 min
- } 25 cycles

PCR tubes were brought to the Genomic Facility of the SCT and sequences were compared to DPP IV sequence with the FASTA database (Ncbi).

1.1.3. Gene cloning into pFastBac EGT-N or pFastBac EGT-C

1.1.3.1. Digestion and purification of DPP IV sequence

pGEM-T Easy containing the DPP IV for the Nterm His-tag was digested with BamHI and SpeI restriction enzymes at the same time. In one eppendorf, it was added the following:

- pGEM-T Easy-DPP IV for Nterm His-tag: 45 µL
 - BamHI (Roche): 5.63 µL
 - SpeI (Roche): 5.63 µL
 - 10xBuffer M(Roche): 11.25 µL
 - H₂O: 45 µL
- (Total Volume = 112.51 µL)

The eppendorf was incubated at T = 37°C during t = 2h

pGEM-T Easy containing the DPP IV for the Cterm His-tag was digested with AatII and XhoI restriction enzymes in two steps to avoid buffer interference. One eppendorf was prepared as following:

- pGEM-T Easy-DPP IV for Cterm His-tag: 45 µL
 - AatII (Roche): 5.63 µL
 - 10xBuffer A(Roche): 11.25 µL
 - H₂O: 50.63 µL
- (Total Volume = 112.51 µL)

The eppendorf was incubated at T = 37°C during t = 2h

The digestion product was then purified with the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” from GE Healthcare. Two eppendorfs of 50 µL/each were obtained. Right after, the second DNA digestion was carried out.

- pGEM-T Easy-DPP IV for Cterm His-tag: 45 µL
- XhoI (Roche): 5.63 µL
- 10xBuffer H(Roche): 11.25 µL
- H₂O: 50.63 µL
(Total Volume = 112.51 µL)

DPP IV gene with cohesive ends was purified by an agarose gel and cutting the corresponding band.

Agarose gel(x2): 50 mL of 1% agarose in TAE Buffer + 4 µL Ethidium Bromide.

- Markers (DNA Molecular Weight Marker III (0.12-21.2 kbp) and DNA Molecular Weigh Marker VIII (19-1114 bp) from Roche): 4 µL of each marker + 1 µL 6xLoading Buffer.
- DPP IV for Nterm His-tag construct: 112.51 µL + 20 µL of 6xLoading Buffer
- DPP IV for Cterm His-tag construct: 100 µL + 20 µL of 6xLoading Buffer
- V = 80 V t ≈ 40 min

Bands corresponding to the amplified sequence of DPP IV (2184 bp) were excised and purified using the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” from GE Healthcare.

1.1.3.2. Digestion and purification of pFastBac EGT-N and pFastBac EGT-C

pFastBac EGT-N was digested with BamHI and SpeI restriction enzymes at the same time. In an eppendorf, it was added the following:

- pFastBac EGT-N: 20 µL
- BamHI (Roche): 2.5 µL
- SpeI (Roche): 2.5 µL
- 10xBuffer M(Roche): 5 µL
- H₂O: 20 µL
(Total Volume = 50 µL)

The eppendorf was incubated at T = 37°C during t = 2h

pFastBac EGT-C was digested with AatII and XhoI restriction enzymes in two steps to avoid buffer interference. One eppendorf was prepared as following:

- pFastBac EGT-C: 20 µL
- AatII (Roche): 2.5 µL
- 10xBuffer A(Roche): 5 µL
- H₂O: 22.5 µL
(Total Volume = 50 µL)

The eppendorf was incubated at T = 37°C during t = 2h

The digestion product was then purified with the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” from GE Healthcare. Two eppendorfs of 50 µL/each were obtained. Right after, the second DNA digestion was carried out.

- Linear pFastBac EGT-C: 50 µL
- XhoI (Roche): 6.25 µL

- 10xBuffer H(Roche): 12.5 μ L
 - H₂O: 56.25 μ L
- (Total Volume = 125 μ L)

pFastBac EGT-C with cohesive ends was purified by an agarose gel and cutting the corresponding band.

Agarose gel: 50 mL of 1% agarose in TAE Buffer + 4 μ L Ethidium Bromide.

- Markers (DNA Molecular Weight Marker III (0.12-21.2 kbp) from Roche): 4 μ L + 1 μ L 6xLoading Buffer.
- Linear pFastBac EGT-C with cohesive ends: 125 μ L + 25 μ L of 6xLoading Buffer
- V = 80 V t \approx 50 min

Bands corresponding to the digested pFastBac EGT-C (4861 bp) were excised and purified using the "Illustra™ GFX PCR DNA and Gel Band Purification Kit" from GE Healthcare.

1.1.3.3. DPP IV gene ligation into pFastBac EGT-N and pFastBac EGT-C

Concentration of DPP IV gene and linear pFastBac's was estimated by loading a small amount in an agarose gel.

Agarose gel: 50 mL of 1% agarose in TAE Buffer + 4 μ L Ethidium Bromide.

- Markers (DNA Molecular Weight Marker III (0.12-21.2 kbp) and DNA Molecular Weight Marker VIII (19-1114 bp) from Roche): 4 μ L of each marker + 1 μ L 6xLoading Buffer.
- Samples: 5 μ L + 1 μ L of 6xLoading Buffer
- V = 80 V t \approx 50 min

DPP IV gene fragments were concentrated by Speed-Vac until dried. Then, ligation reaction was performed directly on the same eppendorf. To this tube, it was added:

- 2xRapid Ligation Buffer, T4 DNA Ligase: 5 μ L
 - Linear pFastBac EGT-N or pFastBac EGT -C: 5 μ L of 1/50 dilution
 - T4 Ligase (1U/ μ L) (Roche): 1 μ L
 - Buffer 10x (Roche): 1 μ L
 - H₂O: 3 μ L
- (Total Volume = 10 μ L)

Incubate at T = 4°C for t = O/N.

Then, XL-Gold cells (Agilent Technologies) were transformed with the ligated products. For each ligation reaction, a 14 mL pre-culture tube was prepared with the following:

- XL-Gold cells : 100 μ L
- β -mercaptoethanol: 4 μ L

Tubes were swirled and incubated on ice for t = 10 min, swirling every 2 min by finger-flicking.

After that, to each tube:

- Ligated product: 10 μ L

Tubes were swirled and incubated on ice for t = 30 min. After that, cells were heat-shocked at T = 42°C for t = 30 s, and then, incubated on ice for t = 2 min.

Right after, 900 μ L of warm LB were added to each tube, and cells were incubated at T = 37°C during t = 1 hour. Next, 200 μ L of each transformant were plated out in LB Agar plates supplemented with 50 μ g/mL carbenicillin.

Next day, plates were checked for colonies. For each construct, 12 pre-culture tubes of 15 mL were prepared with 3 mL LB cultures supplemented with 50 µg/mL carbenicillin. Then, media-containing tubes were inoculated with picked single white colonies. Tubes were grown at $T = 37^{\circ}\text{C}$ during $t = 1\text{h}$ and constant agitation of 250 rpm. Right after, the plasmid was purified with the "Illustra™ plasmid Prep Mini Spin Kit" (GE Healthcare). For each colony, 50 µL of product were obtained.

1.2. Bacmid and baculovirus production

1.2.1. Bacmid generation

For each construct, 100 µL of MAX Efficiency ® DH10Bac™ competent cells were aliquoted in a pre-culture tube. Then, 5 µL of DPP IV-pFastBac EGT-N or DPP IV-pFastBac EGT-C were added to the labelled tube. Samples were mixed by finger-flicking and incubated on ice during $t = 30\text{ min}$.

Right after, cells were heat-shocked at $T = 42^{\circ}\text{C}$ during 45 seconds. Next, tubes were cooled on ice for 2 minutes. Finally, 900 µL of SOC media were added and cells were grown at $T = 37^{\circ}\text{C}$ during $t = 4\text{ hour}$.

After that, 100 µL of a 1:100 dilution of the cell culture, were plated on LB Agar plates supplemented with 50 µg/mL kanamycin, 10 µg/mL gentamicin, 50 µg/mL X-Gal and 0.2 mM IPTG. Plates (one for each DPP IV construct) were incubated at $T = 37^{\circ}\text{C O/N}$.

Next day, plates were checked for white colonies. For each construct, 8 pre-culture tubes of 15 mL were prepared with 2 mL LB cultures supplemented with 50 µg/mL kanamycin and 7 µg/mL gentamicin. Then, media-containing tubes were inoculated with picked single white colonies. Tubes were grown at $T = 37^{\circ}\text{C}$ during $t = \text{O/N}$ and constant agitation of 250 rpm.

On the following day, the plasmid was purified using an in-house method.

First, for each cell culture, 1 mL was centrifuged on eppendorfs at 14,000 g during $t = 1\text{ min}$. This step was repeated for the rest of cell culture. Supernatant was removed and 300 µL of Solution I (15mM Tris-HCl, 10 mM EDTA, 100 µg/mL RNase A, pH 8.0 -previously filter-sterilized-) was added to each tube. Tubes were vortexed until pellet was resuspended. Next, 300 µL of Solution II (0.2N NaOH, 1% SDS - previously filter-sterilized-) were added to each eppendorf. Right after, tubes are mixed by inverting 10 times. Samples were incubated at r.t. during $t = 5\text{ min}$.

Next, 300 µL of 3 M Potassium acetate pH = 5.5, were added slowly and tubes were inverted 10 times. Eppendorfs were then incubated on ice for $t = 10\text{ min}$.

After that, tubes were centrifuged at 14000 g for $t = 10\text{ min}$. Meantime, for each sample, an eppendorf containing 800 µL of isopropanol was prepared. The supernatant was then transferred to the isopropanol eppendorf and tubes were inverted 10 times and incubated on ice for $t = 10\text{ min}$. Following that, eppendorfs were centrifuged at 14000g during $t = 15\text{ min}$. Supernatant was discarded and 500 µL of ethanol were added to each pellet. Tubes were inverted until no pellet was visible. Right after, eppendorfs were centrifuged at 14000g during $t = 5\text{ min}$. Next, supernatant was discarded and pellets were air-dried for

10 min. Finally, pellets were dissolved very carefully in 40 μ L of TE Buffer (10 mM Tris-HCl, 1 mM EDTA, pH = 8.0). Samples were stored at 4°C.

Recombinant bacmids were checked by PCR.

PCR conditions were as follow:

- Primers (10 μ M- FP3 and RP3 for Nterm His-Tag construct and FP4 and RP4 for C term His-tag construct-): 2.5 μ L each
- dNTP's (10 mM): 2 μ L
- PCR Buffer 10x : 10 μ L
- H₂O: 73 μ L
- Bacmid: 5 μ L
- Taq Polymerase (1U/ μ L): 5 μ L
(Total Volume = 100 μ L)

PCR reaction was carried out in a Thermocycler (MiniCycler, MJ Research), with the conditions:

- 94°C 5 min
 - 94 °C 30 s
 - 50°C 60 s
 - 72°C 3 min
 - 72°C 10 min
- } 30 cycles

1.2.2. Sf9 insect cell culture

A 1.5 mL aliquot of Sf9 insect cells adapted in Sf-900 III SFM (GIBCO) was thawed at 37°C and then transferred to a 25 cm² Corning flask containing 4 mL of Sf-900 III SFM.

The plate was incubated at 27°C for 1 hour. Then, cells were checked under the inverted microscope. Media was removed with a Pasteur pipette and 5 mL of fresh pre-warmed Sf-900 III SFM were added. Media was supplemented with 50 units/mL penicillin and 50 μ g/mL streptomycin (PS). The plate was then incubated at 27°C.

Cell growing was checked every 48 hours. If cells were confluent, they were detached with a sterile plastic Pasteur pipette and a 1/5 dilution was done in fresh pre-warmed media. If cells weren't needed, they were discarded. Once the cell culture was stable in plate (4-5 passages), cells were started to grow in solution. 5 mL of detached cells from a confluent 25 cm² plate were diluted in 25 mL of pre-warmed Sf-900 III SFM (with PS) in a 150 mL sterile flask with vented tap (Corning). Cells were grown at 27°C and 150rpm constant agitation. Cell viability was checked every 48 hours with Trypan Blue. 5 μ L of cells were mixed with 5 μ L of PBS and 40 μ L of 0.3125% Trypan Blue, giving a final dilution of 1/10 for cells and 0.250% of Trypan blue. From this mixture, 10 μ L were loaded into a Neubauer chamber. The number of alive cells/mL was the number of non-stained cells x 10⁵. The number of dead cells/mL was the number of blue-stained cells x 10⁵. Viability (percentage of alive cells versus total cells) was calculated. If it was lower than 90%, cell culture was discarded. When needed, cell culture volume was expanded from 30 mL (125 mL flask), to 60 mL (250 mL flask), 120 mL (500 mL) flask and 250 mL (1L flask).

When cell culture reached the 35th passage number, cells were discarded.

1.2.3. Baculovirus production

1.2.3.1. Transfection

Sf9 cells were seeded in a 6-well tissue culture plate at 9×10^5 cells/mL in 2 mL of Sf-900 III SFM supplemented with PS. One plate for each construct was prepared. Then, plates were incubated 1 hour at 27°C.

Meantime, DNA:lipid complexes were prepared. Mixture A was the DNA preparation. For the control wells (2/plate), 4 eppendorf tubes were prepared containing 5 μ L of H₂O and 100 μ L of unsupplemented Grace's medium (Invitrogen). For Nterm and Cterm His-tag DPP IV bacmid (4 wells/plate), 5 μ L of recombinant bacmid were mixed with 100 μ L of unsupplemented Grace's medium separately (4 eppendorf/construct). Mixture B was the lipid preparation. For the control wells, 6 μ L of Cellfectin(1:1.5(M/M) liposome formulation of N, N^I,N^{II}, N^{III}-Tetramethyl-N, N^I, N^{III}-tetrapalmitylspermine (TM-TPS) and dioleoyl phosphatidylethanolamine (DOPE) -Invitrogen) were mixed with 100 μ L of unsupplemented Grace's Medium (4 eppendorfs). For the recombinant bacidms 6 μ L of Cellfectin were mixed with 100 μ L of unsupplemented Grace's medium. (4 eppendorfs/construct). Mixtures A and B for control, Nterm and Cterm His-tag DPP IV bacmid samples were mixed, respectively. Eppendorfs were incubated at r.t. during t = 30 min. (4 eppendorfs for control, 4 eppendorfs for Nterm His-tag DPP IV bacmid and 4 eppendorfs for Cterm His-tag DPP IV bacmid)

After cells were incubated for 1 hour, the media was removed: Cells were washed with 2 mL of unsupplemented Grace's medium and then, washing media was aspirated. To each DNA:lipid complex tube, 800 μ L of unsupplemented Grace's medium was added. Tubes were inverted and the mixture was overlaid onto the cells. Then, cells were incubated at T = 27°C during t = 15 h.

Transfection mixtures were removed and 2 mL of PS supplemented Sf-900 III SFM was added to each well. Cells were incubated at T = 27°C during t = 3 days.

After that time, media from sample wells (4/construct) was transferred to 15-mL tubes and centrifuged at 500 g for t = 5 min. Pellet was discarded and supernatant was transferred to a glass snap-cap tube (Corning). Tubes were protected from light with aluminium foil and kept at T = 4°C. They were labelled as P1 baculovirus stock.

1.2.3.2. Virus amplification

In order to amplify the baculovirus stock, Sf9 cells were grown in solution at 1×10^6 cells/mL and infected with 1 mL of P1 baculovirus. Cell culture was incubated at 27°C and 150 rpm during 3 days. After that, media was collected in 50-mL falcon tubes and centrifuged at 500g for t = 5 min. Supernatant was transferred to glass snap-cap tubes, which were covered with aluminium foil and kept at T = 4°C.

P3 stocks were obtained by infecting 60 mL of in solution Sf9 cells at 1×10^6 cells/mL with 1 mL of P2 baculovirus stock. After 72 hours, media was centrifuged in 50-mL falcons at 500g for t = 5 min. Supernatant was transferred to 15-mL tubes. Tubes were frozen and kept at T = -20°C covered with aluminium foil.

1.3. DPP IV expression

1.3.1. Cell infection

250 mL of Sf9 cells at 3×10^6 cells/mL in Sf-900 III SFM were infected with 5 mL of P3 Baculovirus stock and were incubated for 4 days at 27°C and 150 rpm. After that time, cell culture was centrifuged at 1000 g for $t = 10$ min. Supernatant was filtered through a 0.2 μ m bottle-top filtration system and then diafiltered against 2L of 50 mM Tris-HCl, pH= 7.9, 400 mM NaCl buffer (Binding Buffer) in a Labscale (Millipore) at $T = 4^\circ\text{C}$. This process lasted for 2 days. After that, approximately 50 mL of concentrated supernatant were obtained.

1.3.2. Protein purification

Purification was performed at an ÄKTA Purifier (GE Healthcare) at $T = 4^\circ\text{C}$. A 1 ml Histrap HP column (GE Healthcare) was washed with 2 column volumes of Milli-Q H₂O and then, pre-equilibrated with 5 column volumes of Binding Buffer. After, supernatant was loaded into the column. A small amount of supernatant was kept for further analysis. Flow-through was collected in 50mL falcons and stored at 4 °C. Next, column was washed with 50 mM Tris-HCl, pH= 7.9, 400 mM NaCl, 50 mM Imidazole (Washing buffer) and washing fraction was collected as well. Finally, protein was eluted with 50 mM Tris-HCl, pH = 7.9, 400 mM NaCl, 200 mM Imidazole (Elution buffer). Aliquots of 1 mL were collected and kept at 4 °C.

1.3.3. SDS-PAGE electrophoresis

7.5% acrylamide/bisacrylamide gels of 1.5 mm were done using the recipe:

| | Resolving | Stacking |
|--------------------|-------------|-------------|
| Acrylamide/Bis 40% | 1.875 ml | 488 μ l |
| H ₂ O | 5.625 ml | 3.5 ml |
| Buffer | 2.6 ml | 1.25 ml |
| SDS 10% | 100 μ l | 50 μ l |
| APS 10% | 100 μ l | 500 μ l |
| TEMED | 10 μ l | 5 μ l |

Acrylamide/Bisacrylamide solution was 37.1:1 at 40% (w/v) (Amresco, Berlabo) SDS, TEMED and APS (Sigma).

Resolving buffer(or lower buffer) was 1.5 M Tris-HCl pH = 8.8, 0.4% SDS and stacking buffer (or upper buffer) was 0.5M Tris-HCl pH = 6.5, 0.4 % SDS.

Gel was then placed in a Tetrapak (Bio-Rad) cuvette and running buffer (1.92 M Glycine, 0.25 M Tris-HCl, 1 % SDS) was added into it.

For samples, 10 µL were mixed with 1.5 µL of 1M DTT and 3 µL of 5xLoading Buffer (312.5 mM Tris-HCl pH = 6.8, 50% Glycerol, 10% SDS, 0.1% Bromophenol Blue). Then, the eppendorfs were heated at 37°C during t = 10 min.

After that, gel was loaded. As a marker, 20 µL of Benchmark pre-stained protein ladder (Invitrogen) were loaded on the acrylamide/bisacrylamide gel. Then, samples were loaded.

Gel was run at 120V until the front band reached the bottom of the gel (approx. 1 h 30 min). In order to stain, the gel was then cleaned with Milli-Q H₂O three times and then incubated with Coomassie staining (0.25 g Coomassie blue G250, 100 mL Acetic acid and 900 mL of H₂O) for 1 hour at r.t. and mild orbital agitation. Gel was destained with decolourant solution (10% acetic acid in H₂O) for 1 hour at r.t. and mild orbital agitation. A piece of paper was added in order to help to destain the gel.

If dried, the gel was loaded on Whatman paper (GE Healthcare), then covered with transparent foil and brought to the dissecator. If not dried, gel was directly scanned.

1.3.4. Blue Native-PAGE electrophoresis

1.3.4.1. 4%, 5% and 7.5% acrylamide gels

7.5% acrylamide/bisacrylamide gels of 1.5 mm were done using the same recipe as in the SDS-PAGE electrophoresis but without SDS.

Regarding 4% and 5% gels, instructions were the following:

| | 4% | | 5% | |
|--------------------|-----------|----------|-----------|----------|
| | Resolving | Stacking | Resolving | Stacking |
| Acrylamide/Bis 40% | 1 ml | 488 µl | 1.25 ml | 1.464 µl |
| H ₂ O | 6.375 ml | 3.5 ml | 6.125 ml | 10.5 ml |
| Buffer | 2.6 ml | 1.25 ml | 2.6 ml | 3.75 ml |
| APS 10% | 100 µl | 500 µl | 100 µl | 500 µl |
| TEMED | 10 µl | 5 µl | 10 µl | 5 µl |

Resolving buffer was 1.5 M Tris-HCl pH = 8.8 and the stacking buffer was 0.5M Tris-HCl pH = 6.5.

Running buffer was 1.92 M Glycine, 0.25 M Tris-HCl.

Sample loading buffer (5x) was 312.5 mM Tris-HCl pH = 6.8, 50% Glycerol, 0.1% Bromophenol Blue)

Gel was run and stained in the same conditions as the SDS-PAGE electrophoresis.

7% acidic acrylamide gel

For the acidic native electrophoresis at 7%, the conditions were:

| | Resolving | Stacking |
|--------------------|-----------|----------|
| Acrylamide/Bis 30% | 3.2 ml | 1.875 ml |
| H ₂ O | 3.3 ml | 5.625 ml |
| Buffer | 3.35 ml | 2.6 ml |
| APS 10% | 100 µl | 500 µl |
| TEMED | 20 µl | 10 µl |
| Glycerol 50% | 3.0 mL | - |

Running and staining conditions were the same as for non-acidic Native-PAGE electrophoresis.

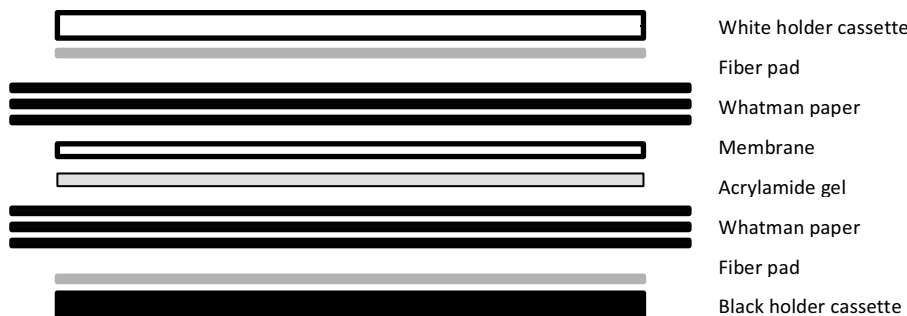
1.3.5. Schiff's staining

Acrylamide gels (both SDS- and Native-) were stained with Schiff's reagent (Fuchsin-sulfite reagent, Sigma) following the instructions given by the provider.

1.3.6. Western Blot

Acrylamide/Bisacrylamide gel was equilibrated in transfer buffer (25 mM Tris, 192 mM Glycine, 20% Methanol) for t = 15 min.

The sandwich for the transference was prepared as indicated in the figure:



Transference was run at $I = 300$ mA during $t = 1$ h 30 min.

After, membrane is transferred to a box and 40mL of 10% milk in PBST (PBS+0.1 Tween-20) are added. Membrane is blocked O/N at 4°C and constant agitation in an orbital shaker.

