

Romina Paula de Aguiar Guedes

# Virtual Screening e Avaliação Biológica de Novos Inibidores do Proteassoma

Dissertação de Mestrado em Química Farmacêutica Industrial, orientada pelo Professor Doutor Jorge António Ribeiro Salvador e pela Professora Doutora Rita Alexandra Nascimento Cardoso Guedes e apresentada à Faculdade de Farmácia da Universidade de Coimbra

Setembro 2013



UNIVERSIDADE DE COIMBRA

## AGRADECIMENTOS

Aos meus orientadores, Professor Doutor Jorge António Ribeiro Salvador e Professora Doutora Rita Alexandra Nascimento Cardoso Guedes, pelo papel essencial que desempenharam no desenvolvimento da minha tese e por todo o apoio e disponibilidade demonstrados.

À minha família por todo o apoio que me deram.

Aos meus colegas da Faculdade de Farmácia da Universidade de Coimbra e da Faculdade de Farmácia da Universidade de Lisboa por todo o apoio e paciência que tiveram comigo.

Aos meus amigos por estarem presentes sempre que possível e por todo o apoio.

Aos meus colegas e amigos de corrida por partilharem comigo tão bons momentos e que em tanto me ajudaram em dias menos bons.

## RESUMO

A via ubiquitina-proteassoma desempenha um papel importante na homeostase celular, exercendo também um papel relevante na regulação de diversas vias celulares, incluindo no crescimento e na proliferação celular, na apoptose, na reparação do ADN, na transcrição e na resposta imune.

O proteassoma é um complexo multienzimático que contém vários centros ativos, sendo a sua função principal degradar proteínas desnecessárias ou danificadas.

Algumas patologias humanas devem-se a alterações nesta via. Como tal, a inibição da via ubiquitina-proteassoma por inibidores do proteassoma pode ser uma abordagem terapêutica racional para várias doenças, tais como o cancro e doenças inflamatórias.

Com o objetivo de obter novos inibidores do proteassoma para o local ativo CT-L, foram utilizados os seguintes métodos computacionais: modelação por homologia, farmacóforos, *docking* e *virtual screening*.

Foi desenvolvido e validado um modelo de homologia para o local ativo CT-L que teve como base o código PDB 3UN8.

Na modelação farmacofórica, obtivemos e validámos dois modelos farmacofóricos: um modelo com base na estrutura de inibidores conhecidos do local ativo  $\beta 5c$  e um modelo com base na estrutura do recetor.

Na etapa de *virtual screening* foi efetuado o *docking* molecular tendo como base a estrutura de ligandos (no qual se utilizou o modelo farmacofórico obtido com base na estrutura de ligandos) e tendo como base a estrutura do recetor (sem a utilização de farmacóforos como filtro). Os resultados foram analisados de modo a selecionarmos potenciais compostos *hit*.

**Palavras-chave:** via ubiquitina-proteassoma, proteassoma, inibidores do proteassoma, *virtual screening*.

## ABSTRACT

The ubiquitin proteasome pathway plays an important role in cellular homeostasis and also it exerts a critical role in regulating a wide variety of cellular pathways, including cell growth and proliferation, apoptosis, DNA repair, transcription and immune response.

The proteasome is a multienzyme complex, containing several active centers and the main function of the proteasome is to degrade unneeded or damaged proteins.

Defects in this pathway have been implicated in a number of human pathologies. Inhibition of the ubiquitin-proteasome pathway by proteasome inhibitors may be a rational therapeutic approach for various diseases, such as cancer and inflammatory diseases.

To obtain new proteasome inhibitors for the CT-L active site, were used the following computational methods: homology modeling, pharmacophores, docking and virtual screening.

Using the PDB code 3UN8, it was developed and validated an homology model for the CT-L active site.

In pharmacophoric modeling, we obtained and validated two pharmacophoric models: a model based in the structure of known inhibitors of the  $\beta 5c$  active site and a model based in the receptor structure.

In the virtual screening step, the molecular docking was made based in ligand structure (in which was used the pharmacophoric model ligand based) and based in the receptor structure (without the use of the pharmacophore as a filter). The results were analysed so we can select potential hit compounds.

**Key-words:** ubiquitin-proteasome pathway, proteasome, proteasome inhibitors, virtual screening.

# ÍNDICE

Agradecimentos	i
Resumo	ii
Abstract	iii
Lista de abreviaturas	vi
Lista de figuras	viii
Lista de gráficos	xi
Lista de tabelas	xii
<b>I – O proteassoma</b>	<b>1</b>
1.1 – Considerações gerais	2
1.2 – O sistema ubiquitina-proteassoma	2
1.3 – A partícula nuclear 20S	4
1.3.1 – As subunidades $\alpha$	6
1.3.2 – As subunidades $\beta$	7
1.4 – Regulação da partícula nuclear 20S	8
1.5 – Outros tipos de proteassoma	10
1.5.1 – O imunoproteassoma	10
1.5.2 – O proteassoma híbrido	11
1.5.3 – O proteassoma específico do timo (timoproteassoma)	11
1.6 – Topologia e mecanismo das subunidades catalíticas do proteassoma	11
<b>II – Inibidores do proteassoma</b>	<b>15</b>
2.1 – Considerações gerais	16
2.2 – Principais classes de inibidores da partícula nuclear 20S	18
2.3 – Inibidores da partícula reguladora 19S	20
<b>III – O desenvolvimento de fármacos assistido por computador</b>	<b>21</b>
3.1 – O processo de descoberta de novos fármacos e o desenho de fármacos assistido por computador	22
3.2 – Modelação por homologia	26
3.3 – Modelação farmacofórica	30
3.3.1 – Considerações gerais	30
3.3.2 – Definição e representação das características químicas	33
3.3.2.1 – Grupos doadores de ligações de hidrogénio	34