Next day, membrane was cleaned with PBST in an orbital shaker at r.t. (3 times x 5 min). Right after, 1st antibody incubation was done. The antibody CD26 antibody-spacer region (Abcam) was diluted 1/2500 in 5% Milk in PBST (4 μ L 1st antibody + 10 mL 5% milk in PBST). The membrane was placed in a sealed plastic bag, where the 1st antibody was added. The bag was pinned in a rotary wheel during 1 hour at r.t. Next, the bag was opened and the membrane was cleaned with PBST in an orbital shaker at r.t. (3 times x 5 min). After, 2nd antibody incubation was done. The antibody ECL Rabbit IgG, HRP-linked, whole Ab (from donkey) (GE Healthcare) was diluted 1/10000 in 5% Milk in PBST (1 μ L 1st antibody + 10 mL 5% milk in PBST). The membrane was placed in a sealed plastic bag and incubated with the 2nd antibody in constant agitation with the rotary wheel during 1 hour at r.t. Then, bag was opened and the membrane was cleaned with PBST in an orbital shaker at r.t. (2 times x 5 min + 1 time x 10 min). Finally the reaction is developed with the kit ECL Western Blotting Detection Reagents 1000 cm² (GE Healthcare). Films (Fujifilm) were exposed at different times and then developed.

1.3.7. Protein concentration and buffer exchange

Buffer exchange was achieved by using a HiPrep 26/10 Desalting column (GE Healthcare) in the ÅKTA Purifier at $T = 4^\circ\text{C}$. Column was equilibrated in Binding Buffer (50 mM Tris-HCl, pH = 7.9, 400 mM NaCl) and then, selected elution tubes were injected.

Protein elution peak was then, concentrated using the Vivaspin 20 100 KDa MWCO (Millipore). First, the membranes were cleaned with Milli-Q H₂O and then, 20 mL of Milli-Q H₂O were centrifuged at 3000 g. Next, membrane was adapted in Binding buffer (50 mM Tris-HCl, pH = 7.9, 400 mM NaCl) by centrifuging 20 mL. This process was repeated 3 times. Then, sample was loaded into the concentrator and centrifuged at 3000 g until the volume was between 1 mL-2 mL. In order to further reduce the volume, Vivaspin 500 30 KDa MWCO (Millipore), were used. Concentrators were clean and buffer adapted by cleaning with Milli-Q H₂O, centrifuging 1 time 500 μ L of Milli-Q H₂O, and then 3 times 500 μ L of Binding buffer. Sample was transferred to it, affording volumes around 500 μ L.

Finally, sample concentration was checked using the nanodrop. To correct the values of absorbance an ϵ factor of 1.94 mL/(mg · cm) was used.

1.4. DPP IV characterization

1.4.1. Size-exclusion chromatography

Superdex200 HR 10/300(GE Healthcare) was run at ÅKTA Purifier (GE Healthcare) at $T = 4^\circ\text{C}$. Column was cleaned with 1 column volume of Milli-Q H₂O and then equilibrated with

3 column volumes of Binding Buffer (50 mM Tris-HCl, pH = 7.9, 400 mM NaCl). 100 μ L of concentrated sample (between 1-2 mg/mL) were loaded into the column. Sample was collected in eppendorfs in order to perform SDS-PAGE and Blue native-PAGE electrophoresis.

1.4.2. K_M vs DPP IV substrate and DPP IV Specific activity

Enzymatic assays were carried out in 96-well black microplates (Costar). Buffer was 100 mM Sodium/Potassium Phosphate pH = 8.0. Substrate used was H-Gly-Pro-AMC (Bachem) solved in 40% Dioxane/60% H₂O at 3 mM or 750 μ M. To each well, buffer and DPP IV were added, and then, the substrate was added. Final volume was 150 μ L. Fluorescence was read in a BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) and the software used was KC4 (Bio-Tek). Excitation and emission wavelengths were 360 nm and 485 nm respectively.

1.4.2.1. Optimization of the conditions

First, kinetic assays at various DPP IV and substrate concentrations were performed. After addition of substrate, the plate was introduced in the fluorimeter. The temperature was set at T= 37°C and the fluorescence was read from t= 0 to t= 1 h, with a lecture every 5 min and mild shaking before each one.

Data were analyzed in the Excel program (Microsoft) and graphics were done with GraphPad Prism 4 (GraphPad). For each DPP IV concentration, the Michaelis-Menten graphic of fluorescence versus time was plotted. From that graphic, DPP IV concentration for enzymatic assays, and its correspondent linearity was determined. This decision was taken depending on the linearity of the curves. Next, initial velocities correspondent to this DPP IV concentration were plotted versus substrate concentration and K_M value was determined as the substrate concentration that yield half of the maximum initial velocity. Based on that, substrate concentration for enzymatic assays was set to be around the K_M value. Finally, to ensure that the enzyme concentration was in the appropriate range, relative velocities for distinct DPP IV concentrations at the determined substrate value were calculated and plotted versus the enzyme concentration. Enzyme concentration value should be in the linear range.

1.4.2.2. Specific activity

An enzymatic assay at the defined DPP IV and substrate concentrations was set. In the same plate, a standard of AMC was added. Initial velocity was calculated. Specific activity was defined as the number of μ mol of substrate that were cleaved in 1 min by 1 mg of protein.

1.4.3 Mass spectrometry

1.4.3.1. ESI: Denatured DPP IV

Experiments were run at a LCT-Premier XE (Waters-Micromass). A UPLC Acquity (Waters) unit was coupled. Conditions that afforded DPP IV detection were the following:

Column: Biosuite pPhenyl 1000 10 μ m RPC 2.0x75 mm

Gradient: H₂O, 0.1% formic acid - ACN, 0.1% formic acid // 5%-80% in t= 60 min

Flow = 0.1 mL/min

Capillary cone voltage = 2000 V

Sample cone voltage = 50 V

Mode = W optics (enhance resolution)

M/z range = 500/3000

1.4.3.2. MALDI

Conditions used for DPP IV detection by MALDI were the following:

1 μ l 1 mg/mL DPP IV + 1 μ l ACH Matrix

1 μ l 1 mg/mL DPP IV + 1 μ l ACH Matrix + 0.2 μ l Formic acid

0.5 μ l 2 mg/mL DPP IV + 1.5 μ l ACH Matrix + 0.2 μ l Formic acid

1 μ l 0.5 mg/mL DPP IV + 1 μ l ACH Matrix

1 μ l 0.5 mg/mL DPP IV + 1 μ l ACH Matrix + 0.2 μ l Formic acid

0.5 μ l 1 mg/mL DPP IV + 1 μ l ACH Matrix + 0.2 μ l Formic acid

1.4.4. Glycosylation pattern

1.4.4.1. Trypsinization of DTP-1 to DTP-4 samples

For treatments DTP-1 to DTP-4 an eppendorf containing 450 μ g of DPP IV was used. First, sample was dried at SpeedVac. Afterwards, protein was resuspended in 100 μ l of 10mM Ammonium bicarbonate (ABC)/6M Urea.

Then, samples were reduced by adding 5 μ l of 200 mM dithiothreitol (DTT) in 25mM ABC/6M Urea during 1 hour at r.t.

After, protein was alkylated with 4 μ l of 200 mM iodoacetamide (IAA) in 25mM ABC/6M Urea during 1 hour at r.t. and protected from light.

Alkylation was stopped by dilution by adding 20 μ l of 200 mM DTT in 25mM ABC/6M Urea. Before adding trypsin, 775 μ l of H₂O were added to the eppendorf in order to dilute urea to trypsin-compatible concentrations. Then, 20 μ g of trypsin were added to the eppendorf (1:20 ratio). Trypsinization was carried O/N at T = 37°C and 50 rpm.

Next day, samples were cleaned with PD-10 columns, and peptides were collected. To reduce the volume, speed-vac was used, obtaining all trypsinized peptides in a single eppendorf.

1.4.4.2. Enzymatic deglycosylation

The kit E-DEGLY (Sigma) was used for the DPP IV enzymatic deglycosylation. Protocol was the one provided by Sigma.

Dried peptides in the eppendorf were resuspended in 80 μ l of H₂O, 40 μ l of 5x Reaction buffer of E-DEGLY kit and 40 μ l of 5x Reaction buffer of α -fucosidase enzyme. The total of 160 μ l were distributed into 4 eppendorfs, named DTP-1 to DTP-4.

To each one of them 2.5 μ l of Triton X-100 were added. Then, combinations of enzymes were added as following:

| | DTP-1 | DTP-2 | DTP-3 | DTP-4 |
|--|-------|-------|-------|-------|
| 1 μ l PNGase F | Yes | No | No | Yes |
| 1 μ l O-Glycosidase | Yes | Yes | Yes | No |
| 1 μ l α -neuraminidase | Yes | Yes | Yes | Yes |
| 1 μ l β (1-4)-Galactosidase | Yes | Yes | Yes | Yes |
| 1 μ l β -N-Acetylglucosaminidase | Yes | Yes | Yes | Yes |
| 1 μ l α -Fucosidase | Yes | No | No | Yes |
| 2 μ l α -Fucosidase | No | No | Yes | No |

Reactions were carried O/N at T= 37°C and 50 rpm of agitation.

Next day, glycosidases were removed by centrifugation of samples in a 30KDa Vivaspin. Peptides passed through, while enzymes were retained. Volume was reduced by speed-vac. Finally, samples were desalted with Zip-Tip C18 and eluted with H₂O, 0.1% Formic acid/Acetonitrile (ACN), 0.1% Formic acid at 50/50% for their processing by MS.

1.4.4.3. Chemical deglycosylation

Chemical deglycosylation was done with the GlycoProfile IV Chemical Deglycosylation kit (Sigma) which is composed by trifluoromethanesulfonic acid. All treatment was done in a fume hood.

In a reaction vial 500 μ g of DPP IV were freeze-dried. Next day, once sample was totally dried, the vial was placed in a H₂O/ice bath at T = 2-8°C

In the same bath, TFMS and anisole were cooled. Once cold, 140 μ l of TFMS were mixed with 10 μ l of anisole in a reaction vial giving a yellow colour.

Then, the 150 μ l TFMS/anisole mixture was added to the DPP IV dried sample and the vial was softly shaken until the protein was dissolved. Afterwards, the vial was cooled at T = 2-8°C for 3 hours.

Parallel, the 60% Pyridine solution was cooled in a methanol/dry ice bath at T = -15°C, with occasional inversion to avoid solidification.

To the sample vial, 4 μ l of 0.2% Bromophenol Blue solution were added, turning the sample into red colour. Right after, the cooled 60% Pyridine solution was added drop-wised (20 μ l each time). Since reaction is highly exothermic, sample was cooled between

each drop in a methanol/dry ice bath at $T = -20^{\circ}\text{C}$. After adding 40 μl of pyridine, a precipitate was formed. Thus, 20 μl of H_2O were added. The precipitate was also obtained with the two following additions of pyridine and was dissolved in the same manner (by adding H_2O). With every pyridine addition, sample turned from red to palid yellow, then to palid blue, light blue and finally to purple.

Finally, protein was dialyzed against 10 mM ABC O/N at $T = 4^{\circ}\text{C}$.

1.4.4.4. Trypsinization of DTP-5 samples

DTP-5 sample was in 10 mM ABC buffer after dialysis, thus sample was dried at SpeedVac. Afterwards, protein was resuspended in 100 μl of 10mM Ammonium bicarbonate (ABC)/6M Urea.

Then, samples were reduced by adding 5 μl of 200 mM dithiothreitol (DTT) in 25mM ABC/6M Urea during 1 hour at r.t.

After, protein was alkylated with 4 μl of 200 mM iodoacetamide (IAA) in 25mM ABC/6M Urea during 1 hour at r.t. and protected from light.

Alkylation was stopped by dilution by adding 20 μl of 200 mM DTT in 25mM ABC/6M Urea. Before adding trypsin, 775 μl of H_2O were added to the eppendorf in order to dilute urea to trypsin-compatible concentrations. Then, 20 μg of trypsin were added to DTP-5 eppendorf (1:25 ratio). Trypsinization was carried O/N at $T = 37^{\circ}\text{C}$ and 50 rpm.

Next day, sample volume was reduced speed-vac.

Finally, sample was desalted with Zip-Tip C18 and eluted with H_2O , 0.1% Formic acid/Acetonitrile (ACN), 0.1% Formic acid at 50/50% for its processing by MS.

1.4.4.5. Data analysis

Deglycosylated and trypsinized peptides were analyzed in a Fourier transform ion cyclotron resonance instrument (FT-ICR).

Dynamic modifications included in data analysis were the following: (Monoisotopic – average mass)

DTP-1: Asn to Asp

DTP-5: HexNac/N

DTP-4: HexNac/S T

DTP-4: HexNacHex/S T (+365.132)

DTP-3:HexNac (2)/N (+406.158745)

DTP-2: HexNac(2)dHex(1) (+552.216654-552.5262)

DTP-2: HexNac(2)dHex(2) (+698.275)

DTP-2: HexNac(2)dHex(3) (+844.332472-844.8086)

DTP-2: HexNac(2)dHex(4) (+990.390381-990.9498)

1.4.4.6. GluC treatment of DTP samples

Ten μg of DTP were used for each treatment. For DTP-1 to DTP-4 18 μL were placed in separate eppendorfs, while for DTP-5 2.5 μl were enough. Samples were dried O/N at the speed-vac.

Next day, peptides were dissolved in 50 μL of 25 mM ABC pH = 7.8. Then, Glu-C enzyme (Sigma) was added to a final concentration of 1 $\mu\text{g}/\mu\text{l}$. Reaction was led O/N at $T = 25^{\circ}\text{C}$.

1.4.4.7. Data analysis

DTCP samples were processed using the following dynamic modifications:

DTCP-1: Asn to Asp + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(1)dHex(1) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(1)dHex(1) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(1)dHex(2) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(1)dHex(3) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(1)dHex(4) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(2)dHex(1) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(2)dHex(2) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(2)dHex(3) + Met oxidation + Gln to piroglutamate
 DTCP-2: HexNAc(2)dHex(4) + Met oxidation + Gln to piroglutamate
 DTCP-2: Hex (1)HexNAc(2)dHex(2) + Met oxidation + Gln to piroglutamate
 DTCP-2: Hex (2)HexNAc(2) + Met oxidation + Gln to piroglutamate
 DTCP-5: HexNac/N + Met oxidation + Gln to piroglutamate
 DTCP-4: HexNacHex/S T + Met oxidation + Gln to piroglutamate
 DTCP-4: HexNacHex/S T + Hex/S T + Met oxidation + Gln to piroglutamate
 DTCP-4: HexNacHex/S T + HexNac/S T + Met oxidation + Gln to piroglutamate
 DTCP-3:HexNAc (2)/N + Met oxidation + Gln to piroglutamate

1.5. Methyl-¹³C Methionine selective labeling and NMR experiments

1.5.1. Methyl-¹³C Methionine labeling

Sf9 cells were grown in Sf-900 III SFM until enough cells were obtained (around 15×10^8 cells). Cultures were centrifuged at 500 g for 5 min. Meanwhile, 500 mL of Sf-900 II SFM, no methionine, no cysteine (Invitrogen) were supplemented with 1 g/L methyl-¹³C Methionine (Cambridge Isotope Laboratories) and 0.150 g/l L-cystine (Sigma Aldrich). Pre-warming of the media and agitation with a magnetic stirrer were required in order to totally solubilize the aminoacids. After culture centrifugation, supernatant was discarded and pellets were resuspended in the labelled media giving a total of 500 mL at 3×10^6 cells/mL. Then, each flask of 250 mL of cells was infected with 5 mL of P3 Baculovirus stock. After 4 days of incubation at 27°C and 150 rpm, cell culture was centrifuged at 1000 g for $t = 10$ min. Treatment of supernatant and protein purification were performed following the same protocol as in the non-labelled DPP IV.

Imidazole removal was achieved by using a HiPrep 26/10 Desalting column (GE Healthcare) in the ÄKTA Purifier at $T = 4^\circ\text{C}$. Column was equilibrated in Binding Buffer (50 mM Tris-HCl, pH = 7.9, 400 mM NaCl) and then, selected elution tubes from the Nickel column were injected.

Protein elution peak was concentrated using the Vivaspin 20 30 KDa MWCO (Millipore). First, the membranes were cleaned with Milli-Q H₂O and then, 20 mL of Milli-Q H₂O were centrifuged at 3000 g. This process was repeated a total of 5 times. Next, membrane was adapted in DPP IV NMR buffer: 50 mM D11 Tris-HCl, pH = 7.9 (Cambridge Isotope Laboratories), 50 mM NaCl; 90% D₂O/10% H₂O; by centrifuging 20 mL. This process was repeated 5 times. Then, sample was loaded into the concentrator and centrifuged at 3000 g until the volume was between 1 mL-2 mL. Following that, 15 mL of DPP IV NMR Buffer were added to the concentrator and centrifuged again. This process was repeated 5 times. In order to further reduce the volume, Vivaspin 500 30 KDa MWCO (Millipore), were used. Concentrators were clean and buffer adapted by cleaning with Milli-Q H₂O, centrifuging 5 time 500 µL of Milli-Q H₂O, and then 5 times 500 µL of DPP IV NMR buffer. Sample was transferred to it, affording volumes around 300 µL and concentrations around 100 µM. Finally, sample concentration was checked using the nanodrop. To correct the values of absorbance an ϵ factor of 1.94 mL/(mg · cm) was used. Sample was finally transferred into a Shigemi tube and the piston was positioned avoiding the appearance of bubbles in the sample.

1.5.2. Methyl-TROSY experiment

Experiments were performed at 600 and 800 MHz Bruker Digital Advance NMR instruments fitted with triple-resonance z-axis gradient cryoprobes. Sequence used was a ¹H-¹³C- HMQC^[170] and data were acquired at T = 35°C for 35 h. 1-D proton NMR spectra were done before and after each HMQC in order to evaluate sample integrity. Data were processed with Topsin 2.0.

1.5.3. Signal assignment: Site-directed mutagenesis

Methionine residues for mutagenesis were selected on basis of their proximity to the active site centre of DPP IV. Methionine codon ATG was modified to CTG, corresponding to Leucine.

Forward and reverse primers were designed with the QuickChange Primer Design (Agilent Technologies) and were bought at Sigma.

Sequences were as follow:

M425L FP: TAATGAATATAAAGGACTGCCAGGAGGAAGGAA

Length: 33 nucleotides; T_m = 60.7°C; MW = 10322; GC% = 39.4

M425L RP: TTCCTTCCTCCTGGCAGTCCTTTATATTCATTA

Length: 33 nucleotides; T_m = 60.7°C; MW = 10027.5; GC% = 39.4

M503L FP: ATTCAGCTTTGGATAAACTGCTGCAGAATGTC

Length: 32 nucleotides; T_m = 60.5°C; MW = 9919.475; GC% = 40.6

M503L RP: GACATTCTGCAGCAGTTTATCCAAAGCTGAAT

Length: 32 nucleotides; T_m = 60.5°C; MW = 9888.465; GC% = 40.6

M509L FP: CAGAATGTCCAGCTGCCCTCCAAAAACTG

Length: 30 nucleotides; T_m = 63°C; MW = 9209.965; GC% = 50

M509L RP: CAGTTTTTTGGAGGGCAGCTGGACATTCTG

Length: 30 nucleotides; T_m = 63°C; MW = 9365.015; GC% = 50

M591L FP: AAGGAGATAAGATCCTGCATGCAATCAACAG

Length: 31 nucleotides; T_m = 60.4°C; MW = 9651.33; GC% = 41.9

M591L RP: CTGTTGATTGCATGCAGGATCTTATCTCCTT

Length: 31 nucleotides; T_m = 60.4°C; MW = 9539.14; GC% = 41.9

M671L FP: GAACGTTACCTGGGTCTCCCAACTC

Length: 25 nucleotides; T_m = 61°C; MW = 7658.75; GC% = 56

M671L RP: GAGTTGGGAGACCCAGGTAACGTTC

Length: 25 nucleotides; T_m = 61°C; MW = 7827.89; GC% = 56

1.5.3.1. Gene fragment amplification

For each mutant, two PCR reactions were set up. One amplified the 5' extreme of the DPP IV gene until the mutation point (Reaction a) and the second from the mutation until the 3' extreme (Reaction b).

PCR conditions were as follow: (Total Volume = 50 µL)

Mix 1

- Primers (5 µM) : 5 µL each
 - Reaction a = FP wt + RP mut
 - Reaction b = FP mut + RP wt
- dNTP's (10 mM): 1 µL
- Template: 2 µL (100 ng/µL)
- H₂O: 12 µL

Mix 2 (For each reaction tube)

- H₂O: 19.75 µL
- PCR Buffer 10x : 5 µL
- Taq Polymerase (1U/ µL): 0.25 µL

PCR reaction was carried out in a Thermocycler (MiniCycler, MJ Research), with the conditions:

- 94°C 2 min
 - 94 °C 30 s
 - 58°C 60 s
 - 72°C 3 min
 - 72°C 7 min
- } 30 cycles

Then, 1% agarose gel was loaded with the PCR reactions samples. Bands were excised and DNA was purified using the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” from GE Healthcare. Each band was eluted with 50 µL of TAE buffer.

1.5.3.2. Mutant DPP IV gene amplification

Second PCR was prepared using as templates the PCR results purified in the previous step.

PCR conditions were as follow: (Total Volume = 50 µL)

- Primers (5 µM) : 5 µL each
- dNTP's (10 mM): 1 µL
- Templates: 10 µL each
- PCR Buffer 10x : 5 µL
- H₂O: 13.75 µL
- Taq Polymerase (1U/ µL): 0.25 µL

PCR reaction was carried out in a Thermocycler (MiniCycler, MJ Research), with the following conditions:

- 94°C 2 min
 - 94 °C 30 s
 - 58°C 60 s
 - 72°C 3 min
 - 72°C 7 min
- } 30 cycles

Then, 1% agarose gel was loaded with the PCR reactions samples. Bands were excised and DNA was purified using the “Illustra™ GFX PCR DNA and Gel Band Purification Kit” from GE Healthcare. Each DPP IV gene corresponding band was purified with 50 µL of TAE buffer.

1.5.3.3. Mutant DPP IV gene ligation in pGEM-T Easy and pFastBac EGT-C

Mutant DPP IV genes were ligated into pGEM-T Easy following the same procedure applied for the wild type DPP IV gene.

Further mutant DPP IV gene insertion in pFastBac EGT-C was also carried out following the steps of the wild type DPP IV gene.

1.5.3.4. Mutant DPP IV bacmid and virus generation

For each construct, 50 µL of MAX Efficiency ® DH10Bac™ competent cells were aliquoted in a pre-culture tube. Then, 5 µL of mutant DPP IV-pFastBac EGT-C were added to the

labelled tube. Samples were mixed by finger-flicking and incubated on ice during $t = 30$ min.

Right after, cells were heat-shocked at $T = 42^{\circ}\text{C}$ during 45 seconds. Next, tubes were cooled on ice for 2 minutes. Finally, 500 μL of SOC media were added and cells were grown at $T = 37^{\circ}\text{C}$ during $t = 4$ hour.

After that, 100 μL of a 1:100 dilution of the cell culture, were plated on LB Agar plates supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin, 10 $\mu\text{g}/\text{mL}$ gentamicin, 50 $\mu\text{g}/\text{mL}$ X-Gal and 0.2 mM IPTG. Plates (one for each DPP IV construct) were incubated at $T = 37^{\circ}\text{C}$ O/N.

Next day, plates were checked for white colonies. For each construct, several pre-culture tubes of 15 mL were prepared with 2 mL LB cultures supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin and 7 $\mu\text{g}/\text{mL}$ gentamicin. Then, media-containing tubes were inoculated with picked single white colonies. Tubes were grown at $T = 37^{\circ}\text{C}$ during $t = \text{O}/\text{N}$ and constant agitation of 250 rpm.

Mutant DPP IV recombinant bacmids were purified following the same procedure as in the wild type DPP IV recombinant bacmid.

1.5.4. Inhibitor-DPP IV Interaction

1.5.4.1. NMR Experiments

NMR experiments were carried out in 600 and 800 MHz Bruker Digital Avance NMR instruments fitted with triple-resonance z-axis gradient cryoprobes. Data were acquired at $T = 35^{\circ}\text{C}$ for 35 h.

DPP IV was at concentrations around 100 μM in NMR Buffer: 50 mM D11 Tris-HCl, pH = 7.9 (Cambridge Isotope Laboratories), 50 mM NaCl; 90% $\text{D}_2\text{O}/10\%$ H_2O . Sample was introduced in a Shigemi tube, and a control $^1\text{H}-^{13}\text{C}$ - HMQC [1] experiment was performed. Then, 10 equivalents of inhibitor were added to the tube and the $^1\text{H}-^{13}\text{C}$ - HMQC [1] experiment was repeated. 1-D proton NMR spectra were done before and after each HMQC in order to evaluate sample integrity.

Data were processed with Topspin 2.0.

Chapter 2. Discovery of DPP IV inhibitors from plant extracts

2.1. Selection, extraction and test

2.1.1 Selection

2.1.1.1. Common antidiabetic plants

Selection of plants was based on bibliography. Plants that were described to possess antidiabetic properties were selected. Then, dried plants were acquired at herbalist stores in Barcelona (Spain). Also fresh plants were bought at local stores.

2.1.1.2. Traditional Chinese Medicinal plants

Selection of plants was based in the book of Chinese Herbal Medicine^[180] which enabled the understanding of the relationship of traditional with modern Chinese medicine. Other bibliography sources were used as well. ^[179, 181] Plants that were reported as antidiabetic agents were selected. One hundred grams of each plant were then acquired at Dasherb in dried format.

2.1.1.3. Brazilian plants

Selection of antidiabetic properties possessing plants was based in the books of Brazilian Traditional Medicine ^[186, 187]. Dried plants were acquired at herbalist stores in Santa Maria (Rio Grande do Sul, Brazil), (Aroma Benestar, Flor do campo, Chá Dermapelle). Besides, fresh plant was collected in the area of Caçapava (Rio Grande do Sul, Brazil). In this case, plant parts (leaves, fruits, roots or bark) were deposit in a newspaper covered tray. The tray was left in the oven at 150°C for 2-3 days until the plant part was completely dried. Afterwards, the dried plant was grilled.

2.1.1.4. Mediterranean plants

In this case, there was not a previous selection of plants versus antidiabetic properties.

2.1.2. Extraction

2.1.2.1. Aqueous extraction

Aqueous extracts were performed using Soxhlet extraction. Thirty grams of dried plant were grinded and introduced in a cellulose cartridge. Then, 250 mL of Milli-Q H₂O were distributed between the balloon and the Soxhlet. Temperature was set at 140°C and the

system was left O/N with water refrigeration. All equipment was protected with aluminum foil in order to protect the system from temperature dropping. Next day, aqueous extract was divided into 50 mL falcon tubes, which were then frozen with liquid N₂ and freeze-dried.

2.1.2.2. Ethanolic maceration

Ethanolic macerations were afforded by incubation of 20-30 grams of grinded plant together with 250-300 mL of pure ethanol (Panreac). Macerations were left at r.t. during 30 days, with occasional hand agitation every 3-5 days. After that, extracts were solvent evaporated. Then, a mixture of 50:50 H₂O/Acetonitrile was loaded into the balloon in order to remove all the plant extract. This solution was transferred to falcon tubes, which were then frozen with liquid N₂ and freeze-dried.

2.1.2.3. Hydroalcoholic extraction

Extraction was performed using 70% Ethanol / 30% H₂O and using Soxhlet. Various grams of dried plant were introduced in a paper cone which was placed inside the Soxhlet apparatus. Between 250-300 mL of solvent were distributed between the balloon and the Soxhlet. Temperature was set at 140°C and the system was left for 5 hours with water refrigeration. All equipment was protected with aluminum foil in order to protect the system from temperature dropping. After that, the extract was removed and kept at T = 4°C. Fresh 250-300 mL of solvent were added and left O/N without heating. In the morning, the system was heated for 5 more hours. Finally, extract was removed and together with the first batch of extract (making a total of 500-600 mL), it was solvent evaporated. Then, H₂O was loaded into the balloon in order to remove all the plant extract. This solution was transferred to freeze-drying pots, which were then frozen with liquid N₂ and freeze-dried.

2.1.3. Plant extracts screening

For each extract a stock was done. For aqueous extracts the solvent was Milli-Q H₂O, while for ethanolic and hydroalcoholic extracts Dimethyl sulfoxide (DMSO) (Panreac) was used. When needed, sonication helped to solubilize.

Stock concentrations were 25 mg/mL for TCM and BP extracts; and 5 mg/mL for common antidiabetic and Mediterranean plants.

Activity assay was performed in a 96-well black microplate (Costar). Buffer was 100 mM Sodium/Potassium Phosphate pH = 8.0. Substrate used was H-Gly-Pro-AMC (Bachem) solved in 40% Dioxanne/60% H₂O at 750 μM. DPP IV concentration varied from stock to stock. For each assay, there were 4 types of wells: positive wells (buffer, protein, DMSO and substrate), negative controls (buffer, DMSO and substrate), plant extracts (buffer, protein, plant extract and substrate) and negative of the samples (buffer, plant extract and substrate). Buffer and protein were added. After that, 3 μL of DMSO or stock plant extract was added to its correspondent wells. A pre-incubation of 15 min at T= 37°C and constant agitation of 90rpm was performed. Then, 10 μL of 750 μM substrate was added to all the wells, yielding a final concentration of 50μM. Final volume was 150 μL. Plates were

incubated for 20 min at T= 37°C and constant agitation of 90rpm. After that, fluorescence was read in a BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) with the software KC4 (Bio-Tek). Excitation and emission wavelengths were 360 nm and 485 nm respectively. Calculations were done with Excel (Microsoft) and for the representations, GraphPad Prism 4 (GraphPad).

2.2. *Coutarea Latiflora*

2.2.1. Extract fractionation and compound isolation

Coutarea Latiflora ethanolic extract was fractionated by means of Preparative HPLC (Waters). The column used was a Waters 2545 Symmetry C18 of 3.5 µm diameter; 7.8 x 100 mm. Purification was performed at flow: 3ml/min with a gradient of 10-40% (H₂O, 0.1 % TFA – ACN, 0.1% TFA). Sample injected (copalchi extract) was at 20 mg/mL and each injection was of 500 µL.

Purified fractions were analysed by means of analitic HPLC (Waters). The column used was a Waters 2695 Sunfire C18 of 3.5 µm diameter; 4.6 x 100 mm. Conditions used were flow at 1ml/min and gradient of 15-40% (H₂O, 0.045 % TFA – ACN, 0.036 % TFA)

2.2.2. Molecule characterization

Structures were elucidated by Albert Puigpinós while he was a technician at IRB Barcelona. He did a series of NMR spectra (+H, DEPT, COSY, HSQC and HMBC).

2.2.2.1. AP-3-a

MW: 448 [M+H]: 449.10784 C₂₁H₂₁O₁₁

¹H NMR (N,N-Dimethylformamide-d₇, 500 MHz) δ (ppm): 10.95 (s, 1H) 21, 9.28 (s, 1H) 19, 9.11 (s, 1H) 18, 6.91 (d, 1H) 12, 6.85 (d, 1H) 16, 6.76 (dd, 1H) 15, 6.64 (d, 1H) 2, 6.53 (d, 1H) 6, 5.84 (s, 1H) 8, 4.83 (d, 1H) 22, 3.83 (dd, 1H) 31, 3.66 (dd, 1H) 31, 3.41 (m, 1H) 24, 3.34 (m, 1H) 26, 3.28 (m, 1H) 25, 2.78 (m, 1H) 27.

¹³C NMR (N,N-Dimethylformamide-d₇, 500 MHz) δ (ppm): 162.54 1, 160.13 9, 157.19 5, 156.75 13, 156.30 3, 146.52 14, 144.79 11, 131.34 7, 119.64 15, 116.87 12, 114.84 16, 111.67 8, 102.96 4, 100.74 22, 99.35 2, 97.20 6, 77.77 24, 77.13 26, 73.76 27, 70.25 25, 61.57 31.

2.2.2.2. AP-3-b

MW: 462 [M+H]: 463.12353 C₂₂H₂₃O₁₁

¹H NMR (N,N-Dimethylformamide-d₇, 500 MHz) δ (ppm): 9.3 ppm (m, 2H) 18 19 , 6.93 (m, 1H) 12, 6.87 (m, 1H) 16, 6.79 (m, 1H) 15, 6.77 (m, 1H) 2, 6.68 (m, 1H) 6, 5.91 (d, 1H) 8,

3.93 (d, 3H) 33, 3.87 (m, 1H) 31, 3.77 (m, 1H) 31, 3.48 (m, 1H) 24, 3.34 (m, 1H) 26, 3.23 (m, 1H) 25, 2.78 (m, 1H) 27.

¹³C NMR (N,N-Dimethylformamide-d₇, 500 MHz) δ (ppm): 163.48 1, 163.42 1, 160.0 9, 159.94 9, 157.05 5, 156.97 5, 156.07 3, 156.03 3, 131.28 7, 131.13 7, 116.0 12, 114.94 16, 114.86 16, 112.80 8, 112.57 8, 104.0 4, 103.91 4, 101.94 22, 100.72 22, 99.81 2, 99.51 2, 95.67 6, 95.64 6, 77.90 24, 77.20 26, 73.86 27, 71.14 25, 63.60 31, 55.80 33, 55.74 33.

2.2.3. Inhibitor characterization

2.2.3.1. Study of the inhibition

AP-3-a and AP-3-b IC₅₀ values were calculated by enzymatic assays. Concentrations of molecules ranged from 10 μ M to 1 mM. Assays were performed following the protocol of “2.1.3 Plant extracts screening”. Data calculations were done with Excel (Microsoft) and graphs were plotted with GraphPad Prism 4 (GraphPad).

Aggregation studies were performed by addition of 0.01% of Triton (Fluka).

Reversibility of AP-3-a was evaluated by modifying the preincubation times of enzyme with molecule (time-course experiment). Times were 0, 15, 30 and 60 minutes. After that time, substrate was added, incubation of 20 min was done and fluorescence values were read.

Aglycone of AP-3-a was obtained by hydrolysis. Two mg of molecule were placed in a balloon together with 10 mL of 2M TFA. It was then heated for 1 hour at T = 120°C. After, it was let to cool and TFA was evaporated under pressure at T = 40°C. Finally, it was frozen with liquid N₂ and freeze-dried.

2.2.3.2. Inhibition mechanism

Kinetic experiments were performed. For Copalchi A, concentrations used were; 0, 10, 50, 100, 200 and 400 μ M. Pre-incubation with enzyme was of 15 min. Substrate concentrations were: 20, 30, 40 and 50 μ M. Final incubation was performed with the plate inside the BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) fluorimeter in order to monitor the full-length reaction. Fluorimeter temperature was set to T = 37°C and shaking was performed before each lecture, which was every 5 min from 0 to 60 min. Excitation and emission wavelengths were 360 nm and 485 nm respectively. Calculations were done with Excel (Microsoft) and for the representations, GraphPad Prism 4 (GraphPad). Lineweaver-Burk and Dixon plot representations were done and then, secondary plots were obtained by representing the slopes and intercepts of Lineweaver-Burk versus inhibitor concentration. Five equations were obtained from the slope plot and the factor relationships from enzymatic reaction. By an iteration process, the six unknown values were solved.

2.2.3.4. NMR Analysis

AP-3-a was analyzed by NMR, following the same procedure as in “1.5.2 Methyl-TROSY experiment”.

Chapter 3. Discovery of POP inhibitors by High Throughput Screening

3.1. Set-up Fluorescence Polarization Experiment

3.1.1. Probe synthesis

Synthesis of H-(Ahx)₃-Arg-(Ahx)₂-Pro-Pro-OH (cold peptide) was performed as previously described.^[202] The solid-phase peptide synthesis was done following the Fmoc/^tBu strategy and using PyBOP/HOAt/DIEA reagents for the aminoacid coupling. Resin was 2-chlorotritil (f = 1.5 mmol/g).

Purification of cold peptide was performed in an HPLC (Waters) using the Sunfire C18 column 10x150 mm 5 μ m with a gradient of 10-40 % (H₂O-ACN) during t= 15 min.

Coupling of carboxyfluorescein to the resin-immobilized peptide was performed by a two-step process. First reaction consisted on 3 eq. of carboxyfluorescein, 3 eq. of PyBOP, 9 eq. of HOAt and 9 eq. of DIEA in DMF for 1 hour at r.t. Reaction was done twice. Afterwards, second reaction was performed: 3 eq. of carboxyfluorescein, 3 eq. of DPCDI 9 eq. of HOAt in DMF:DCM (1:1) during O/N. at r.t.

Carboxyfluorescein-labeled probe was purified in an HPLC (Waters) using the Sunfire C18 column 10x150 mm 5 μ m with a gradient of 10-70 % (H₂O-ACN) during t= 15 min.

TAMRA coupling was performed in the same way as carboxyfluorescein with one exception: First reaction was one during 1.5 hours at r.t. and was not repeated.

TAMRA-labeled probe was purified in an HPLC (Waters) using the Sunfire C18 column 10x150 mm 5 μ m with a gradient of 25-35 % (H₂O-ACN) during t= 15 min.

3.1.2. FP assay optimization

First, probes were assayed for binding to POP.

Experiments were carried out in 384 black-well plates (Corning). In the last well of a row (number 24), 60 μ L of the mixture 1 μ M POP/50 nM probe were added. In wells 1 to 23, 30 μ L of 50 nM probe were added. Then, 30 μ L of well 24 were removed and added to well 23. After pipet mixing, 30 μ L of well 23 were transferred to well 22, and so on, until well 2. The 30 μ L that were removed from this well were discarded.

As a result, well 24 had a POP concentration of 1 μ M, well 23 was 500 nM, well 2 was 250 nM and so on. But probe concentration remained constant.

After plate preparation, FP and TF were read Envision 1 and Envision 2 fluorescence plate readers at the appropriate wavelengths for carboxyfluorescein (λ_{Abs} = 494 nm; λ_{Em} = 519 nm) and TAMRA (λ_{Abs} = 541 nm; λ_{Em} = 565 nm).

In the case of Carboxyfluorescein, the experiment was repeated using 30 nM concentration of probe. Calculations were done with Excel (Microsoft) and data was represented with GraphPad Prism 4 (GraphPad).

For the confidence assay, half of a 384 black-well plate was filled with 30 μ L/well of the mixture POP+ probe (1 μ M POP/30 nM carboxyfluorescein-probe or 50 nM TAMRA-probe), and the other half with 30 μ L/well of 30 nM carboxyfluorescein-probe or 50 nM TAMRA-probe. Then, plate was read for FP and TF at the appropriate wavelengths for carboxyfluorescein ($\lambda_{Abs} = 494$ nm; $\lambda_{Em} = 519$ nm) and TAMRA ($\lambda_{Abs} = 541$ nm; $\lambda_{Em} = 565$ nm). Data was represented with GraphPad Prism 4 (GraphPad). Calculations were done with Excel (Microsoft) and z-score was calculated with the formula:

$$z\text{-score} = 1 - \frac{(3 \times SD_{\text{free probe}} + 3 \times SD_{\text{complex}})}{(FP_{\text{complex}} - FP_{\text{free probe}})}$$

Competition experiments of probe with cold peptide and ZPP (Biomol) were performed in 384 black-well plates (Corning). In the last well of a row (number 24), 60 μ L of the mixture 1 μ M POP/30 or 50 nM probe/100 μ M competitor were added. In wells 1 to 23, 30 μ L of 1 μ M POP/30 nM or 50 probe were added. (30 nM carboxyfluorescein-probe and 50 nM TAMRA-probe).

Then, 30 μ L of well 24 were removed and added to well 23. After pipet mixing, 30 μ L of well 23 were transferred to well 22, and so on, until well 2. The 30 μ L that were removed from this well were discarded.

As a result, well 24 had a competitor concentration of 100 μ M, well 23 was 50 μ M, well 2 was 25 μ M and so on. But POP and probe concentrations remained constant.

After plate preparation, FP and TF were read Envision 1 and Envision 2 fluorescence plate readers at the appropriate wavelengths for carboxyfluorescein ($\lambda_{Abs} = 494$ nm; $\lambda_{Em} = 519$ nm) and TAMRA ($\lambda_{Abs} = 541$ nm; $\lambda_{Em} = 565$ nm).

Calculations were done with Excel (Microsoft) and data was represented with GraphPad Prism 4 (GraphPad).

3.2. High throughput screening

Stocks for POP and probes were prepared at the double of the final concentration. Same volume for each solution (protein and probe) was mixed, and filtered through 0.2 μ m, affording final concentrations of 1 μ M_{POP} and 30 nM_{Carboxyfluorescein probe} or 50 nM_{TAMRA probe}. Then, in each well of the 384 black-well plates, 30 μ L of the POP-probe mixture was added. Next, a 100 nL pin transfer of library compounds was done. Thus, final tested concentrations of compounds were 33 μ M (if 10 mM stock) or 6.67 μ g/mL (if 2 mg/mL stock). To have an approximate comparison, if 350 g/mol was considered as a good standard for small molecule weight, then 6.67 μ g/mL would equal to 19 μ M.

Finally, plates were read at Envision 1 and Envision 2 fluorescence plate readers at the appropriate wavelengths for carboxyfluorescein and TAMRA.

3.3. Hit validation

All molecular docking and clustering calculations were done with Schrödinger software.

3.3.1. Molecular docking

Initial coordinates for the closed POP structures were taken from the Protein Data Bank (pdb) entries 1QFS and 2XDW (porcines with covalent inhibitor), 1QFM (porcine without inhibitor) and 3DDU (human with non-covalent inhibitor). Hydrogen atoms and titratable side chains were optimized with the Protein Preparation Wizard tool at physiological pH. The grid was centered on this catalytic active region of POP. All 75 ligands were prepared with LigPrep default protocol and subsequent generation of conformers with ConfGen. Thus 1238 different conformations were obtained and docked to the four protein structures with Glide XP precision. The best score results were received for the human POP structure.

3.3.2. Molecular clustering

All structural and data analysis (including fingerprints, similarity searching, substructure searching, selection by diversity and clustering) was done by Canvas – a cheminformatics package of Schrödinger that provides a range of applications like Scaffold Decomposition, Maximum common substructure and R-Group analysis (all described in the Canvas User Manual). We tried many different ways of searching similarities and trying to separate 75 compounds in a few groups.

3.4. Kinetic studies for candidates

Kinetic experiments were performed. For HTS-43, concentrations used were; 0, 1, 5, 10, and 25 μM . Pre-incubation with enzyme was of 15 min. Substrate concentrations were: 10, 15, 20 and 30 μM . Final incubation was performed with the plate inside the BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) fluorimeter in order to monitor the full-length reaction. Fluorimeter temperature was set to $T = 37^\circ\text{C}$ and shaking was performed before each lecture, which was every 5 min from 0 to 60 min. Excitation and emission wavelengths were 360 nm and 485 nm respectively. Calculations were done with Excel (Microsoft) and the representations of Lineweaver-Burk and Dixon-Plot were done with GraphPad Prism 4 (GraphPad).

For HTS-75, concentrations used were; 0, 0.1, 1, 5, 10 and 25 μM . Experiments conditions were the same as for HTS-43. Calculations were done with Excel (Microsoft) and the representations were done with GraphPad Prism 4 (GraphPad). Lineweaver-Burk and Dixon plot representations were done and then, secondary plots were obtained by representing the slopes and intercepts of Lineweaver-Burk versus inhibitor concentration. Five equations were obtained from the slope plot and the factor relationships from enzymatic reaction. By an iteration process, the six unknown values were solved.

Chapter 4. Characterization of Peptidyl Aryl Vinyl Sulfones as cathepsins L and B inhibitors

4.1. Library screening

Cathepsin L (3.4.22.15) and cathepsin B (3.4.22.1), both from human liver were purchased from Sigma–Aldrich. Enzymatic assays were performed in 96-well black microplates (Costar). Cathepsin L was at a final concentration of 70 ng/mL in the activity buffer: 100 mM NaOAc, containing 0.01% Triton X-100, 5 mM EDTA and 5 mM DTT, pH 5.5 (acetic acid). Cathepsin B was at 6.65 ng/mL final concentration in the activity buffer: 100 mM sodium phosphate, pH 6.2, 1 mM EDTA and 1 mM DTT. Compound stock solutions were prepared at 5 mM or 25 mM in DMSO. Dilutions from the stock solution were used in order to perform the inhibition curves. The mixture was pre-incubated for 15 min at 37°C and constant agitation of 90 rpm. Next, the substrate was added. Cathepsin L substrate used in experiments was Z-Phe-Arg-AMC, and cathepsin B substrate was Z-Arg-Arg-AMC, (both of them from Bachem). Plates were then incubated for 30 min at 37°C and constant agitation of 90 rpm. The reaction was stopped by addition of 150 µL of 1 mM NaOAc (pH 4) to each well. The plate was read in the BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) at $\lambda_{\text{excitation}}=360$ and $\lambda_{\text{emission}}=485\text{nm}$.

Selectivity of selected compounds was assayed against POP and DPP IV.

DPP IV (3.4.14.5) from porcine kidney was purchased from Sigma–Aldrich and POP was obtained by recombinant expression in *E. coli* previously described. [221] Enzymatic assays were performed in 96-well black microplates (Costar). POP final concentration was 10.6 ng/mL and DPP IV final concentration was 0.187 ng/mL. Buffer for both proteins was 100 mM sodium/potassium phosphate, pH 8. After the addition of buffer, enzyme and compound, the mixture was pre-incubated for 15 min at 37°C. Next, the substrate was added. In the case of POP, the substrate was Z-Gly-Pro-AMC, and DPP IV substrate was H-Gly-Pro-AMC, both from Bachem. After addition, plates were incubated for 60 min at 37°C and 90 rpm. Finally, the reaction was stopped by the addition of 150 µL of 1 mM NaOAc (pH 4) to each well. The plate was read in the BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) at $\lambda_{\text{excitation}}=360$ and $\lambda_{\text{emission}}=485\text{nm}$.

4.2. Elucidation of the inhibition mechanism

Time-course experiments were performed for PAVS-20 against cathepsin L activity. Pre-incubation was done for 0, 5, 10, 15, 30, 60, 90, and 120 min. After pre-incubation, substrate was added, and plates were incubated at 37°C for 30 min. Following this incubation, the reaction was stopped and fluorescence readings were read.

Lineweaver-Burk representation was obtained after performing a kinetic experiment. PAVS-20 concentrations were: 0, 2, 5, and 10 nM and substrate concentrations were: 20, 50, 100, 150, 200 and 800 μ M. Pre-incubation with enzyme was of 15 min. Final incubation was performed with the plate inside the BIO-TEK FL600 Microplate Fluorescence Reader (Bio-Tek Instruments) fluorimeter in order to monitor the full-length reaction. Fluorimeter temperature was set to $T = 37^{\circ}\text{C}$ and shaking was performed before each lecture, which was every 5 min from 0 to 30 min. Excitation and emission wavelengths were 360 nm and 485 nm respectively. Calculations were done with Excel (Microsoft) and for the representations, GraphPad Prism 4 (GraphPad).

4.3. Docking

Compounds were drawn using MarvinSketch and translated to three-dimensional viewing with MarvinView. For Cathepsin L, the X-ray crystal structure corresponding to PDB ID: 1CS8 (1.80 resolution). For cathepsin B, the X-ray crystal structure corresponding to PDB ID: 1HUC (2.10 resolution) was used. Water molecules and propeptide (only present in 1CS8) were previously removed from the file.

Docking calculations were performed with AutoDock version 4.^[222] The protein was set as rigid while the compounds were defined as flexible. A grid map of 60 points with a point spacing of 0.3750 (generated using AutoGrid, version 4) was placed in the active center of the enzyme, using AutoDock tools. For each compound, 100 docking runs were carried out using 50 individuals and the Lamarckian genetic algorithm. After calculations, the results were evaluated, and the best-docked conformations were selected for further analysis

RESUMEN

INTRODUCCIÓN

La siguiente introducción está dividida en cuatro secciones:

1. Las proteasas
2. Las proteasas cómo dianas terapéuticas
3. Los inhibidores de proteasas cómo agentes terapéuticos
4. Casos de estudio: DPP IV, POP y catepsinas L y B

1. Las proteasas

Las proteasas son enzimas que liberan amino ácidos, péptidos o proteínas de péptidos o proteínas mayores. [1]

1.1. Clasificación

Las proteasas se pueden clasificar en base a una serie de criterios.

Dependiendo del tipo de sustrato, las proteasas se clasifican en exo-proteasas o endo-proteasas. En base al mecanismo catalítico, estas enzimas se dividen en seis grupos: serina-, cisteína-, treonina-, aspártico-, metalo- y glutámico- proteasas

Por último, dependiendo del origen evolutivo, proteasas homólogas, bien en secuencia aminoácida o en estructura tridimensional, se agrupan en familias siempre y cuando compartan el mismo mecanismo catalítico. A su vez, el conjunto de familias de proteasas con un mismo origen evolutivo se denomina clan. En este caso, la homología entre los miembros ha de existir, pero el mecanismo catalítico puede diferir.[2]

1.2. Sustratos de proteasas

Los sustratos endógenos son aquellos que, además de contener una secuencia de reconocimiento para la enzima, son degradados por la enzima en ensayos *in vitro* y coexisten espacial y temporalmente *in vivo* con la proteasa activa.

La identificación del degradoma de una proteasa es una tarea ardua.

Tradicionalmente, en vez de localizar los sustratos de una enzima, se realizaba el proceso contrario: dado un sustrato se localizaba la proteasa responsable de su procesamiento. Los sustratos de la enzima convertidora de angiotensina (ACE) fueron descubiertos mediante este método.[6]

Hoy en día se dispone de una batería de técnicas para la identificación de los sustratos de las proteasas, incluyendo herramientas bioinformáticas,[7] el doble híbrido de levadura,[8]

técnicas genéticas (cómo “knockin”, “knockout” o RNA de interferencia)^[9] o la proteómica (espectrometría de masas).^[3]

1.3. Funciones y regulación de las proteasas

Básicamente, las proteasas pueden degradar proteínas o procesarlas.

La primera función es llevada a cabo por proteasas que son poco específicas y están localizadas en compartimentos subcelulares.^[5]

En cambio, el procesamiento de proteínas es realizado por proteasas muy específicas y que suelen compartir el mismo espacio subcelular.^[5] Como resultado, la función del sustrato se modifica (se activa, inactiva o altera). De esta forma las proteasas pueden regular ciclos biológicos, activando o desactivando dicha ruta. La habilidad de las proteasas para regular estos procesos biológicos se denomina señalización de proteasas. A diferencia de otro tipo de señalizaciones, éste es irreversible.

Por este motivo, las proteasas están altamente reguladas, desde el nivel de transcripción hasta la inhibición de la enzima activa.

2. Las proteasas cómo dianas terapéuticas

Las proteasas participan en una multitud de procesos biológicos cruciales y por eso, cualquier fallo en su actividad o expresión es causa de rutas y ciclos aberrantes. Así, las proteasas están involucradas en un gran número de enfermedades.

2.1. Identificación de la diana

En enfermedades como el cáncer o enfermedades neurodegenerativas, se ha observado una proteólisis no adecuada, tanto inferior cómo superior al control.

En otras patologías, la actividad o expresión de la proteasa pueden ser correctas. Sin embargo, dicha enzima puede estar involucrada en ciclos biológicos desregulados. Así, la modificación de la proteasa mediante agentes terapéuticos puede restaurar el correcto funcionamiento de la ruta.

Finalmente, los agentes infecciosos cómo parásitos, virus o bacterias contienen sus propias proteasas que pueden ser objeto de modificación para el tratamiento de la enfermedad.

2.2. Estudio de la diana

Una vez identificada una posible diana, se tiene que estudiar para seleccionarla finalmente cómo diana terapéutica. Este estudio engloba conocer los sustratos, el mecanismo de señalización (si existe) y la regulación de dicha proteasa.

3. Los inhibidores de proteasas cómo agentes terapéuticos

Entre el 5 y el 10% de las dianas farmacéuticas actuales son proteasas.^[3] Es por este motivo que el mercado de inhibidores de proteasas es muy atractivo para las compañías farmacéuticas.

3.1. Estrategias para el descubrimiento de principios activos: obtención de un “hit”

Los inhibidores de proteasas son moléculas que bloquean la entrada del sustrato al centro activo de la proteasa o impiden su procesamiento. Para encontrar inhibidores de proteasas existen varias técnicas, aplicables también en campos externos a las proteasas.

Las estrategias de descubrimiento de fármacos basadas en previo conocimiento de la estructura proteica incluyen el cribado virtual,^[21] o síntesis *de novo* de inhibidores en base a información estructural del centro activo de la proteasa.

Por otro lado los métodos que no requieren previo conocimiento de la estructura tridimensional de la proteína son (principalmente): modificaciones del sustrato o producto, cribado de alto rendimiento (HTS), cribado basado en fragmentos, métodos *in silico* basados en similitud respecto ligandos conocidos o farmacóforos y cribado fenotípico.

3.2. El agente terapéutico perfecto: del “hit” al “lead”

Para que el “hit” sea considerado un “lead”, el inhibidor no sólo tiene que ser potente, sino que además ha de ser selectivo y presentar baja toxicidad.

Referente al diseño del inhibidor, las moléculas pequeñas suelen tener mejor propiedades biofísicas que las grandes (péptidos o anticuerpos). En cuanto al comportamiento cinético del inhibidor, los reversibles son los aconsejados en la mayoría de casos, ya que suelen presentar menos efectos secundarios. Finalmente, respecto al mecanismo de inhibición, los inhibidores tradicionales son competitivos, aunque actualmente existe una tendencia hacia el uso de inhibidores no competitivos, ya que en algunas proteasas, estas moléculas son más selectivas que los dirigidos al centro activo.^[5]

En referencia a las propiedades fisiológicas, el inhibidor ha de tener unas propiedades ADMET adecuadas. ADMET incluye: administración, distribución, metabolismo, excreción y toxicidad. Para la evaluación de dichas propiedades existe una batería de ensayos *in vitro*, cómo por ejemplo: ensayos de células Caco-2 o ensayos de citocromo P450.

4. Casos de estudio

4.1. Dipeptidil peptidasa IV (DPP IV)

La dipeptidil peptidasa IV (DPP IV, CD26, EC 3.4.14.5) es una serina exopeptidasa que pertenece a la subfamilia B de la familia S9 (clan SC). Procesa substratos con N-terminal X-Pro o X-Ala, cortando después de prolina o alanina siempre y cuando el P1' no sea prolina y el extremo N terminal esté libre.^[52]

Es una enzima de 220 KDa, homodimérica y glicosilada.^[53] El extremo N terminal de la DPP IV está en el citosol, mientras que la mayor parte de la proteína es extracelular.

El ectodominio de cada monómero está compuesto por dos dominios estructurales: el hidrolasa- α/β y la turbina- β .^[53] El centro activo de la enzima se encuentra entre ambos dominios y es accesible por dos aberturas: una localizada en el centro de la turbina- β y otra presente entre los dos dominios. Actualmente, se cree que es la última la que utiliza el sustrato.

La DPP IV hidroliza un amplio abanico de substratos, incluyendo hormonas gastrointestinales, citoquinas secretadas y péptidos la familia pancreática.

Además de su actividad proteolítica, la DPP IV interacciona con otras proteínas como la adenosina deaminasa (ADA) o la seprasa (FAP).

La DPP IV está involucrada en patologías como la diabetes de tipo 2, cáncer o artritis reumatoide, pero es en el primer caso donde la DPP IV es una diana validada. La relación de la enzima con la enfermedad es en base a dos de los substratos de la proteasa.

Las moléculas GLP-1 (péptido similar al glucagón 1) y GIP (péptido inulinotrópico dependiente de glucosa) son hormonas incretinas producidas por las células L del duodeno (GLP-1) y por las células K (GIP) del intestino. Después de la ingestión, se secretan al torrente sanguíneo, donde circulan hasta llegar al páncreas. Una vez allí, las incretinas estimulan tanto la producción como secreción de insulina. Sin embargo, una vez en el torrente sanguíneo las hormonas son rápidamente inactivadas por la DPP IV. Así pues, los inhibidores de dicha proteasa son agentes terapéuticos para el tratamiento de la diabetes de tipo 2.^[52]

Actualmente hay cinco inhibidores de la DPP IV en el mercado, bajo el nombre genérico de sitagliptin, vildagliptin, saxagliptin, alogliptin y linagliptin.^[32]

4.2. Prolil oligopeptidasa (POP)

La prolil oligopeptidasa (POP, EC 3.4.21.26) es una serina exopeptidasa que pertenece a la subfamilia A de la familia S9 (clan SC). Procesa péptidos de alrededor 30 residuos, cortando después del extremo C terminus de prolina.^[99]

POP es una proteína citosólica de 81 KDa. Está compuesta por dos dominios: hidrolasa- α/β y turbina- β . El centro activo se encuentra entre ambos dominios y es accesible para el sustrato mediante una abertura lateral entre ambos dominios. Hay otra abertura en el centro de la turbina- β , pero es de un diámetro demasiado pequeño para permitir la entrada del sustrato.

En cuanto a los sustratos de POP, ensayos *in vitro* han demostrado que la enzima puede hidrolizar un gran número de neuropéptidos. Sin embargo, no se ha podido demostrar *in vivo*, ya que POP es citosólica pero estos sustratos son extracelulares.

La proteasa también interacciona con otras proteínas como la proteína asociada al crecimiento 43 (GAP-43) o la α -sinucleína.^[99]

POP está relacionada con desórdenes cognitivos (esquizofrenia y desorden bipolar). Actualmente, no se ha podido esclarecer el mecanismo de acción de POP en estas patologías. No obstante, estudios recientes han demostrado que inhibidores de POP provocan un aumento de la concentración de IP3 (inositol trifosfato). Esta molécula estimula la liberación de ion calcio del retículo endoplasmático al citosol. Una vez ahí, el ion calcio está involucrado en procesos de memoria y aprendizaje. Así pues, aunque no se conoce cuál es la relación directa entre POP e IP3, los inhibidores de la proteasa son fármacos potenciales para el tratamiento de desórdenes cognitivos.

Por otro lado, la interacción de POP con α -sinucleína promueve la agregación de esta última proteína.^[115] Además, los inhibidores de POP son capaces de revertir el proceso de agregación. Por lo tanto, los inhibidores de POP pueden ser fármacos para esta patología también, pese a que la proteasa no es una diana validada.

En la actualidad, no hay ningún inhibidor de POP en el mercado y sólo existe uno bajo investigación en desarrollo, la molécula epigallocatequina, proveniente del extracto de hojas de té verde japonés.^[32]

4.3. Catepsinas L y B

La catepsina L (EC 3.4.22.15) y la catepsina B (EC 3.4.22.1) son cisteína endopeptidasas pertenecientes a la subfamilia A de la familia C1 (clan CA).

La catepsina L se expresa en forma de proenzima de 30 KDa y tiene dos dominios: uno predominantemente formado por hélices α y el otro por láminas β .^[119] El centro activo se encuentra entre ambos dominios y está protegido por un segmento del propéptido. La maduración de la enzima se produce cuando el propéptido se libera y el centro activo queda accesible a los sustratos.

La catepsina B, al igual que la catepsina L, se expresa en forma de proenzima y contiene los mismos dos dominios.^[120]

La expresión de ambas enzimas es ubicua y está altamente regulada en todas las fases del ciclo celular. Se encuentran particularmente concentradas en los lisosomas y en la superficie celular de linfocitos T citotóxicos y células "natural killer".^[121]

La catepsina B hidroliza una amplia variedad de sustratos, con preferencia por moléculas con residuos básicos o fenilalanina en la posición P2 y amino ácidos con cadena lateral poco voluminosa en P1. Por el contrario, catepsina L es más restrictiva e hidroliza moléculas con amino ácidos hidrofóbicos en P2 y residuos polares en la posición P1.

Entre las funciones de ambas proteasas destacan el reciclaje proteico, la activación de enzimas, la maduración de hormonas y la presentación de antígeno.

Aunque las catepsinas L y B no son dianas validadas para el tratamiento del cáncer, se ha observado que estas enzimas son redirigidas a la superficie celular e incluso secretadas al exterior, en células tumorales. Una vez en el exterior, las catepsinas degradan las proteínas de la matriz extracelular, favoreciendo el crecimiento del tumor y la metástasis.^[122]

Por otro lado, se han detectado altos niveles de estas enzimas en el líquido sinovial de pacientes con artritis reumatoide. Se ha demostrado que, ante procesos de inflamación, se induce secreción de catepsinas, y una vez en el exterior celular, provocan la destrucción del tejido. De esta forma, las catepsinas L y B son también potenciales dianas para el tratamiento de la artritis reumatoide.^[131]

En el panorama actual, no hay ningún inhibidor de estas enzimas que esté en el mercado y únicamente hay dos moléculas en investigación activa, uno en fase I y otro en preclínica.^[32]

OBJETIVOS

Las proteasas están involucradas en un alto número de enfermedades y por lo tanto, son dianas terapéuticas relevantes.

Por este motivo, nuestra principal meta es el descubrimiento de inhibidores de proteasas como agentes terapéuticos. Hemos centrado nuestro estudio en cuatro proteasas: la dipeptidil peptidasa IV, la prolil oligopeptidasa y las catepsinas L y B.

Para la búsqueda de inhibidores, se han seleccionado tres estrategias: cribado de plantas medicinales, cribado de alto rendimiento y caracterización de una biblioteca proveniente de la química combinatoria.

Para conseguir nuestra meta se han fijado cuatro objetivos:

1. Expresar DPP IV y establecer una metodología para su estudio por NMR
 - a) Poner a punto un sistema de expresión de la DPP IV recombinante
 - b) Establecer una metodología de marcaje que permita estudiar la DPP IV mediante resonancia magnética nuclear (NMR)
2. Descubrir inhibidores de DPP IV de origen botánico
 - a) Seleccionar plantas con propiedades reductoras de la concentración de azúcar en sangre
 - b) Aislar inhibidores de la DPP IV de extractos de plantas y caracterizar el mecanismo de inhibición
3. Descubrir inhibidores de POP mediante HTS
 - a) Poner a punto el ensayo de polarización de la fluorescencia para la detección de ligandos del centro activo de POP
 - b) Cribar compuestos no tóxicos como ligandos de POP y validarlos como inhibidores.
 - c) Caracterizar el mecanismo de inhibición de los inhibidores de POP.
4. Identificar inhibidores de catepsinas L y B a partir de una biblioteca de química combinatoria.
 - a) Cribar una biblioteca, especialmente diseñada para la inhibición de las cisteína proteasas.
 - b) Caracterizar el mecanismo de inhibición de los inhibidores de catepsina.

RESULTADOS

Los resultados obtenidos durante la presente tesis están divididos en cuatro capítulos. Los dos primeros hacen referencia a DPP IV, el tercero a POP y el último a las catepsinas L y B. Los capítulos son los siguientes:

Capítulo 1: Metodología para el marcaje de la DPP IV y la interacción de la proteína con sus inhibidores

Capítulo 2: Descubrimiento de inhibidores de la DPP IV a partir de extractos de plantas

Capítulo 3: Descubrimiento de inhibidores de la POP a partir del cribado de alto rendimiento (HTS)

Capítulo 4: Caracterización de las peptidil aril vinil sulfonas como inhibidores de las catepsinas L y B

Capítulo 1: Metodología para el marcaje de la DPP IV y estudio de la interacción de la proteína con sus inhibidores

1.1. Expresión recombinante de la DPP IV en células de insecto

El uso de sistemas procariotas para la expresión de DPP IV no era una opción debido a que DPP IV tiene modificaciones post-traduccionales vitales tanto para una correcta estructura tridimensional como para una actividad proteolítica adecuada.

Dentro de los sistemas eucariotas disponibles, se seleccionaron las células de insecto, dadas la facilidad de su cultivo y la simplicidad de modificación para la expresión de proteínas recombinantes, mediante la técnica de baculovirus. Además, las células de insecto permiten la obtención de altos rendimientos en comparación a la células de mamífero, y realizan modificaciones post-traduccionales más parecidas a los humanos que las que realizan las levaduras.

1.1.1. Obtención y amplificación de los baculovirus

Para la obtención de baculovirus recombinantes, que contuvieran la secuencia de DPP IV, se utilizó la técnica de transposición del gen de DPP IV en el DNA de un baculovirus parental (bácmido parental). Dicha transposición es posible gracias a las bacterias *E. Coli* DH10Bac (figura r.1.).

Previamente a la transposición, la secuencia de DNA correspondiente al ectodominio de DPP IV fue clonada en un plásmido dador. En este caso, dos plásmidos, denominados pFastBac EGT-N y pFastBac EGT-C, fueron utilizados.

Transfomación de las bacterias *E. Coli* DH10Bac con el plásmido dador/DPP IV permitió la obtención de los b́acmidos recombinantes/DPP IV (2 constructos, dependiendo del plásmido dador). Una vez aislados los b́acmidos,

Después, las células de insecto Sf9 se transfectaron con los b́acmidos recombinantes. Como resultado, las células eucariotas secretaron los correspondientes baculovirus recombinantes.

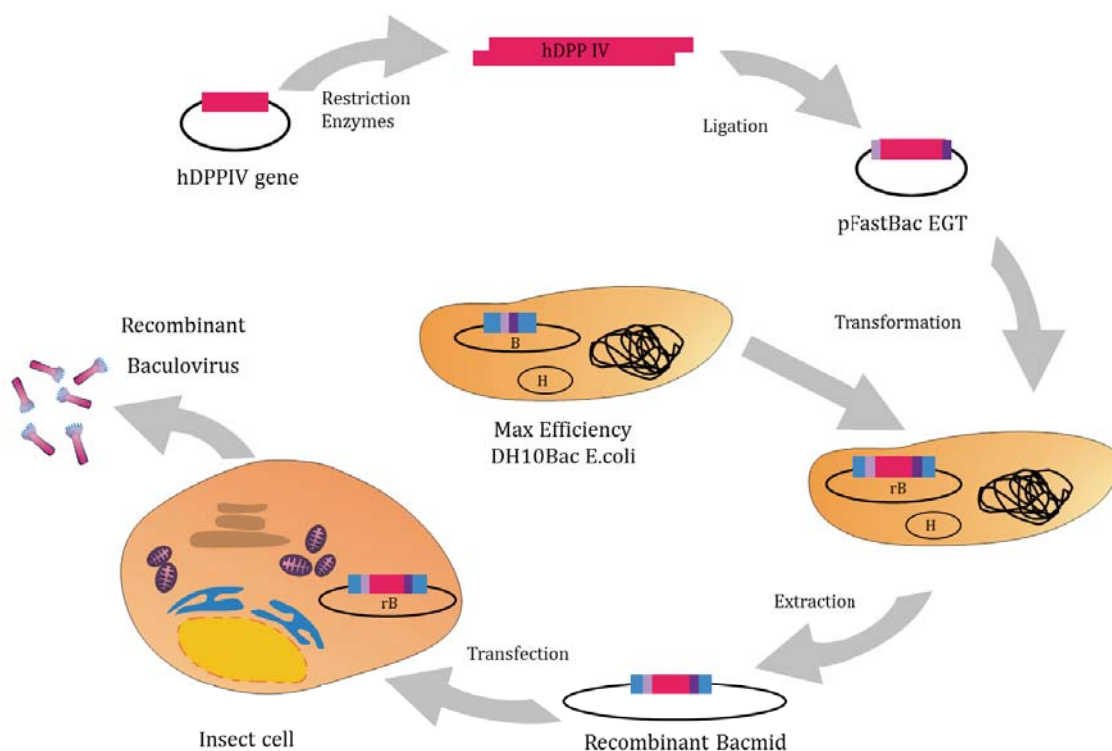


Figura r.1. : Esquema de la producción de los baculovirus recombinantes.

Los baculovirus obtenidos de la transfección carecen de potencia infectiva, y por ello, se amplificaron mediante sucesivas infecciones. La tercera población (resultado del segundo ciclo infectivo) fue utilizada para la expresión de la DPP IV.

1.1.2. Expresión y purificación de la DPP IV

Ambos procesos se optimizaron de forma paralela (figura r.2.).

Para la expresión de la proteína, se probaron dos líneas celulares (Sf9 y Hi5), dos constructos de DPP IV (uno por cada plásmido dador), diferentes condiciones de infección (concentración inicial de células y ratio de baculovirus/células) y varios tiempos de incubación (entre 2 y 4 días). De todas las condiciones probadas, la óptima fue la siguiente: células Sf9 a una concentración inicial de 3×10^6 células/mL; infección con el constructo proveniente del plásmido dador pFastBac EGT-C; e incubación durante 4 días a $T = 27^\circ\text{C}$ y 130 rpm.

La purificación de la proteína del medio de cultivo (ya que se secretaba) se realizaba después de un paso de diafiltración para la eliminación de elementos quelantes y la reducción del volumen a purificar.

Se probaron diferentes columnas de purificación: afinidad (columna de níquel), intercambio aniónico y de exclusión molecular. Finalmente, se optimizó la purificación y se obtuvo DPP IV pura en un único paso utilizando la columna de níquel.

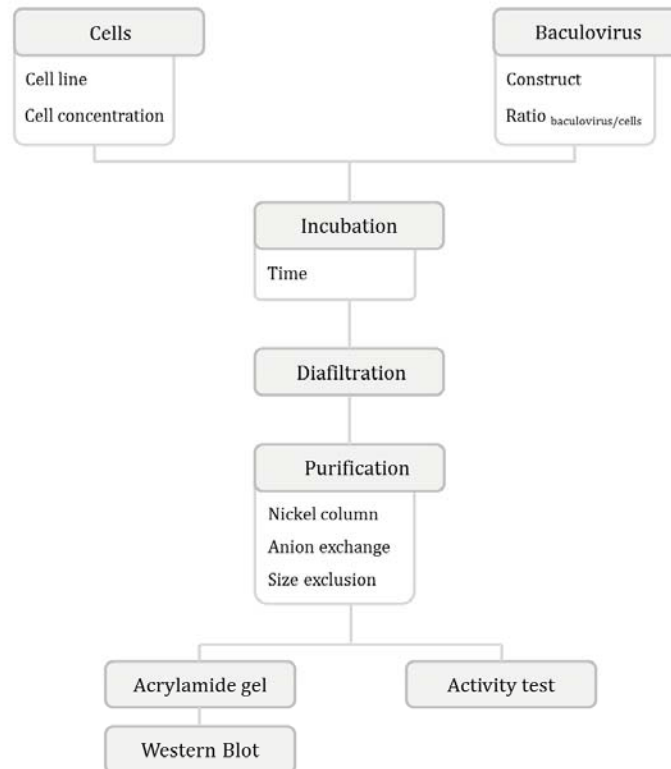


Figura r.2. : Esquema de la expresión y purificación de DPP IV.

Cómo resultado, la proteína obtenida fue pura y activa, con una actividad específica entre 1 y 5 Unidades de sustrato / mg DPP IV.

1.2. Estudio de la DPP IV mediante NMR

DPP IV es una proteína grande para ser estudiada mediante resonancia magnética nuclear (NMR). Por este motivo, se utilizó una estrategia que combinaba el marcaje selectivo de la proteína con el uso de experimentos TROSY-HSQC. Mediante el marcaje selectivo se evitaba la superposición de señales en el espectro, lo que facilita el análisis del experimento. Por otro lado, los experimentos de NMR basados en secuencias de pulsos TROSY (espectroscopia con relajación transversa optimizada) permiten aumentar la sensibilidad respecto secuencias no-TROSY. Esto es debido a que estos experimentos seleccionan la componente con relajación lenta de cada señal, en vez de realizar la

interconversión de las componentes lenta y rápida (cómo hacen los experimentos tradicionales de HSQC).

Además del problema que presenta estudiar una proteína tan grande mediante NMR, otro inconveniente en el presente estudio es el limitado número de técnicas de marcaje que existen en células de insecto, en comparación al amplio espectro de posibilidades que existen en bacterias.

1.2.1. Marcaje selectivo de la DPP IV

Se seleccionó un marcaje selectivo con ^{13}C de los metilos de los residuos de metionina. Este marcaje fue posible gracias a un medio específico de células de insecto que contiene todos los nutrientes necesarios para el correcto crecimiento de estos organismos, pero carece de metionina y cistina. Por lo tanto, se suplementó dicho medio con cistina y metionina marcada. La infección, incubación y purificación de la DPP IV marcada se realizaron de la misma forma que la proteína no marcada.

1.2.2. Experimento Methyl-TROSY

El Methyl-TROSY consiste en un experimento HMQC dónde la secuencia de pulsos ha sido optimizada para proteínas grandes.

Los experimentos se realizaron en los equipos de NMR Bruker de 600 y 800 MHz con criosonda.

De los 14 residuos de metionina que contiene DPP IV, se esperaban 14 señales en el espectro de NMR. Sin embargo, se detectaron 11 señales (figura r.3.)

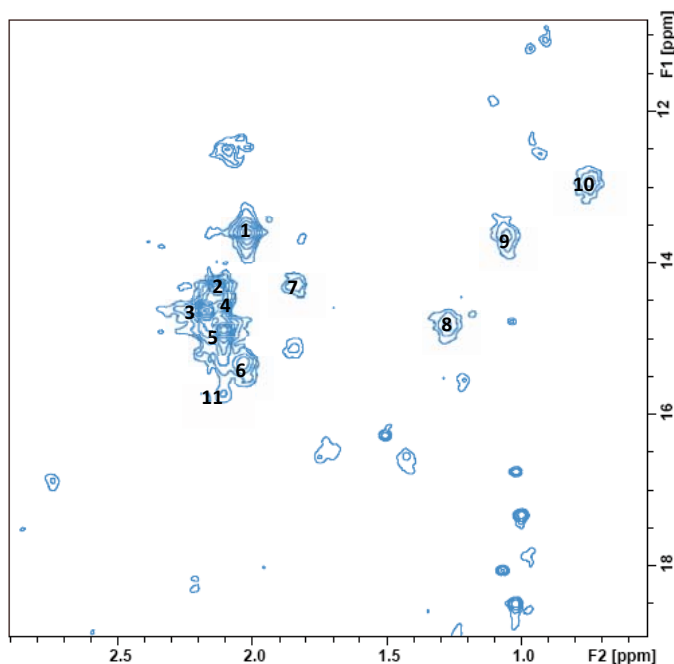


Figura r.3. : Espectro del experimento Methyl-TROSY de la muestra de DPP IV marcada selectivamente con ^{13}C en los metilos de los residuos de metionina.

1.2.3. Estudio de la interacción de DPP IV con sus inhibidores mediante NMR

Dado que la metodología de marcaje de la DPP IV estaba optimizada y los espectros de NMR permitían la observación de once señales correspondientes a once residuos de metionina, se estudió cómo la inhibición de la proteasa afectaba al resultante espectro.

En caso de detectar diferencias, la técnica de NMR sería una herramienta adecuada para estudiar el efecto de los inhibidores en la DPP IV en solución (y no cristalizada).

Para ello, se realizaron los experimentos correspondientes utilizando como inhibidores modelo las moléculas: P32/98 (Isoleucina-tiazolidida), NVP DPP 728 y berberina.

El primer compuesto, es un inhibidor competitivo no-covalente de la DPP IV. La IC_{50} fue determinada en 520 nM. El espectro resultante (figura r.4.) presentaba las mismas once señales observadas en el espectro de la proteína no inhibida, pero 4 señales adicionales fueron detectadas. De estas cuatro, tres eran propias del inhibidor, pero la cuarta era de origen proteico (señalada con una "D" a la izquierda de la señal en la figura r.4.)

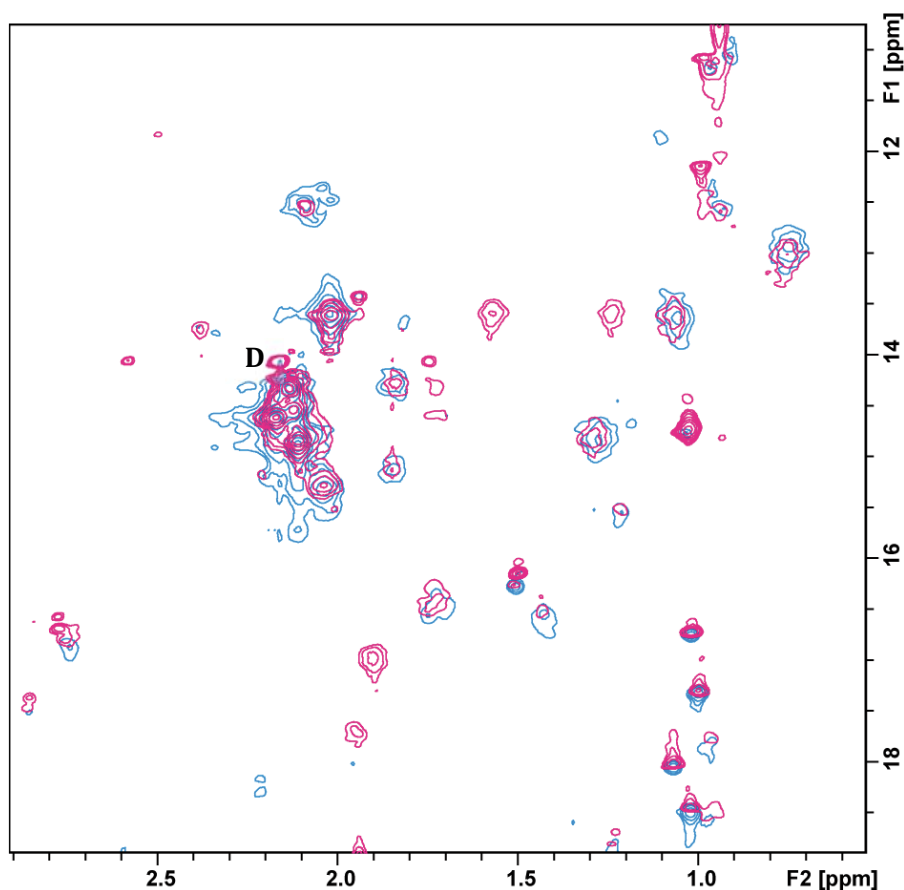


Figura r.3. : Espectro del experimento Methyl-TROSY de la muestra de DPP IV marcada selectivamente con ^{13}C en los metilos de los residuos de metionina libre (en azul) e inhibida con la molécula P32/98 (en rosa).

En los experimentos realizadas con NVP DPP 728, un inhibidor competitivo covalente de la DPP IV y con berberina, inhibidor competitivo no covalente, también se detectó la señal extra (señal D).

Teniendo en cuenta que esta señal extra se observa en los tres casos, creemos que su aparición es consecuencia de un pequeño cambio estructural que sufre la proteasa después de ser inhibida. No obstante, no disponemos de las suficientes pruebas para demostrarlo, ya que la señal no pudo ser asignada (relacionar la señal en el espectro con el residuo de metionina que la causa). Por lo tanto, más experimentos son necesarios para demostrar o refutar dicha hipótesis.

Capítulo 2: Descubrimiento de inhibidores de la DPP IV a partir de extractos de plantas

2.1. Selección, extracción y test

Se seleccionó un conjunto de plantas cuya acción antidiabética había sido previamente reportada. Esta colección incluye: plantas antidiabéticas comunes, plantas de la medicina tradicional china y plantas brasileñas. Dado que en el laboratorio se disponía de una colección de plantas mediterráneas, éstas también se seleccionaron, aunque no habían sido previamente relacionadas con desórdenes del nivel de azúcar en sangre.

Se realizaron tres tipos de extracción: acuosa en soxhlet, maceración etanólica y extracción hidro-alcohólica en soxhlet.

En el caso de las plantas mediterráneas, sólo se realizó la primera extracción (acuosa en soxhlet). Los extractos de plantas brasileñas se obtuvieron mediante extracción hidro-alcohólica (70% de etanol) en soxhlet. Finalmente, para las plantas antidiabéticas comunes y las de la medicina tradicional china se realizaron las extracciones acuosas en soxhlet y la maceración etanólica.

Una vez obtenidos los extractos, se testó su potencia inhibidora de la DPP IV, mediante ensayos enzimáticos. Los resultados se muestran en las figuras r.4. a r.7.

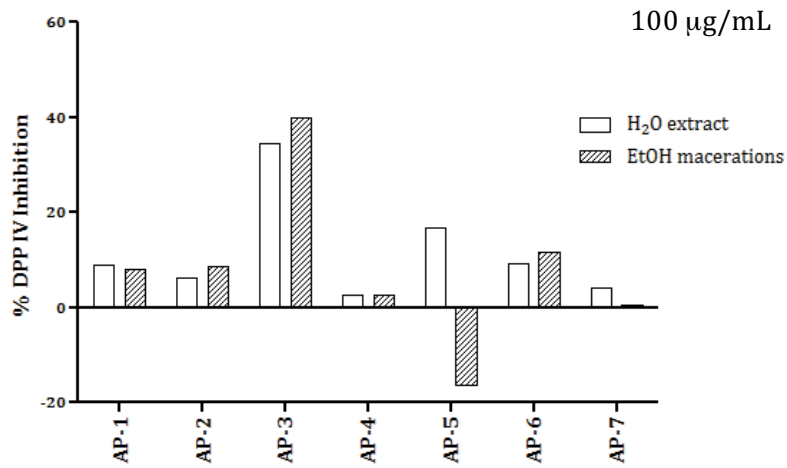


Figura r.4. : Potencia de inhibición de la DPP IV de los extractos acuosos y etanólicos de las plantas antidiabéticas comunes.

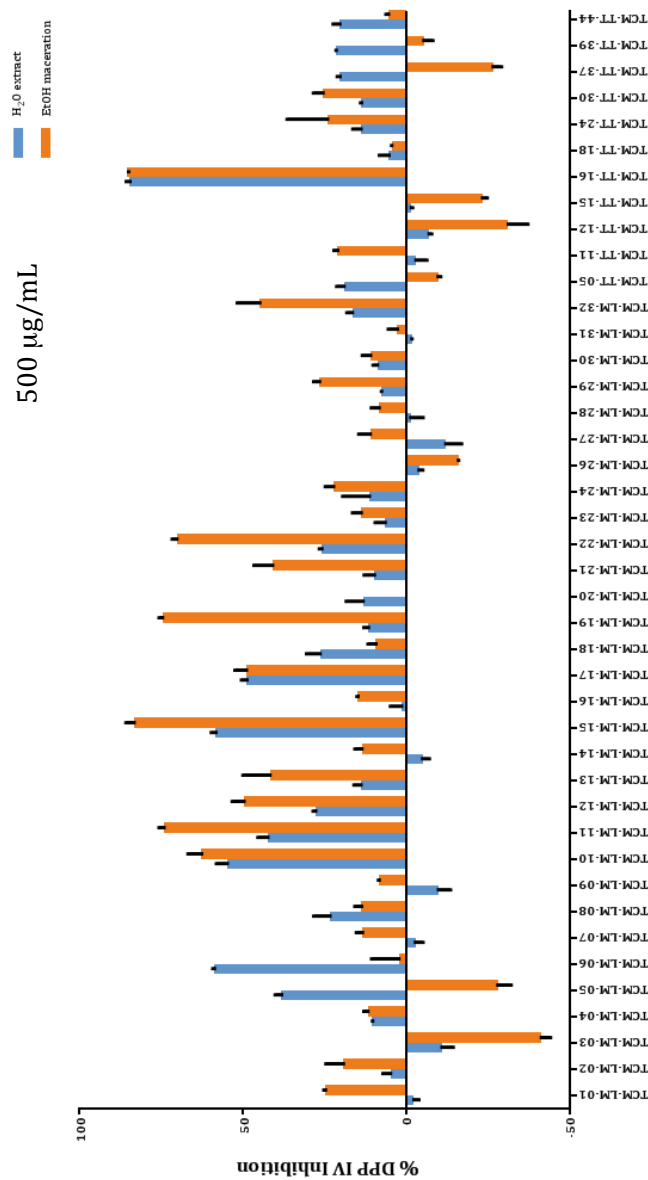


Figura r.5. : Potencia de inhibición de la DPP IV de los extractos de las plantas de la medicina tradicional china: extractos acuosos en azul y etanólicos en naranja.

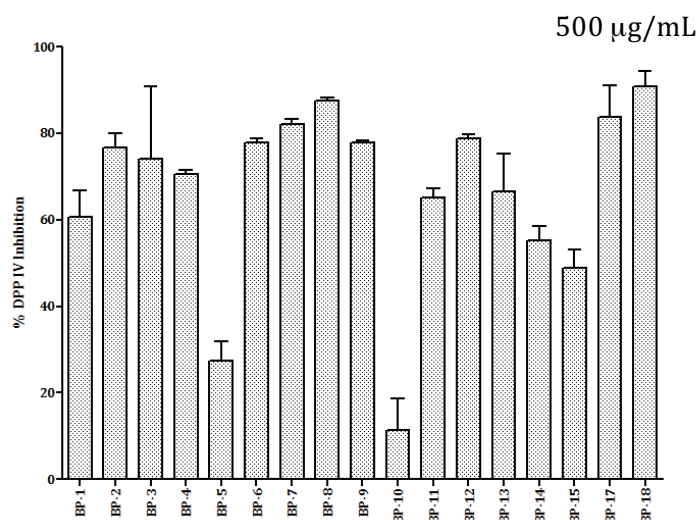


Figura r.6. : Potencia de inhibición de la DPP IV de los extractos hidro-alcohólicos de las plantas brasileñas.

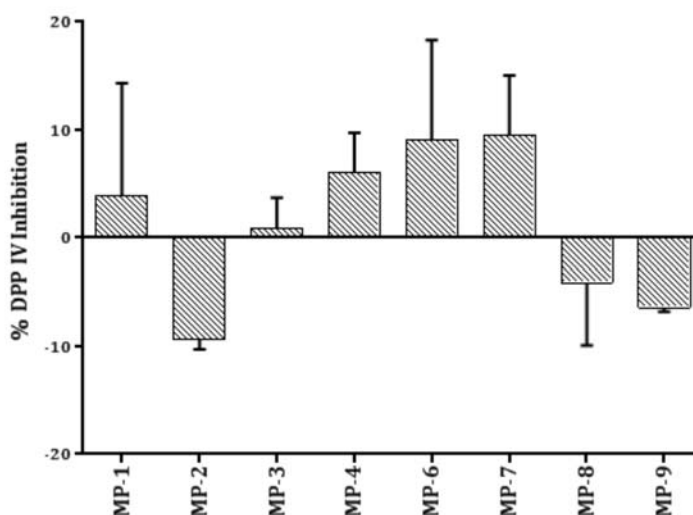


Figura r.7. : Potencia de inhibición de la DPP IV de los extractos acuosos de las plantas mediterráneas.

En total, extractos se testaron 123 extractos provenientes de 7 plantas comunes antidiabéticas, 42 plantas de la medicina tradicional china, 17 plantas brasileñas y 8 mediterráneas.

2.2. *Coutarea latiflora*

El extracto etanólico de la planta *Coutarea latiflora* (código AP-3) demostró una gran potencia de inhibición de la DPP IV: alrededor del 80% de inhibición con una concentración de extracto de 500 µg/mL. Dadas las perspectivas positivas de encontrar un inhibidor de la proteasa, el extracto etanólico se seleccionó para un estudio a fondo.

Después de fraccionar el extracto y testar dichas fracciones, se purificaron e identificaron dos flavonoides como inhibidores de la DPP IV, denominadas AP-3-a y AP-3-b. La que presentó más potencia fue AP-3-a (figura r.8.) con una IC_{50} de 115 μM , por lo que se decidió un análisis del mecanismo de inhibición de dicha molécula.

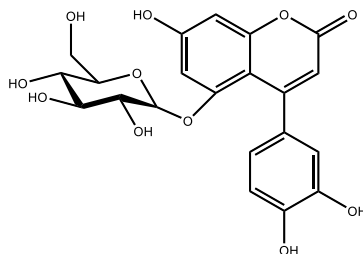


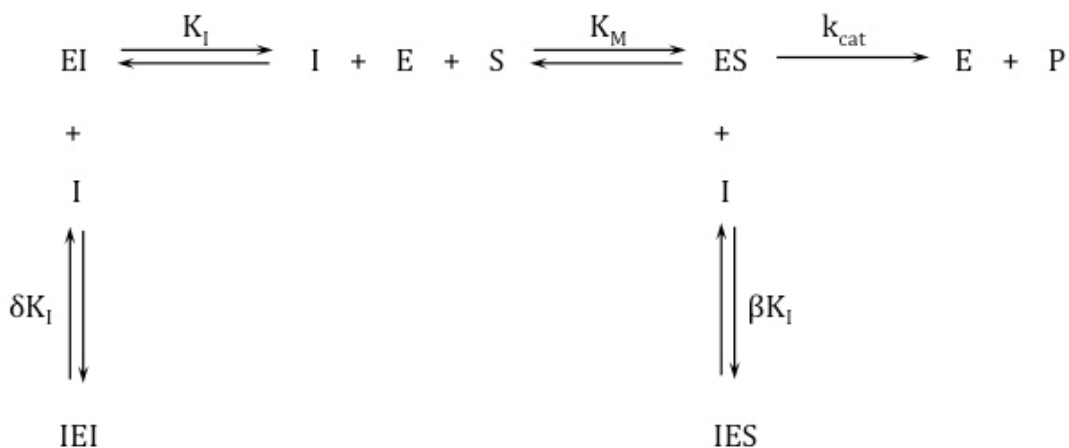
Figura r.8. : Estructura química de la molécula AP-3-a, identificada como un inhibidor de la DPP IV.

La molécula se caracterizó como un inhibidor reversible. Además, se testó la inhibición en presencia de detergente (Triton X-100). Este experimento demostró que la molécula no inhibía a la proteasa debido a la formación de micelas (fenómeno común en flavonoides). También se comprobó la selectividad de AP-3-a frente a la serina proteasa POP. Aunque la inhibición de POP era menor, la ratio determinada fue de 1.13, de modo que para aumentar la selectividad de AP-3-a sobre DPP IV respecto a otras proteasas, se deberían realizar modificaciones en la estructura de la molécula.

El mecanismo de inhibición fue elucidado mediante ensayos cinéticos. Sorprendentemente, AP-3-a inhibía DPP IV de forma parabólica (no lineal). Este comportamiento implica la presencia de dos puntos de unión de la molécula en la proteína, uno de ellos competitivo y el otro no-competitivo. Así, el inhibidor puede unirse en dos posiciones diferentes al mismo tiempo.

Para caracterizar el sistema y calcular las constantes de inhibición, se realizaron las representaciones secundarias de Lineweaver-Burk (gráficos de i) las pendientes e ii) los puntos de intersección con el eje y, de las rectas de Lineweaver-Burk).

Los valores de las constantes se encuentran en la figura r.9.



$$K_M = 92.59 \mu\text{M}; K_i = 157.6 \mu\text{M}; \beta = 1.49; \delta = 1.97$$

Figura r.9. : Mecanismo de inhibición de la DPP IV que realiza la molécula AP-3-a.

Finalmente, se estudió el efecto de la inhibición de la DPP IV que provocaba AP-3-a, mediante NMR.

El espectro obtenido (figura r.10.), pese a que fue de baja resolución, muestra la ausencia de la señal "D", observada en los espectros de DPP IV incubada con los inhibidores competitivos.

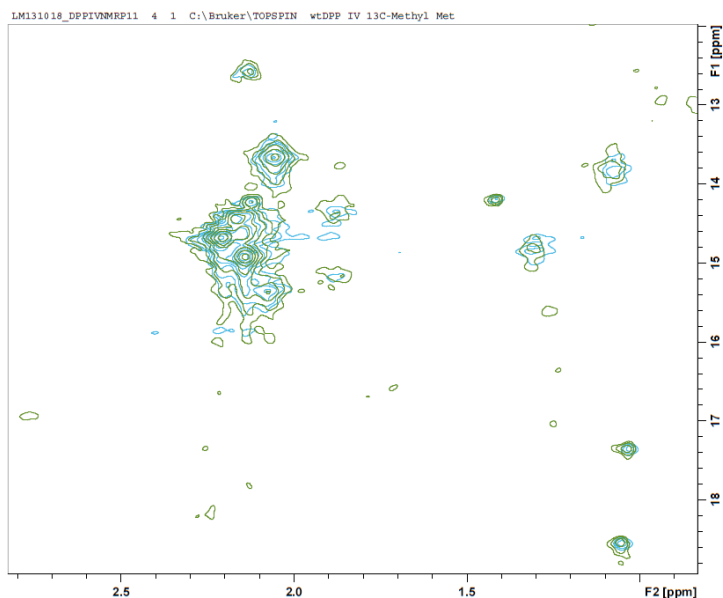


Figura r.10. : Espectro del experimento Methyl-TROSY de la muestra de DPP IV marcada selectivamente con ^{13}C en los metilos de los residuos de metionina libre (en azul) e inhibida con la molécula AP-3- (en verde).

Así pues, se estableció una hipótesis, relacionada con el comportamiento del inhibidor, para explicar la ausencia de la señal "D". Si la señal "D" es resultado de una inhibición

competitiva, una molécula como AP-3-a, con un comportamiento parabólico no sería capaz de promover la aparición de dicha señal en el espectro.

Para demostrar o refutar cualquiera de las dos, se necesitan más experimentos.

Capítulo 3: Descubrimiento de inhibidores de la POP a partir del cribado de alto rendimiento (HTS)

La estrategia para la obtención de inhibidores de POP se centró en el cribado de compuestos no tóxicos mediante la técnica de cribado de alto rendimiento (HTS) y el ensayo de polarización de la fluorescencia.

3.1. Puesta a punto del ensayo de polarización de fluorescencia

El ensayo de polarización de fluorescencia (FP) se basa en dos conceptos. Por un lado, la rotación de las moléculas sigue el modelo browniano: moléculas pequeñas rotan más rápidamente que las grandes. El segundo concepto está relacionado con la naturaleza intrínseca del fluoróforo: la excitación de un fluoróforo con fluorescencia polarizada provoca la emisión de fluorescencia también polarizada.

Entonces, después de excitar el fluoróforo, las moléculas rotan y el plano de fluorescencia cambia al mismo tiempo. Las moléculas pequeñas rotan más rápidamente y por lo tanto, el plano cambia y se detecta poca FP. En cambio, las moléculas grandes apenas rotan y más FP se detecta.

La aplicación de la FP para la detección de ligandos de proteína utiliza estos conceptos. Una molécula con afinidad por un punto específico de la proteína se marca con un fluoróforo para la obtención de una sonda. La unión de sonda y proteína es un complejo grande en comparación a la sonda libre. Así el complejo tendrá valores de FP superiores a los de la sonda libre. Si al complejo de proteína y sonda se añade un ligando de POP capaz de desplazar la sonda, ésta quedará libre y los valores de FP serán inferiores.

Por lo tanto, se puede determinar si una molécula es ligando o no de la proteína midiendo los valores de FP.

Para el presente estudio, se diseñó una sonda consistente en un péptido con afinidad para el centro activo unido a un fluoróforo. Dos fluoróforos fueron utilizados para minimizar el número de falsos positivos. Fueron carboxifluoresceína y TAMRA.

Se demostró que la sonda se unía a POP y que además era desplazada después de la adición del inhibidor canónico de POP (ZPP) o bien por el péptido no marcado.

3.2. Cribado de alto rendimiento (HTS)

Las bibliotecas de compuestos seleccionadas para el cribado de alto rendimiento (HTS) fueron aquellas que contenían compuestos cuya toxicidad había sido estudiada y habían sido determinados como seguros para su uso en humanos. Se eligieron 6 bibliotecas con un total de 4,500 compuestos.

El HTS se realizó para los 4,500 compuestos con las dos sondas y para cada uno de ellos se obtuvo 4 valores: FP de sonda 1, FP de sonda 2, TF (Fluorescencia total) de sonda 1 y TF de sonda 2. Dado que el experimento se midió por duplicado, se generó un total de 8 valores/compuesto.

Los datos fueron evaluados en base a la diferencia de los valores de FP y TF respecto a los valores control. Para determinar los positivos se fijaron límites y las moléculas que los superaron (73 moléculas) se definieron como ligandos de POP (figura r. 11). Cabe mencionar que se identificaron 84 positivos, pero varias moléculas se encontraban repetidas en varias placas, de forma que el número final de positivos fue 73.



Figura r.11. : Representación de la potencia de las moléculas del HTS que superaron los límites impuestos para ser considerados ligando de POP. Un valor de z-score implica mayor potencia para desplazar la sonda de FP.

3.3. Validación de los hits

Para validar los ligandos de POP como inhibidores de la proteasa, se tenía que comprobar el porcentaje de inhibición mediante ensayos enzimáticos.

Primero de todo, los compuestos se agruparon en “clusters”, grupos que englobaban moléculas similares en estructura. Como resultado, se generaron 16 clusters.

Seguidamente, se seleccionaron representantes de cada cluster en base al resultado de FP y a la puntuación obtenida en un estudio de docking que se realizó. Así pues, se escogieron 37 moléculas de los 73 ligandos de POP para su estudio de inhibición de POP. Se realizó ensayo enzimático a la concentración de 200 μ M y se calculó el porcentaje de inhibición de la proteasa (figura r.12.)

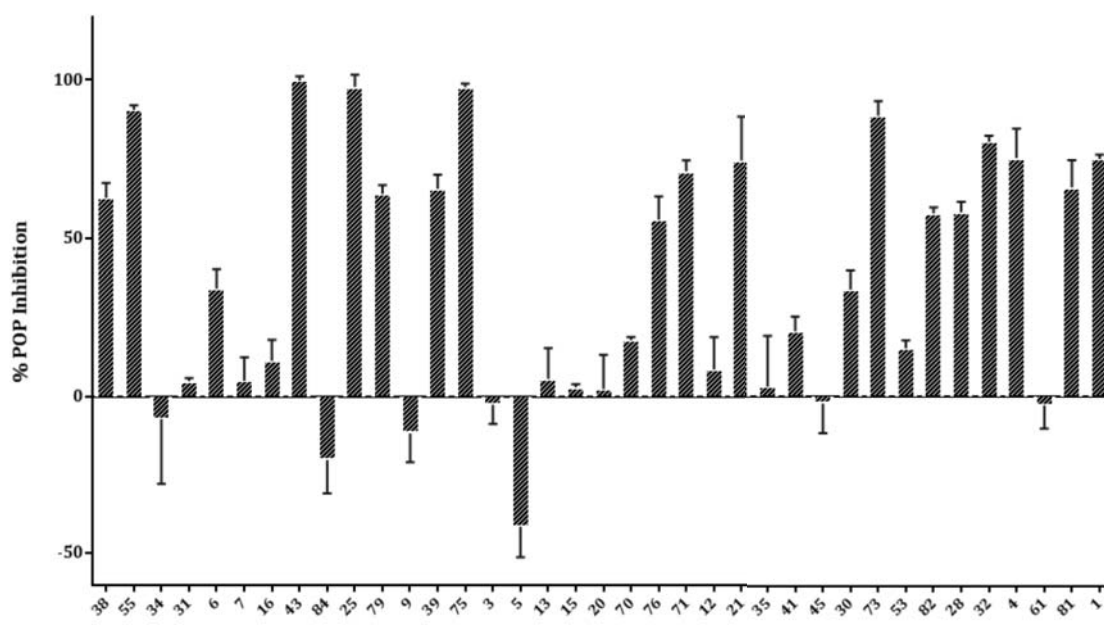


Figura r.12.: Porcentaje de inhibición de POP que realizan los 37 representantes de los clusters.

De las 37 moléculas se seleccionaron las que causaban una inhibición de POP superior al 80% para un estudio enzimático. También se añadió para este análisis la molécula que presentaba el mejor resultado de docking.

Se calculó el valor de IC_{50} respecto a POP de las seis moléculas. En la representación (figura r.13.) se puede observar dos tipos de comportamiento en función de la pendiente de las curvas. Un primer grupo (en colores rosas y morados) tiene una pendiente más acusada que el segundo grupo (en colores verdes).

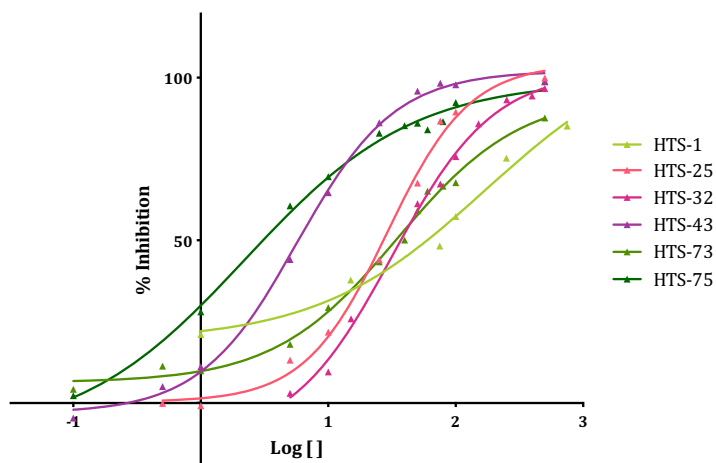


Figura r.13. : Curvas de la IC₅₀ de las seis moléculas seleccionadas.

Dado que los dos mejores inhibidores de POP pertenecían a grupos diferentes (según la pendiente de la curva de IC₅₀), se seleccionaron para un elucidar su mecanismo de inhibición.

La molécula HTS-43 (pendiente acusada) se caracterizó cómo un inhibidor competitivo de POP con una K_i = 3 μM. Sin embargo, el compuesto HTS-75 (pendiente suave) se definió cómo un inhibidor parabólico.

Para caracterizar el sistema y calcular las constantes de inhibición, se realizaron las representaciones secundarias de Lineweaver-Burk (gráficos de i) las pendientes e ii) los puntos de intersección con el eje y, de las rectas de Lineweaver-Burk).

Los valores de las constantes para HTS-75 respecto POP se encuentran en la figura r.14.

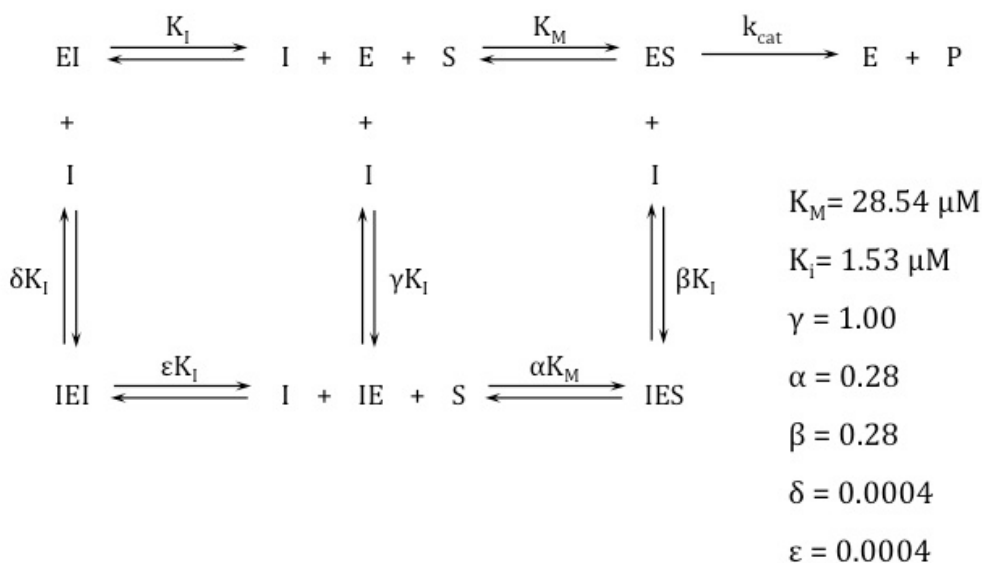


Figura r.14. : Mecanismo de inhibición de la POP que realiza la molécula HTS-75.

Capítulo 4: Caracterización de las peptidil aril vinil sulfonas como inhibidores de las catepsinas L y B

En el siguiente capítulo se identificaron inhibidores de las catepsinas L y B a partir de una biblioteca proveniente de la química combinatoria y diseñada para la inhibición de cisteína proteasas.

La biblioteca se basaba en modificaciones de la estructura presente en la figura r.15.

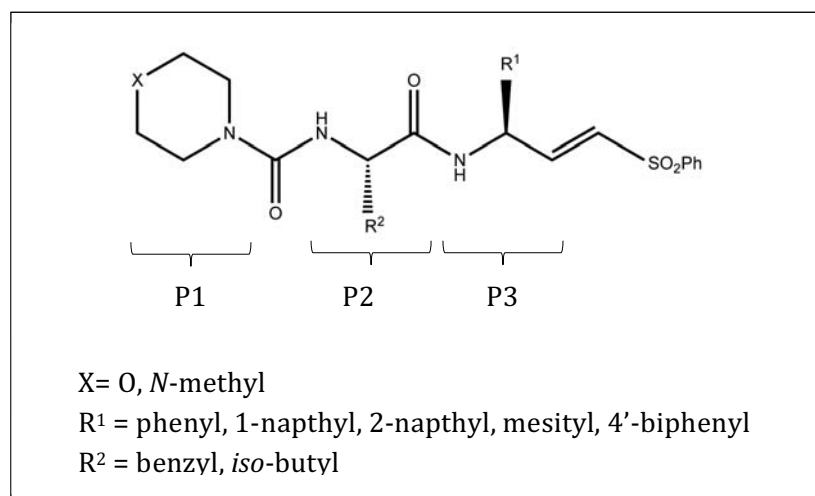


Figura r.15. : Estructura para el diseño de la biblioteca.

4.1. Cribado de la biblioteca

La actividad de los compuestos de la biblioteca como inhibidores de las catepsinas fue evaluada mediante el cálculo de los valores de IC₅₀. Entre las moléculas, se descubrió un potente inhibidor de la catepsina L, el compuesto PAVS-20 con una IC₅₀ de 2.6 nM. Además, se demostró su selectividad frente a la catepsina B (ratio 1/500). La molécula PAVS-20 no mostró inhibición de la serina proteasa DPP IV y una leve inhibición de POP (ratio 1/50,000).

4.2. Análisis y descripción del mecanismo de inhibición

Dado que PAVS-20 era un potente inhibidor de catepsina L con elevada selectividad, se seleccionó para realizar un estudio más exhaustivo.

Se realizaron experimentos cinéticos donde el tiempo de pre-incubación se modificó y se trazaron las gráficas de progreso de reacción. Así, se obtuvo las constantes de inhibición de PAVS-20 sobre catepsina L ($K_{inac} = 19.9$ nM y $k_{2nd} = 181.420$ s⁻¹ M⁻¹).

4.3. Docking

Se realizaron experimentos de docking para proporcionar una base estructural que permitiera interpretar las diferencias en potencias de inhibición entre los miembros de la biblioteca.

En la figura r.16. se puede observar cómo moléculas con sustituyente voluminoso en P3 no pueden interaccionar completamente con la superficie proteica (figura r.16. b). Esto explica la disminución de potencia de PAVS-19 ($IC_{50} > 10^3$) en comparación a PAVS-20 ($IC_{50} = 2.6$ nM), dos moléculas que sólo se diferencian en la posición P3.

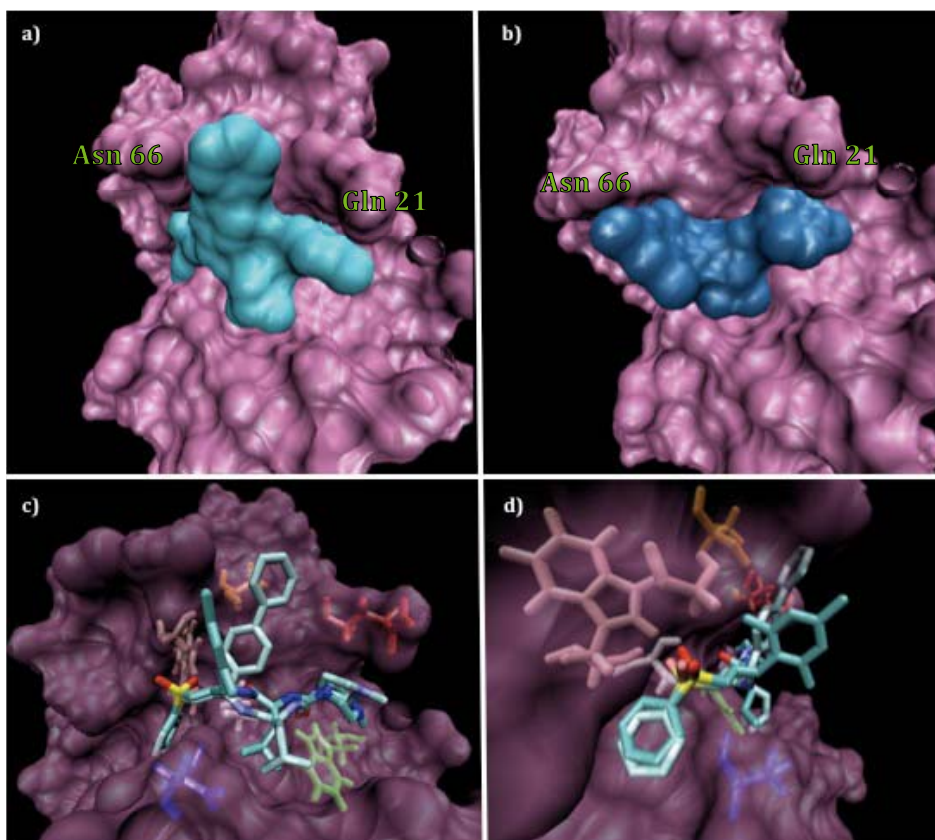


Figura r.16.: a) Docking de PAVS-20 en el centro activa de la catepsina L; b) Docking de PAVS-19 en el centro activo de la catepsina L; c y d) Comparación de los dockings de PAVS-19 y PAVS-20, desde dos puntos de vista diferentes.

DISCUSIÓN

La principal meta de la presente tesis era el descubrimiento de nuevos inhibidores de proteasa. Para las cuatro proteasas mencionadas (DPP IV, POP, catepsina L y B) se han encontrado nuevas moléculas con capacidad inhibitoria.

En base a la experiencia adquirida durante el presente estudio, de los tres proyectos de descubrimiento de fármacos, continuaría con el de los inhibidores de POP (capítulo 3 de resultados) por varios motivos.

Primero, actualmente hay poca competencia en el campo de los inhibidores de POP y no hay ninguno que esté bajo investigación activa en fase clínica.

Segundo, el compuesto identificado en el capítulo 3 de los resultados tiene un mecanismo de inhibición parabólico, un tipo de inactivación que nunca antes ha sido descrito para esta proteasa. Dado que POP puede estar involucrada en patologías mentales mediante interacciones de proteína-proteína, puede ser que el sitio no-competitivo donde se une el inhibidor parabólico sea una región crucial para que POP interaccione con otras proteínas. Así pues, el compuesto descubierto puede ser un potencial agente terapéutico para el tratamiento de condiciones como el Parkinson, el desorden bipolar y la esquizofrenia.

Tercero, estas patologías afectan la calidad de vida del paciente y de la familia. Además, están relacionadas con la edad, de forma que a medida que la población media de países desarrollados envejece, la incidencia de dichas enfermedades aumenta, haciendo que el mercado para los inhibidores de POP sea mayor.

Por otro lado, la principal desventaja del proyecto de POP es que no es una diana validada. Pero la validación, a pesar de ser un reto, no es imposible.

CONCLUSIONES

Objetivo 1: Expresar DPP IV y realizar un estudio mediante NMR

1. La expresión recombinante de la DPP IV en células de insecto se ha realizado mediante la técnica del baculovirus. Se ha obtenido una proteína pura, dimérica y activa.
2. La proteína se ha utilizado como modelo para el estudio del patrón de glicosilación mediante espectrometría de masas. Sin embargo, no se pudo caracterizar las glicosilaciones de DPP IV.
3. Una metodología de marcaje selectivo se ha puesto a punto para DPP IV, consistente en el marcaje con ^{13}C del metilo de los residuos de metionina. La proteína marcada se ha estudiado por NMR utilizando el experimento TROSY-HSQC. Como resultado, se detectaron once de las catorce señales esperadas.
4. El estudio por NMR de la DPP IV en presencia de inhibidores ha sido realizado. Como resultado se ha detectado la aparición de una nueva señal en las muestras de DPP IV incubadas con inhibidores competitivos.

Objetivo 2: Descubrir inhibidores de DPP IV de origen botánico

1. Se ha testado una colección de extractos de plantas, de origen diverso. De esta colección, el extracto más prometedor es el extracto etanólico de la planta *Coutarea Latiflora* (AP-3). Fraccionamiento del extracto ha permitido la identificación de dos inhibidores de DPP IV: AP-3-a ($\text{IC}_{50} = 115 \mu\text{M}$) y AP-3-b ($\text{IC}_{50} = 340 \mu\text{M}$).
2. El mecanismo de inhibición sobre DPP IV de la molécula AP-3-a ha sido caracterizado como un inhibidor reversible con un comportamiento parabólico con una K_i de $158 \mu\text{M}$ (para el sitio competitivo).
3. Se ha realizado el estudio por NMR de la DPP IV inhibida por AP-3-a. En este caso, hay ausencia de la señal extra visualizada en los experimentos control del capítulo 1 de resultados. Una hipótesis para este fenómeno es que la señal extra es característica de inhibición de tipo competitivo, pero desaparece en tipos de inhibición más complejos, como el mecanismo parabólico que tiene AP-3-a.

Objetivo 3: Descubrir inhibidores de DPP IV mediante cribado de alto rendimiento

1. Se ha puesto a punto el ensayo de polarización de la fluorescencia para la identificación de ligandos de POP.
Se han testado un total de seis bibliotecas con 4,500 compuestos no tóxicos. Como resultado, se han obtenido 73 hits.
2. Los hits han sido clasificados en 16 clusters, y moléculas representativas de cada grupo han sido seleccionadas para ser validadas como inhibidores de POP. Cinco moléculas han inhibido POP con un porcentaje mayor del 80%.
3. Estas cinco moléculas, junto al compuesto con la mejor puntuación de docking, han sido seleccionadas para el cálculo del valor de IC_{50} . Así, las 6 moléculas se dividieron en dos grupos: grupo A, con una pendiente de IC_{50} acusada, y grupo B, con una pendiente más suave. Los compuestos más activos de cada grupo se seleccionaron para un análisis cinético. La molécula HTS-43 (grupo A, $IC_{50} = 6\mu M$) se ha definido como un inhibidor competitivo de POP con un valor de K_i de $3\mu M$. En cambio, el compuesto HTS-75 (grupo B, $IC_{50} = 3\mu M$) se ha caracterizado como un inhibidor parabólico de POP con un valor de K_i igual a $1.5\mu M$ (para el sitio competitivo). HTS-75 es el primer inhibidor parabólico de POP reportado.

Objetivo 4: Identificar inhibidores de las catepsinas L y B a partir de una biblioteca de química combinatoria

1. Se ha testado una nueva biblioteca de compuestos como inhibidores de las catepsinas L y B, de la que se ha identificado un potente inhibidor de la catepsina L (PAVS-20; $IC_{50} = 2.6\text{ nM}$)
2. El mecanismo de inhibición de PAVS-20 se ha caracterizado como un inhibidor covalente e irreversible, con valores de inhibición $K_{inac} = 20\text{ nM}$ y $k_{2nd} = 181.420\text{ s}^{-1}M^{-1}$.
3. Finalmente, las diferencias en inhibición de las catepsinas por parte de las moléculas de la biblioteca han sido explicadas mediante análisis de docking. La catepsina L presenta un subsitio S1 estrecho, de forma que sustituyentes voluminosos en P3 no son permitidos. Esto explica la diferencia en inhibición de catepsina L de las moléculas PAVS-19 ($IC_{50}/catL > 10^3\text{ nM}$) y PAVS-20 ($IC_{50}/catL = 2.6\text{ nM}$), que sólo difieren en este sustituyente.

BIBLIOGRAPHY

1. Rawlings, N. D., Barrett, A. J. and Bateman, A. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J Biol Chem.* 286, 38321-38328 (2011).
2. Rawlings, N. D., Barrett, A. J. and Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 40, D343-350 (2012).
3. Drag, M. and Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov.* 9, 690-701 (2010).
4. Turk, B., Turk, D. and Turk, V. Protease signalling: the cutting edge. *EMBO J.* 31, 1630-1643 (2012).
5. Turk, B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov.* 5, 785-799 (2006).
6. Skeggs, L. T., Jr., Kahn, J. R. and Shumway, N. P. The preparation and function of the hypertensin-converting enzyme. *J Exp Med.* 103, 295-299 (1956).
7. Thornberry, N. A., Rano, T. A., Peterson, E. P., Rasper, D. M., Timkey, T., Garcia-Calvo, M., Houtzager, V. M., Nordstrom, P. A., Roy, S., Vaillancourt, J. P., Chapman, K. T. and Nicholson, D. W. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J Biol Chem.* 272, 17907-17911 (1997).
8. McQuibban, G. A., Gong, J. H., Tam, E. M., McCulloch, C. A., Clark-Lewis, I. and Overall, C. M. Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3. *Science.* 289, 1202-1206 (2000).
9. Roberds, S. L., Anderson, J., Basi, G., Bienkowski, M. J., Branstetter, D. G., Chen, K. S., Freedman, S. B., Frigon, N. L., Games, D., Hu, K., Johnson-Wood, K., Kappenman, K. E., Kawabe, T. T., Kola, I., Kuehn, R., Lee, M., Liu, W., Motter, R., Nichols, N. F., Power, M., Robertson, D. W., Schenk, D., Schoor, M., Shopp, G. M., Shuck, M. E., Sinha, S., Svensson, K. A., Tatsuno, G., Tintrup, H., Wijsman, J., Wright, S. and McConlogue, L. BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. *Hum Mol Genet.* 10, 1317-1324 (2001).
10. Coburger, I., Hoefgen, S. and Than, M. E. The structural biology of the amyloid precursor protein APP - a complex puzzle reveals its multi-domain architecture. *Biol Chem.* 395, 485-498 (2014).
11. Macfarlane, R. G. An Enzyme Cascade in the Blood Clotting Mechanism, and Its Function as a Biochemical Amplifier. *Nature.* 202, 498-499 (1964).
12. Consortium, U. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191-198 (2014).
13. Ngo, J. K., Pomatto, L. C. and Davies, K. J. Upregulation of the mitochondrial Lon Protease allows adaptation to acute oxidative stress but dysregulation is associated with chronic stress, disease, and aging. *Redox Biol.* 1, 258-264 (2013).
14. Ogawa, T., Verhamme, I. M., Sun, M. F., Bock, P. E. and Gailani, D. Exosite-mediated substrate recognition of factor IX by factor XIa. The factor XIa heavy chain is required for initial recognition of factor IX. *J Biol Chem.* 280, 23523-23530 (2005).

15. Baugh, R. J., Dickinson, C. D., Ruf, W. and Krishnaswamy, S. Exosite interactions determine the affinity of factor X for the extrinsic Xase complex. *J Biol Chem.* 275, 28826-28833 (2000).
16. Krishnaswamy, S. and Betz, A. Exosites determine macromolecular substrate recognition by prothrombinase. *Biochemistry.* 36, 12080-12086 (1997).
17. Segers, K., Dahlback, B., Bock, P. E., Tans, G., Rosing, J. and Nicolaes, G. A. The role of thrombin exosites I and II in the activation of human coagulation factor V. *J Biol Chem.* 282, 33915-33924 (2007).
18. Fishman, M. C. and Porter, J. A. Pharmaceuticals: a new grammar for drug discovery. *Nature.* 437, 491-493 (2005).
19. de Almeida, H., Bastos, I. M., Ribeiro, B. M., Maigret, B. and Santana, J. M. New binding site conformations of the dengue virus NS3 protease accessed by molecular dynamics simulation. *PLoS One.* 8, e72402 (2013).
20. Nguyen, T. T., Lee, S., Wang, H. K., Chen, H. Y., Wu, Y. T., Lin, S. C., Kim, D. W. and Kim, D. In vitro evaluation of novel inhibitors against the NS2B-NS3 protease of dengue fever virus type 4. *Molecules.* 18, 15600-15612 (2013).
21. Zou, Y., Li, L., Chen, W., Chen, T., Ma, L., Wang, X., Xiong, B., Xu, Y. and Shen, J. Virtual screening and structure-based discovery of indole acylguanidines as potent beta-secretase (BACE1) inhibitors. *Molecules.* 18, 5706-5722 (2013).
22. Clark, D. E. What has virtual screening ever done for drug discovery? *Expert Opin Drug Discov.* 3, 841-851 (2008).
23. Shoichet, B. K. Virtual screening of chemical libraries. *Nature.* 432, 862-865 (2004).
24. Yuriev, E. and Ramsland, P. A. Latest developments in molecular docking: 2010-2011 in review. *J Mol Recognit.* 26, 215-239 (2013).
25. Yuriev, E. Challenges and advances in structure-based virtual screening. *Future Med Chem.* 6, 5-7 (2014).
26. Jhoti, H., Rees, S. and Solari, R. High-throughput screening and structure-based approaches to hit discovery: is there a clear winner? *Expert Opin Drug Discov.* 8, 1449-1453 (2013).
27. Chen, X., Chong, C. R., Shi, L., Yoshimoto, T., Sullivan, D. J., Jr. and Liu, J. O. Inhibitors of Plasmodium falciparum methionine aminopeptidase 1b possess antimalarial activity. *Proc Natl Acad Sci U S A.* 103, 14548-14553 (2006).
28. Chatterjee, A. K. Cell-based medicinal chemistry optimization of high-throughput screening (HTS) hits for orally active antimalarials. Part 1: challenges in potency and absorption, distribution, metabolism, excretion/pharmacokinetics (ADME/PK). *J Med Chem.* 56, 7741-7749 (2013).
29. Jones, A. J. and Avery, V. M. Whole-organism high-throughput screening against Trypanosoma brucei brucei. *Expert Opin Drug Discov.* 8, 495-507 (2013).

30. Fox, J. T. and Myung, K. Cell-based high-throughput screens for the discovery of chemotherapeutic agents. *Oncotarget*. 3, 581-585 (2012).
31. Davies, T. H., M. Fragment-Based Drug Discovery and X-Ray Crystallography. XII, (2012).
32. Reuters, T. Integrity Prous. (2014).
33. Jorgensen, W. L. The many roles of computation in drug discovery. *Science*. 303, 1813-1818 (2004).
34. Wadood, A., Riaz, M., Uddin, R. and Ul-Haq, Z. In silico identification and evaluation of leads for the simultaneous inhibition of protease and helicase activities of HCV NS3/4A protease using complex based pharmacophore mapping and virtual screening. *PLoS One*. 9, e89109 (2014).
35. Melagraki, G. and Afantitis, A. Ligand and structure based virtual screening strategies for hit-finding and optimization of hepatitis C virus (HCV) inhibitors. *Curr Med Chem*. 18, 2612-2619 (2011).
36. Schenone, M., Dancik, V., Wagner, B. K. and Clemons, P. A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol*. 9, 232-240 (2013).
37. Takebe, Y., Saucedo, C. J., Lund, G., Uenishi, R., Hase, S., Tsuchiura, T., Kneteman, N., Ramessar, K., Tyrrell, D. L., Shirakura, M., Wakita, T., McMahon, J. B. and O'Keefe, B. R. Antiviral lectins from red and blue-green algae show potent in vitro and in vivo activity against hepatitis C virus. *PLoS One*. 8, e64449 (2013).
38. Leite, P. E., de Almeida, K. B., Lagrota-Candido, J., Trindade, P., da Silva, R. F., Ribeiro, M. G., Lima-Araujo, K. G., Santos, W. C. and Quirico-Santos, T. Anti-inflammatory activity of *Eugenia punicifolia* extract on muscular lesion of mdx dystrophic mice. *J Cell Biochem*. 111, 1652-1660 (2010).
39. Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J. and Whisstock, J. C. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem*. 276, 33293-33296 (2001).
40. Martin, I., Teixido, M. and Giralt, E. Intracellular fate of peptide-mediated delivered cargoes. *Curr Pharm Des*. 19, 2924-2942 (2013).
41. Pujals, S., Sabido, E., Tarrago, T. and Giralt, E. all-D proline-rich cell-penetrating peptides: a preliminary in vivo internalization study. *Biochem Soc Trans*. 35, 794-796 (2007).
42. Farrera-Sinfreu, J., Giralt, E., Royo, M. and Albericio, F. Cell-penetrating proline-rich peptidomimetics. *Methods Mol Biol*. 386, 241-267 (2007).
43. Pujals, S. and Giralt, E. Proline-rich, amphipathic cell-penetrating peptides. *Adv Drug Deliv Rev*. 60, 473-484 (2008).
44. Teixido, M., Zurita, E., Mendieta, L., Oller-Salvia, B., Prades, R., Tarrago, T. and Giralt, E. Dual system for the central nervous system targeting and blood-brain barrier transport of a selective prolyl oligopeptidase inhibitor. *Biopolymers*. 100, 662-674 (2013).

45. Teixido, M., Zurita, E., Prades, R., Tarrago, T. and Giralt, E. A novel family of diketopiperazines as a tool for the study of transport across the blood-brain barrier (BBB) and their potential use as BBB-shuttles. *Adv Exp Med Biol.* 611, 227-228 (2009).
46. Prades, R., Guerrero, S., Araya, E., Molina, C., Salas, E., Zurita, E., Selva, J., Egea, G., Lopez-Iglesias, C., Teixido, M., Kogan, M. J. and Giralt, E. Delivery of gold nanoparticles to the brain by conjugation with a peptide that recognizes the transferrin receptor. *Biomaterials.* 33, 7194-7205 (2012).
47. Tarrago, T., Kichik, N., Claasen, B., Prades, R., Teixido, M. and Giralt, E. Baicalin, a prodrug able to reach the CNS, is a prolyl oligopeptidase inhibitor. *Bioorg Med Chem.* 16, 7516-7524 (2008).
48. LeCluyse, E. L., Witek, R. P., Andersen, M. E. and Powers, M. J. Organotypic liver culture models: meeting current challenges in toxicity testing. *Crit Rev Toxicol.* 42, 501-548 (2012).
49. Natesh, R., Schwager, S. L., Sturrock, E. D. and Acharya, K. R. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature.* 421, 551-554 (2003).
50. L.K., M. Handbook of Assay Development in Drug Discovery. (2006).
51. Meek, T. D. Inhibitors of HIV-1 protease. *J Enzyme Inhib.* 6, 65-98 (1992).
52. Mentlein, R. Dipeptidyl-peptidase IV (CD26)--role in the inactivation of regulatory peptides. *Regul Pept.* 85, 9-24 (1999).
53. Rasmussen, H. B., Branner, S., Wiberg, F. C. and Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nat Struct Biol.* 10, 19-25 (2003).
54. Lambeir, A. M., Durinx, C., Scharpe, S. and De Meester, I. Dipeptidyl-peptidase IV from bench to bedside: an update on structural properties, functions, and clinical aspects of the enzyme DPP IV. *Crit Rev Clin Lab Sci.* 40, 209-294 (2003).
55. Mentlein, R., Gallwitz, B. and Schmidt, W. E. Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. *Eur J Biochem.* 214, 829-835 (1993).
56. Gorrell, M. D., Gysbers, V. and McCaughan, G. W. CD26: a multifunctional integral membrane and secreted protein of activated lymphocytes. *Scand J Immunol.* 54, 249-264 (2001).
57. Baggio, L. L. and Drucker, D. J. Biology of incretins: GLP-1 and GIP. *Gastroenterology.* 132, 2131-2157 (2007).
58. Holst, J. J., Christensen, M., Lund, A., de Heer, J., Svendsen, B., Kielgast, U. and Knop, F. K. Regulation of glucagon secretion by incretins. *Diabetes Obes Metab.* 13 Suppl 1, 89-94 (2011).
59. Drucker, D. J., Erlich, P., Asa, S. L. and Brubaker, P. L. Induction of intestinal epithelial proliferation by glucagon-like peptide 2. *Proc Natl Acad Sci U S A.* 93, 7911-7916 (1996).

60. Sakurai, Y., Shintani, N., Hayata, A., Hashimoto, H. and Baba, A. Trophic effects of PACAP on pancreatic islets: a mini-review. *J Mol Neurosci.* 43, 3-7 (2011).
61. Chu, J. Y., Cheng, C. Y., Lee, V. H., Chan, Y. S. and Chow, B. K. Secretin and body fluid homeostasis. *Kidney Int.* 79, 280-287 (2011).
62. Oravecz, T., Pall, M., Roderiquez, G., Gorrell, M. D., Ditto, M., Nguyen, N. Y., Boykins, R., Unsworth, E. and Norcross, M. A. Regulation of the receptor specificity and function of the chemokine RANTES (regulated on activation, normal T cell expressed and secreted) by dipeptidyl peptidase IV (CD26)-mediated cleavage. *J Exp Med.* 186, 1865-1872 (1997).
63. Yazbeck, R., Howarth, G. S. and Abbott, C. A. Dipeptidyl peptidase inhibitors, an emerging drug class for inflammatory disease? *Trends Pharmacol Sci.* 30, 600-607 (2009).
64. Kos, K., Baker, A. R., Jernas, M., Harte, A. L., Clapham, J. C., O'Hare, J. P., Carlsson, L., Kumar, S. and McTernan, P. G. DPP-IV inhibition enhances the antilipolytic action of NPY in human adipose tissue. *Diabetes Obes Metab.* 11, 285-292 (2009).
65. Batterham, R. L. and Bloom, S. R. The gut hormone peptide YY regulates appetite. *Ann N Y Acad Sci.* 994, 162-168 (2003).
66. O'Connor, T. M., O'Connell, J., O'Brien, D. I., Goode, T., Bredin, C. P. and Shanahan, F. The role of substance P in inflammatory disease. *J Cell Physiol.* 201, 167-180 (2004).
67. Barreto, S. G., Carati, C. J., Toouli, J. and Saccone, G. T. The islet-acinar axis of the pancreas: more than just insulin. *Am J Physiol Gastrointest Liver Physiol.* 299, G10-22 (2010).
68. Karin, N. The multiple faces of CXCL12 (SDF-1 α) in the regulation of immunity during health and disease. *J Leukoc Biol.* 88, 463-473 (2010).
69. Scanlan, M. J., Raj, B. K., Calvo, B., Garin-Chesa, P., Sanz-Moncasi, M. P., Healey, J. H., Old, L. J. and Rettig, W. J. Molecular cloning of fibroblast activation protein alpha, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers. *Proc Natl Acad Sci U S A.* 91, 5657-5661 (1994).
70. Lee, K. N., Jackson, K. W., Christiansen, V. J., Chung, K. H. and McKee, P. A. A novel plasma proteinase potentiates alpha2-antiplasmin inhibition of fibrin digestion. *Blood.* 103, 3783-3788 (2004).
71. Abbott, C. A., McCaughan, G. W., Levy, M. T., Church, W. B. and Gorrell, M. D. Binding to human dipeptidyl peptidase IV by adenosine deaminase and antibodies that inhibit ligand binding involves overlapping, discontinuous sites on a predicted beta propeller domain. *Eur J Biochem.* 266, 798-810 (1999).
72. Ishii, T., Ohnuma, K., Murakami, A., Takasawa, N., Kobayashi, S., Dang, N. H., Schlossman, S. F. and Morimoto, C. CD26-mediated signaling for T cell activation occurs in lipid rafts through its association with CD45RO. *Proc Natl Acad Sci U S A.* 98, 12138-12143 (2001).
73. Ohnuma, K., Uchiyama, M., Yamochi, T., Nishibashi, K., Hosono, O., Takahashi, N., Kina, S., Tanaka, H., Lin, X., Dang, N. H. and Morimoto, C. Caveolin-1 triggers T-cell activation via CD26 in association with CARMA1. *J Biol Chem.* 282, 10117-10131 (2007).

74. Ohnuma, K., Yamochi, T., Uchiyama, M., Nishibashi, K., Yoshikawa, N., Shimizu, N., Iwata, S., Tanaka, H., Dang, N. H. and Morimoto, C. CD26 up-regulates expression of CD86 on antigen-presenting cells by means of caveolin-1. *Proc Natl Acad Sci U S A.* 101, 14186-14191 (2004).
75. Cheng, H. C., Abdel-Ghany, M. and Pauli, B. U. A novel consensus motif in fibronectin mediates dipeptidyl peptidase IV adhesion and metastasis. *J Biol Chem.* 278, 24600-24607 (2003).
76. Gonzalez-Gronow, M., Grenett, H. E., Weber, M. R., Gawdi, G. and Pizzo, S. V. Interaction of plasminogen with dipeptidyl peptidase IV initiates a signal transduction mechanism which regulates expression of matrix metalloproteinase-9 by prostate cancer cells. *Biochem J.* 355, 397-407 (2001).
77. Girardi, A. C., Degray, B. C., Nagy, T., Biemesderfer, D. and Aronson, P. S. Association of Na(+)-H(+) exchanger isoform NHE3 and dipeptidyl peptidase IV in the renal proximal tubule. *J Biol Chem.* 276, 46671-46677 (2001).
78. Davoodi, J., Kelly, J., Gendron, N. H. and MacKenzie, A. E. The Simpson-Golabi-Behmel syndrome causative glypican-3, binds to and inhibits the dipeptidyl peptidase activity of CD26. *Proteomics.* 7, 2300-2310 (2007).
79. Shimizu, Y. and Shaw, S. Lymphocyte interactions with extracellular matrix. *FASEB J.* 5, 2292-2299 (1991).
80. Yu, D. M., Yao, T. W., Chowdhury, S., Nadvi, N. A., Osborne, B., Church, W. B., McCaughan, G. W. and Gorrell, M. D. The dipeptidyl peptidase IV family in cancer and cell biology. *FEBS J.* 277, 1126-1144 (2010).
81. Martin, M., Huguet, J., Centelles, J. J. and Franco, R. Expression of ecto-adenosine deaminase and CD26 in human T cells triggered by the TCR-CD3 complex. Possible role of adenosine deaminase as costimulatory molecule. *J Immunol.* 155, 4630-4643 (1995).
82. Ahren, B., Simonsson, E., Larsson, H., Landin-Olsson, M., Torgeirsson, H., Jansson, P. A., Sandqvist, M., Bavenholm, P., Efendic, S., Eriksson, J. W., Dickinson, S. and Holmes, D. Inhibition of dipeptidyl peptidase IV improves metabolic control over a 4-week study period in type 2 diabetes. *Diabetes Care.* 25, 869-875 (2002).
83. Gerich, J. DPP-4 inhibitors: what may be the clinical differentiators? *Diabetes Res Clin Pract.* 90, 131-140 (2010).
84. Davidson, J. A. Incorporating incretin-based therapies into clinical practice: differences between glucagon-like Peptide 1 receptor agonists and dipeptidyl peptidase 4 inhibitors. *Mayo Clin Proc.* 85, S27-37 (2010).
85. Gallwitz, B. Small molecule dipeptidylpeptidase IV inhibitors under investigation for diabetes mellitus therapy. *Expert Opin Investig Drugs.* 20, 723-732 (2011).
86. Mendieta, L., Tarrago, T. and Giralt, E. Recent patents of dipeptidyl peptidase IV inhibitors. *Expert Opin Ther Pat.* 21, 1693-1741 (2011).
87. Pro, B. and Dang, N. H. CD26/dipeptidyl peptidase IV and its role in cancer. *Histol Histopathol.* 19, 1345-1351 (2004).

88. Sedo, A., Duke-Cohan, J. S., Balaziová, E. and Sedová, L. R. Dipeptidyl peptidase IV activity and/or structure homologs: contributing factors in the pathogenesis of rheumatoid arthritis? *Arthritis Res Ther.* 7, 253-269 (2005).
89. Ospelt, C., Mertens, J. C., Jungel, A., Brentano, F., Maciejewska-Rodriguez, H., Huber, L. C., Hemmatzad, H., Wuest, T., Knuth, A., Gay, R. E., Michel, B. A., Gay, S., Renner, C. and Bauer, S. Inhibition of fibroblast activation protein and dipeptidylpeptidase 4 increases cartilage invasion by rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum.* 62, 1224-1235 (2010).
90. Stulc, T. and Sedo, A. Inhibition of multifunctional dipeptidyl peptidase-IV: is there a risk of oncological and immunological adverse effects? *Diabetes Res Clin Pract.* 88, 125-131 (2010).
91. Umezawa, H., Aoyagi, T., Ogawa, K., Naganawa, H., Hamada, M. and Takeuchi, T. Diprotins A and B, inhibitors of dipeptidyl aminopeptidase IV, produced by bacteria. *J Antibiot (Tokyo).* 37, 422-425 (1984).
92. Flentke, G. R., Munoz, E., Huber, B. T., Plaut, A. G., Kettner, C. A. and Bachovchin, W. W. Inhibition of dipeptidyl aminopeptidase IV (DP-IV) by Xaa-boroPro dipeptides and use of these inhibitors to examine the role of DP-IV in T-cell function. *Proc Natl Acad Sci U S A.* 88, 1556-1559 (1991).
93. Li, J., Wilk, E. and Wilk, S. Aminoacylpyrrolidine-2-nitriles: potent and stable inhibitors of dipeptidyl-peptidase IV (CD 26). *Arch Biochem Biophys.* 323, 148-154 (1995).
94. Augustyns, K., Van der Veken, P., Senten, K. and Haemers, A. The therapeutic potential of inhibitors of dipeptidyl peptidase IV (DPP IV) and related proline-specific dipeptidyl aminopeptidases. *Curr Med Chem.* 12, 971-998 (2005).
95. A/S, N. N. WO 02/02560 A2. (2002).
96. KG, B. I. P. WO 02/068420 A1. (2002).
97. Merck. US 6,699,871 B2. (2004).
98. Takeda. WO 2005/095381 A1. (2005).
99. Lopez, A., Tarrago, T. and Giralt, E. Low molecular weight inhibitors of Prolyl Oligopeptidase: a review of compounds patented from 2003 to 2010. *Expert Opin Ther Pat.* 21, 1023-1044 (2011).
100. Fulop, V., Bocskei, Z. and Polgar, L. Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell.* 94, 161-170 (1998).
101. Kichik, N., Tarrago, T., Claasen, B., Gairi, M., Millet, O. and Giralt, E. 15N relaxation NMR studies of prolyl oligopeptidase, an 80 kDa enzyme, reveal a pre-existing equilibrium between different conformational states. *Chembiochem.* 12, 2737-2739 (2011).
102. Li, M., Chen, C., Davies, D. R. and Chiu, T. K. Induced-fit mechanism for prolyl endopeptidase. *J Biol Chem.* 285, 21487-21495 (2010).

103. Moreno-Baylach, M. J., Felipo, V., Mannisto, P. T. and Garcia-Horsman, J. A. Expression and traffic of cellular prolyl oligopeptidase are regulated during cerebellar granule cell differentiation, maturation, and aging. *Neuroscience*. 156, 580-585 (2008).
104. Di Daniel, E., Glover, C. P., Grot, E., Chan, M. K., Sanderson, T. H., White, J. H., Ellis, C. L., Gallagher, K. T., Uney, J., Thomas, J., Maycox, P. R. and Mudge, A. W. Prolyl oligopeptidase binds to GAP-43 and functions without its peptidase activity. *Mol Cell Neurosci*. 41, 373-382 (2009).
105. Szeltner, Z., Morawski, M., Juhasz, T., Szamosi, I., Liliom, K., Csizmok, V., Tolgyesi, F. and Polgar, L. GAP43 shows partial co-localisation but no strong physical interaction with prolyl oligopeptidase. *Biochim Biophys Acta*. 1804, 2162-2176 (2010).
106. Lopez, A., Mendieta, L., Prades, R., Royo, S., Tarrago, T. and Giralt, E. Peptide POP inhibitors for the treatment of the cognitive symptoms of schizophrenia. *Future Med Chem*. 5, 1509-1523 (2013).
107. Toide, K., Okamiya, K., Iwamoto, Y. and Kato, T. Effect of a novel prolyl endopeptidase inhibitor, JTP-4819, on prolyl endopeptidase activity and substance P- and arginine-vasopressin-like immunoreactivity in the brains of aged rats. *J Neurochem*. 65, 234-240 (1995).
108. Reichelt, K. L., Hole, K., Hamberger, A., Saelid, G., Edminson, P. D., Braestrup, C. B., Lingjaerde, O., Ledaal, P. and Orbeck, H. Biologically active peptide-containing fractions in schizophrenia and childhood autism. *Adv Biochem Psychopharmacol*. 28, 627-643 (1981).
109. Maes, M., Goossens, F., Scharpe, S., Meltzer, H. Y., D'Hondt, P. and Cosyns, P. Lower serum prolyl endopeptidase enzyme activity in major depression: further evidence that peptidases play a role in the pathophysiology of depression. *Biol Psychiatry*. 35, 545-552 (1994).
110. Maes, M., Goossens, F., Scharpe, S., Calabrese, J., Desnyder, R. and Meltzer, H. Y. Alterations in plasma prolyl endopeptidase activity in depression, mania, and schizophrenia: effects of antidepressants, mood stabilizers, and antipsychotic drugs. *Psychiatry Res*. 58, 217-225 (1995).
111. Jalkanen, A. J., Piepponen, T. P., Hakkarainen, J. J., De Meester, I., Lambeir, A. M. and Forsberg, M. M. The effect of prolyl oligopeptidase inhibition on extracellular acetylcholine and dopamine levels in the rat striatum. *Neurochem Int*. 60, 301-309 (2012).
112. Williams, R. S., Eames, M., Ryves, W. J., Viggars, J. and Harwood, A. J. Loss of a prolyl oligopeptidase confers resistance to lithium by elevation of inositol (1,4,5) trisphosphate. *EMBO J*. 18, 2734-2745 (1999).
113. Komatsu, Y. GABAB receptors, monoamine receptors, and postsynaptic inositol trisphosphate-induced Ca²⁺ release are involved in the induction of long-term potentiation at visual cortical inhibitory synapses. *J Neurosci*. 16, 6342-6352 (1996).
114. Harwood, A. J. and Agam, G. Search for a common mechanism of mood stabilizers. *Biochem Pharmacol*. 66, 179-189 (2003).
115. Brandt, I., Gerard, M., Sergeant, K., Devreese, B., Baekelandt, V., Augustyns, K., Scharpe, S., Engelborghs, Y. and Lambeir, A. M. Prolyl oligopeptidase stimulates the aggregation of alpha-synuclein. *Peptides*. 29, 1472-1478 (2008).

116. Larrinaga, G., Perez, I., Blanco, L., Lopez, J. I., Andres, L., Etxezarraga, C., Santaolalla, F., Zabala, A., Varona, A. and Irazusta, J. Increased prolyl endopeptidase activity in human neoplasia. *Regul Pept.* 163, 102-106 (2010).
117. Myohanen, T. T., Tenorio-Laranga, J., Jokinen, B., Vazquez-Sanchez, R., Moreno-Baylach, M. J., Garcia-Horsman, J. A. and Mannisto, P. T. Prolyl oligopeptidase induces angiogenesis both in vitro and in vivo in a novel regulatory manner. *Br J Pharmacol.* 163, 1666-1678 (2011).
118. Mach, L., Mort, J. S. and Glossl, J. Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. *J Biol Chem.* 269, 13036-13040 (1994).
119. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S. and Cygler, M. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* 15, 5492-5503 (1996).
120. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. and et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* 10, 2321-2330 (1991).
121. Xing, R., Addington, A. K. and Mason, R. W. Quantification of cathepsins B and L in cells. *Biochem J.* 332 (Pt 2), 499-505 (1998).
122. Palermo, C. and Joyce, J. A. Cysteine cathepsin proteases as pharmacological targets in cancer. *Trends Pharmacol Sci.* 29, 22-28 (2008).
123. Shaw, E., Wikstrom, P. and Ruscica, J. An exploration of the primary specificity site of cathepsin B. *Arch Biochem Biophys.* 222, 424-429 (1983).
124. Towatari, T. and Katunuma, N. Selective cleavage of peptide bonds by cathepsins L and B from rat liver. *J Biochem.* 93, 1119-1128 (1983).
125. Lecaille F, K. J., Brömme D. Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chem Rev.* 102, 30 (2002).
126. Vicik, R., Busemann, M., Gelhaus, C., Stiefl, N., Scheiber, J., Schmitz, W., Schulz, F., Mladenovic, M., Engels, B., Leippe, M., Baumann, K. and Schirmeister, T. Aziridide-based inhibitors of cathepsin L: synthesis, inhibition activity, and docking studies. *ChemMedChem.* 1, 1126-1141 (2006).
127. Soderstrom, M., Salminen, H., Glumoff, V., Kirschke, H., Aro, H. and Vuorio, E. Cathepsin expression during skeletal development. *Biochim Biophys Acta.* 1446, 35-46 (1999).
128. Joyce JA, B. A., Chegade K, Meyer-Morse N, Giraudo E, Tsai FY, Greenbaum DC, Hager JH, Bogyo M, Hanahan D. Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell.* 5, 10 (2004).
129. Joyce JA, H. D. Multiple roles for cysteine cathepsins in cancer. *Cell Cycle.* 3, 3 (2004).

130. Gocheva, V., Zeng, W., Ke, D., Klimstra, D., Reinheckel, T., Peters, C., Hanahan, D. and Joyce, J. A. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes Dev.* 20, 543-556 (2006).
131. Cunnane, G., FitzGerald, O., Hummel, K. M., Gay, R. E., Gay, S. and Bresnihan, B. Collagenase, cathepsin B and cathepsin L gene expression in the synovial membrane of patients with early inflammatory arthritis. *Rheumatology (Oxford)*. 38, 34-42 (1999).
132. Lemaire, R., Huet, G., Zerimech, F., Grard, G., Fontaine, C., Duquesnoy, B. and Flipo, R. M. Selective induction of the secretion of cathepsins B and L by cytokines in synovial fibroblast-like cells. *Br J Rheumatol.* 36, 735-743 (1997).
133. Lopez-Castejon, G., Theaker, J., Pelegrin, P., Clifton, A. D., Braddock, M. and Surprenant, A. P2X(7) receptor-mediated release of cathepsins from macrophages is a cytokine-independent mechanism potentially involved in joint diseases. *J Immunol.* 185, 2611-2619 (2010).
134. Kolkhorst, V., Sturzebecher, J. and Wiederanders, B. Inhibition of tumour cell invasion by protease inhibitors: correlation with the protease profile. *J Cancer Res Clin Oncol.* 124, 598-606 (1998).
135. Hiramatsu, H., Kyono, K., Shima, H., Fukushima, C., Sugiyama, S., Inaka, K., Yamamoto, A. and Shimizu, R. Crystallization and preliminary X-ray study of human dipeptidyl peptidase IV (DPPIV). *Acta Crystallogr D Biol Crystallogr.* 59, 595-596 (2003).
136. Iwaki-Egawa, S., Watanabe, Y., Kikuya, Y. and Fujimoto, Y. Dipeptidyl peptidase IV from human serum: purification, characterization, and N-terminal amino acid sequence. *J Biochem.* 124, 428-433 (1998).
137. Durinx, C., Lambeir, A. M., Bosmans, E., Falmagne, J. B., Berghmans, R., Haemers, A., Scharpe, S. and De Meester, I. Molecular characterization of dipeptidyl peptidase activity in serum: soluble CD26/dipeptidyl peptidase IV is responsible for the release of X-Pro dipeptides. *Eur J Biochem.* 267, 5608-5613 (2000).
138. Aertgeerts, K., Ye, S., Shi, L., Prasad, S. G., Witmer, D., Chi, E., Sang, B. C., Wijnands, R. A., Webb, D. R. and Swanson, R. V. N-linked glycosylation of dipeptidyl peptidase IV (CD26): effects on enzyme activity, homodimer formation, and adenosine deaminase binding. *Protein Sci.* 13, 145-154 (2004).
139. Thoma, R., Loffler, B., Stihle, M., Huber, W., Ruf, A. and Hennig, M. Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV. *Structure.* 11, 947-959 (2003).
140. Baer, J. W., Gerhartz, B., Hoffmann, T., Rosche, F. and Demuth, H. U. Characterisation of human DP IV produced by a *Pichia pastoris* expression system. *Adv Exp Med Biol.* 524, 103-108 (2003).
141. Smith, G. E., Summers, M. D. and Fraser, M. J. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol.* 3, 2156-2165 (1983).
142. Assenberg, R., Wan, P. T., Geisse, S. and Mayr, L. M. Advances in recombinant protein expression for use in pharmaceutical research. *Curr Opin Struct Biol.* 23, 393-402 (2013).

143. Dobers, J., Zimmermann-Kordmann, M., Leddermann, M., Schewe, T., Reutter, W. and Fan, H. Expression, purification, and characterization of human dipeptidyl peptidase IV/CD26 in Sf9 insect cells. *Protein Expr Purif.* 25, 527-532 (2002).
144. Jarvis, D. L. Baculovirus-insect cell expression systems. *Methods Enzymol.* 463, 191-222 (2009).
145. Weihofen, W. A., Liu, J., Reutter, W., Saenger, W. and Fan, H. Crystal structure of CD26/dipeptidyl-peptidase IV in complex with adenosine deaminase reveals a highly amphiphilic interface. *J Biol Chem.* 279, 43330-43335 (2004).
146. Kawasaki, N., Lin, C. W., Inoue, R., Khoo, K. H., Ma, B. Y., Oka, S., Ishiguro, M., Sawada, T., Ishida, H., Hashimoto, T. and Kawasaki, T. Highly fucosylated N-glycan ligands for mannan-binding protein expressed specifically on CD26 (DPPVI) isolated from a human colorectal carcinoma cell line, SW1116. *Glycobiology.* 19, 437-450 (2009).
147. Henzler-Wildman, K. and Kern, D. Dynamic personalities of proteins. *Nature.* 450, 964-972 (2007).
148. Tarrago, T., Martin-Benito, J., Sabido, E., Claasen, B., Madurga, S., Gairi, M., Valpuesta, J. M. and Giral, E. A new side opening on prolyl oligopeptidase revealed by electron microscopy. *FEBS Lett.* 583, 3344-3348 (2009).
149. Ferentz, A. E. and Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q Rev Biophys.* 33, 29-65 (2000).
150. Meyer, B. and Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl.* 42, 864-890 (2003).
151. Takahashi, H. and Shimada, I. Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J Biomol NMR.* 46, 3-10 (2010).
152. Fan, Y., Shi, L., Ladizhansky, V. and Brown, L. S. Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment. *J Biomol NMR.* 49, 151-161 (2011).
153. Sugiki, T., Shimada, I. and Takahashi, H. Stable isotope labeling of protein by *Kluyveromyces lactis* for NMR study. *J Biomol NMR.* 42, 159-162 (2008).
154. Gossert, A. D. and Jahnke, W. Isotope labeling in insect cells. *Adv Exp Med Biol.* 992, 179-196 (2012).
155. Sastry, M., Xu, L., Georgiev, I. S., Bewley, C. A., Nabel, G. J. and Kwong, P. D. Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy. *J Biomol NMR.* 50, 197-207 (2011).
156. Matsuda, T., Watanabe, S. and Kigawa, T. Cell-free synthesis system suitable for disulfide-containing proteins. *Biochem Biophys Res Commun.* 431, 296-301 (2013).
157. Vinarov, D. A., Newman, C. L., Tyler, E. M., Markley, J. L. and Shahan, M. N. Wheat germ cell-free expression system for protein production. *Curr Protoc Protein Sci.* Chapter 5, Unit 5 18 (2006).

158. Breukels, V., Konijnenberg, A., Nabuurs, S. M., Doreleijers, J. F., Kovalevskaya, N. V. and Vuister, G. W. Overview on the use of NMR to examine protein structure. *Curr Protoc Protein Sci.* Chapter 17, Unit17 15 (2011).
159. Rasia, R. M., Lescop, E., Palatnik, J. F., Boisbouvier, J. and Brutscher, B. Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J Biomol NMR.* 51, 369-378 (2011).
160. Roy Choudhury, A., Perdih, A., Zuperl, S., Sikorska, E., Solmajer, T., Jurga, S., Zhukov, I. and Novic, M. Structural elucidation of transmembrane transporter protein bilirubin translocase: conformational analysis of the second transmembrane region TM2 by molecular dynamics and NMR spectroscopy. *Biochim Biophys Acta.* 1828, 2609-2619 (2013).
161. Aberle, D., Muhle-Goll, C., Burck, J., Wolf, M., Reisser, S., Luy, B., Wenzel, W., Ulrich, A. S. and Meyers, G. Structure of the Membrane Anchor of Pestivirus Glycoprotein E(rns), a Long Tilted Amphipathic Helix. *PLoS Pathog.* 10, e1003973 (2014).
162. Liu, Z., Gong, Z., Guo, D. C., Zhang, W. P. and Tang, C. Subtle Dynamics of holo Glutamine Binding Protein Revealed with a Rigid Paramagnetic Probe. *Biochemistry.* (2014).
163. Vogeli, B., Kazemi, S., Guntert, P. and Riek, R. Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat Struct Mol Biol.* 19, 1053-1057 (2012).
164. Aramini, J. M., Hamilton, K., Ma, L. C., Swapna, G. V., Leonard, P. G., Ladbury, J. E., Krug, R. M. and Montelione, G. T. F NMR Reveals Multiple Conformations at the Dimer Interface of the Nonstructural Protein 1 Effector Domain from Influenza A Virus. *Structure.* (2014).
165. Reichheld, S. E., Muiznieks, L. D., Stahl, R., Simonetti, K., Sharpe, S. and Keeley, F. W. Conformational Transitions of the Crosslinking Domains of Elastin During Self-assembly. *J Biol Chem.* (2014).
166. Munari, F., Gajda, M. J., Hiragami-Hamada, K., Fischle, W. and Zweckstetter, M. Characterization of the effects of phosphorylation by CK2 on the structure and binding properties of human HP1beta. *FEBS Lett.* (2014).
167. Brister, M. A., Pandey, A. K., Bielska, A. A. and Zondlo, N. J. O-GlcNAcylation and Phosphorylation Have Opposing Structural Effects in tau: Phosphothreonine Induces Particular Conformational Order. *J Am Chem Soc.* (2014).
168. Konuma, T., Lee, Y. H., Goto, Y. and Sakurai, K. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins.* 81, 107-118 (2013).
169. Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A.* 94, 12366-12371 (1997).

170. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. and Kay, L. E. Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc.* 125, 10420-10428 (2003).
171. Pederson, R. A., White, H. A., Schlenzig, D., Pauly, R. P., McIntosh, C. H. and Demuth, H. U. Improved glucose tolerance in Zucker fatty rats by oral administration of the dipeptidyl peptidase IV inhibitor isoleucine thiazolidide. *Diabetes.* 47, 1253-1258 (1998).
172. Girardi, A. C., Knauf, F., Demuth, H. U. and Aronson, P. S. Role of dipeptidyl peptidase IV in regulating activity of Na^+/H^+ exchanger isoform NHE3 in proximal tubule cells. *Am J Physiol Cell Physiol.* 287, C1238-1245 (2004).
173. Hughes, T. E., Mone, M. D., Russell, M. E., Weldon, S. C. and Villhauer, E. B. NVP-DPP728 (1-[[[2-[(5-cyanopyridin-2-yl)amino]ethyl]amino]acetyl]-2-cyano-(S)-pyrrolidine), a slow-binding inhibitor of dipeptidyl peptidase IV. *Biochemistry.* 38, 11597-11603 (1999).
174. Al-masri, I. M., Mohammad, M. K. and Tahaa, M. O. Inhibition of dipeptidyl peptidase IV (DPP IV) is one of the mechanisms explaining the hypoglycemic effect of berberine. *J Enzyme Inhib Med Chem.* 24, 1061-1066 (2009).
175. Ganesan, A. The impact of natural products upon modern drug discovery. *Curr Opin Chem Biol.* 12, 306-317 (2008).
176. Newman, D. J. and Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod.* 75, 311-335 (2012).
177. Prabhakar, P. K. and Doble, M. A target based therapeutic approach towards diabetes mellitus using medicinal plants. *Curr Diabetes Rev.* 4, 291-308 (2008).
178. Hung, H. Y., Qian, K., Morris-Natschke, S. L., Hsu, C. S. and Lee, K. H. Recent discovery of plant-derived anti-diabetic natural products. *Nat Prod Rep.* 29, 580-606 (2012).
179. Li, W. L., Zheng, H. C., Bukuru, J. and De Kimpe, N. Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus. *J Ethnopharmacol.* 92, 1-21 (2004).
180. Chongyun Liu, A. T. Chinese herbal medicine: modern applications of traditional formulas. (2003).
181. Jia, W., Gao, W. and Tang, L. Antidiabetic herbal drugs officially approved in China. *Phytother Res.* 17, 1127-1134 (2003).
182. Balbani, A. P., Silva, D. H. and Montovani, J. C. Patents of drugs extracted from Brazilian medicinal plants. *Expert Opin Ther Pat.* 19, 461-473 (2009).
183. Farias, R. A., Rao, V. S., Viana, G. S., Silveira, E. R., Maciel, M. A. and Pinto, A. C. Hypoglycemic effect of trans-dehydrocrotonin, a nor-clerodane diterpene from *Croton cajucara*. *Planta Med.* 63, 558-560 (1997).
184. Graim, J. F., Lopes Filho Gde, J., Brito, M. V. and Brasil Matos, L. T. Histologic evaluation of rats' liver after *Croton cajucara* Benth (sacaca) administration. *Acta Cir Bras.* 23, 130-134 (2008).

185. Rodrigues, G., Marcolin, E., Bona, S., Porawski, M., Lehmann, M. and Marroni, N. P. Hepatic alterations and genotoxic effects of *Croton cajucara* Benth (SACACA) in diabetic rats. *Arq Gastroenterol.* 47, 301-305 (2010).
186. Simoes, C. M. O. Plantas da Medicina Popular no Rio Grande do Sul. (1988).
187. Harri Lorenzi, F. J. d. A. M. Plantas medicinais no Brasil: nativas e exóticas. (2002).
188. Pinto, A., Capasso, A. and Sorrentino, L. Experimental animal studies on the hypoglycemic effects of a Copalchi extract. *Arzneimittelforschung.* 47, 829-833 (1997).
189. Korec, R., Heinz Sensch, K. and Zoukas, T. Effects of the neoflavonoid coumestrol, one of the antidiabetic active substances of *Hintonia latiflora*, on streptozotocin-induced diabetes mellitus in rats. *Arzneimittelforschung.* 50, 122-128 (2000).
190. Guerrero-Analco, J., Medina-Campos, O., Brindis, F., Bye, R., Pedraza-Chaverri, J., Navarrete, A. and Mata, R. Antidiabetic properties of selected Mexican copalchis of the Rubiaceae family. *Phytochemistry.* 68, 2087-2095 (2007).
191. Cristians, S., Guerrero-Analco, J. A., Perez-Vasquez, A., Palacios-Espinosa, F., Ciangherotti, C., Bye, R. and Mata, R. Hypoglycemic activity of extracts and compounds from the leaves of *Hintonia standleyana* and *H. latiflora*: potential alternatives to the use of the stem bark of these species. *J Nat Prod.* 72, 408-413 (2009).
192. Mata, R., Cristians, S., Escandon-Rivera, S., Juarez-Reyes, K. and Rivero-Cruz, I. Mexican antidiabetic herbs: valuable sources of inhibitors of alpha-glucosidases. *J Nat Prod.* 76, 468-483 (2013).
193. Andrade-Cetto, A. and Heinrich, M. Mexican plants with hypoglycaemic effect used in the treatment of diabetes. *J Ethnopharmacol.* 99, 325-348 (2005).
194. Pharma, G. G. C. K. Process of manufacturing an extract from the Copalchi stem barks and the use of neoflavonoids for the treatment of diabetes. (1999).
195. Feng, B. Y. and Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc.* 1, 550-553 (2006).
196. Kren, V. and Martinkova, L. Glycosides in medicine: "The role of glycosidic residue in biological activity". *Curr Med Chem.* 8, 1303-1328 (2001).
197. Lorey, S., Stockel-Maschek, A., Faust, J., Brandt, W., Stiebitz, B., Gorrell, M. D., Kahne, T., Mrestani-Klaus, C., Wrenger, S., Reinhold, D., Ansorge, S. and Neubert, K. Different modes of dipeptidyl peptidase IV (CD26) inhibition by oligopeptides derived from the N-terminus of HIV-1 Tat indicate at least two inhibitor binding sites. *Eur J Biochem.* 270, 2147-2156 (2003).
198. Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U. and Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov.* 10, 188-195 (2011).
199. Fox, S., Farr-Jones, S., Sopchak, L., Boggs, A., Nicely, H. W., Khoury, R. and Biro, M. High-throughput screening: update on practices and success. *J Biomol Screen.* 11, 864-869 (2006).

200. Bleicher, K. H., Bohm, H. J., Muller, K. and Alanine, A. I. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov.* 2, 369-378 (2003).
201. Ashburn, T. T. and Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 3, 673-683 (2004).
202. E., S. Identificació de proteases mitjançant proteòmica basada en activitat. (2008).
203. Yasuma T, O. S., Choh N, Nomura T, Furuyama N, Nishimura A, Fujisawa Y, Sohda T. Synthesis of peptide aldehyde derivatives as selective inhibitors of human cathepsin L and their inhibitory effect on bone resorption. *J Med Chem.* 41, 7 (1998).
204. Yamamoto M, I. S., Kondo H, Inoue S. Design and synthesis of dual inhibitors for matrix metalloproteinase and cathepsin. *Bioorg Med Chem Lett.* 12, 3 (2002).
205. Falguyret JP, O. R., Okamoto O, Wesolowski G, Aubin Y, Rydzewski RM, Prasit P, Riendeau D, Rodan SB, Percival MD. Novel, nonpeptidic cyanamides as potent and reversible inhibitors of human cathepsins K and L. *J Med Chem.* . 44, 10 (2001).
206. Greenspan PD, C. K., Tommasi RA, Cowen SD, McQuire LW, Farley DL, van Duzer JH, Goldberg RL, Zhou H, Du Z, Fitt JJ, Coppa DE, Fang Z, Macchia W, Zhu L, Capparelli MP, Goldstein R, Wigg AM, Doughty JR, Bohacek RS, Knap AK. Identification of dipeptidyl nitriles as potent and selective inhibitors of cathepsin B through structure-based drug design. *J Med Chem.* 44, 4524 (2001).
207. Brömme D, K. J., Okamoto K, Rasnick D, Palmer JT. Peptidyl vinyl sulphones: a new class of potent and selective cysteine protease inhibitors: S2P2 specificity of human cathepsin O2 in comparison with cathepsins S and L. *Biochem J.* 315, 5 (1996).
208. McKerrow JH, E. J., Caffrey CR. Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med Chem.* 7, 6 (1999).
209. Engel JC, D. P., Hsieh I, McKerrow JH. Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection. *J Exp Med.* 188, 10 (1998).
210. Pauly TA, S. T., Ammirati M, Sivaraman J, Danley DE, Griffor MC, Kamath AV, Wang IK, Laird ER, Seddon AP, Ménard R, Cygler M, Rath VL. Specificity determinants of human cathepsin s revealed by crystal structures of complexes. *Biochemistry.* 42, 11 (2003).
211. McGrath ME, K. J., Barnes MG, Brömme D. Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nat Struct Biol.* 4, 5 (1997).
212. Somoza JR, P. J., Ho JD. The crystal structure of human cathepsin F and its implications for the development of novel immunomodulators. *J Mol Biol.* 322, 10 (2002).
213. Picó A, M. A., Pericàs MA. Enantiodivergent, catalytic asymmetric synthesis of gamma-amino vinyl sulfones. *J Org Chem.* 68, 9 (2003).
214. Mendieta L, P. A., Tarragó T, Teixidó M, Castillo M, Rafecas L, Moyano A, Giralt E. Novel peptidyl aryl vinyl sulfones as highly potent and selective inhibitors of cathepsins L and B. *ChemMedChem.* 5, 12 (2010).

215. Scheidt KA, R. W., McKerrow JH, Selzer PM, Hansell E, Rosenthal PJ. Structure-based design, synthesis and evaluation of conformationally constrained cysteine protease inhibitors. *Bioorg Med Chem.* 6, 18 (1998).
216. Palmer JT, R. D., Klaus JL, Brömme D. Vinyl sulfones as mechanism-based cysteine protease inhibitors. *J Med Chem.* 38, 4 (1995).
217. Larson ET, P. F., Huynh MH, Giebel JD, Kelley AM, Zhang L, Bogyo M, Merritt EA, Carruthers VB. Toxoplasma gondii cathepsin L is the primary target of the invasion-inhibitory compound morpholinurea-leucyl-homophenyl-vinyl sulfone phenyl. *J Biol Chem.* 284, 12 (2009).
218. Ettari R, N. E., Di Francesco ME, Dude MA, Pradel G, Vicík R, Schirmeister T, Micale N, Grasso S, Zappalà M. Development of peptidomimetics with a vinyl sulfone warhead as irreversible falcipain-2 inhibitors. *J Med Chem.* 51, 9 (2008).
219. Götz MG, C. C., Hansell E, McKerrow JH, Powers JC. Peptidyl allyl sulfones: a new class of inhibitors for clan CA cysteine proteases. *Bioorg Med Chem.* 12, 9 (2004).
220. Musil D, Z. D., Turk D, Engh RA, Mayr I, Huber R, Popovic T, Turk V, Towatari T, Katunuma N, et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* 10, 10 (1991).
221. Tarrago, T., Frutos, S., Rodriguez-Mias, R. A. and Giralt, E. Identification by ¹⁹F NMR of traditional Chinese medicinal plants possessing prolyl oligopeptidase inhibitory activity. *Chembiochem.* 7, 827-833 (2006).
222. Morris, G. M., Goodsell, D. S., Huey, R. and Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des.* 10, 293-304 (1996).