3.3.2.2 – Grupos aceitadores de ligações de hidrogénio	34
3.3.2.3 – Grupos carregados positivamente e negativamente	34
3.3.2.4 – Grupos hidrofóbicos	35
3.4 – <i>Docking</i> molecular	35
3.4.1 – Considerações gerais	35
3.4.2 – Algoritmos de procura	36
3.4.2.1 – Métodos sistemáticos	37
3.4.2.2 – Métodos estocásticos ou aleatórios	37
3.4.2.3 – Métodos de simulação	37
3.4.3 – Funções de <i>scoring</i>	38
3.4.3.1 – Funções com base em campos de força	38
3.4.3.2 – Funções empíricas	39
3.4.3.3 – Funções <i>knowledge-based</i>	39
3.5 – <i>Virtual screening</i>	39
3.5.1 – O <i>software</i> GOLD	42
<b>IV – Resultados e discussão</b>	<b>44</b>
4.1 – Desenvolvimento de modelos de homologia das subunidades catalíticas do proteassoma	45
4.2 – Desenvolvimento de modelos farmacofóricos tridimensionais	52
4.2.1 – Desenvolvimento de modelos farmacofóricos tridimensionais com base na estrutura de inibidores da subunidade catalítica $\beta 5c$ do proteassoma	56
4.2.2 – Desenvolvimento de modelos farmacofóricos tridimensionais com base na estrutura tridimensional da subunidade catalítica $\beta 5c$ do proteassoma	69
4.3 – <i>Virtual screening</i>	77
4.3.1 – <i>Virtual screening</i> com base na estrutura de ligandos	78
4.3.2 – <i>Virtual screening</i> com base na estrutura do recetor	80
4.3.3 – Discussão de resultados do <i>virtual screening</i>	81
<b>V – Considerações finais e perspetivas futuras</b>	<b>92</b>
<b>VI – Anexos</b>	<b>95</b>
<b>VII – Lista de referências</b>	<b>121</b>

## LISTA DE ABREVIATURAS

Å – Angstrom

Acc – Aceitador de ligações de hidrogénio

Acc2 – Aceitador de ligações de hidrogénio projetado

Aro – Aromático

ASP – *Astex Statistical Potential*

AVC – Acidente vascular cerebral

ADME – Absorção, distribuição, metabolização, excreção

BLAST – *Basic Local Alignment Search Tool*

CADD – *Computer aided drug design* (desenho de fármacos assistido por computador)

C-L – *Caspase-like*

CT-L – *Quimotripsina-like*

DM – Dinâmica molecular

Don – Doador de ligações de hidrogénio

Don2 – Doador de ligações de hidrogénio projetado

DUB – *Deubiquitinating enzymes*

FDA – *Food and Drug Administration*

FEP – *Free energy perturbation* (perturbação de energia livre)

GA – *Genetic algorithm* (algoritmo genético)

GOLD – *Genetic optimisation for ligand docking*

HTS – *High throughput screening* (*screening* de elevado rendimento)

Hyd – Hidrofóbico

IC<sub>50</sub> – concentração inibitória a 50%

INF-γ – Interferão gama

IUPAC – *International Union of Pure and Applied Chemistry*

K<sub>i</sub> – Constante de inibição

LE – *Ligand efficiency* (eficiência do ligando)

LMP – *Low-molecular-weight protein*

MC – Monte Carlo

MECL – *Multicatalytic endopeptidase-like-complex*

MF – Modelo(s) farmacofórico(s)

MHC-I – *Major histocompatibility complex class I* (complexo de histocompatibilidade major classe I)

ML – Ligando metálico  
ML2 – Ligando metálico projetado  
MM – Mecânica molecular  
MMFF – *Merck molecular force field*  
MOE – *Molecular Operating Environment*  
MQ – Mecânica quântica  
NCI – *National Cancer Institute*  
PA – *Post acidic*  
PiR – Anel pi  
QSAR – *Quantitative structure activity relationship* (relação estrutura-atividade quantitativa)  
RMN – Ressonância magnética nuclear  
RMSD – Desvio quadrático médio  
S – Svedberg  
SI – Sistema internacional  
T-L – *Tripsina-like*  
TNF  $\alpha$  – Fator de necrose tumoral alfa  
UBD – *Ubiquitin binding domains*  
UniProt – *Universal Protein Resource*  
VS – *Virtual screening*  
3D-QSAR – *Three-dimensional quantitative structure-activity relationship* (relação estrutura-atividade quantitativa tridimensional)

**Abreviaturas de aminoácidos:**

Ala/A – Alanina	Phe/F – Fenilalanina
Arg/R – Arginina	Ser/S – Serina
Asn/N – Asparagina	Thr/T – Treonina
Asp/D – Aspartato	Tyr/Y – Tirosina
Cys/C – Cisteína	Val/V – Valina
Gln/Q – Glutamina	
Glu/E – Glutamato	
Gly/G – Glicina	
His/H – Histidina	
Lys/K – Lisina	
Leu/L – Leucina	

## LISTA DE FIGURAS

<b>Figura 1</b> – O sistema ubiquitina-proteassoma.	4
<b>Figura 2</b> – Estrutura dos proteassomas 20S de <i>archae</i> e de eucariotas	5
<b>Figura 3</b> – Os centros catalíticos do proteassoma.	7
<b>Figura 4</b> – Estrutura do proteassoma 26S eucariótico.	9
<b>Figura 5</b> – Os <i>pockets</i> de especificidade do substrato e a respectiva nomenclatura dos resíduos do substrato.	12
<b>Figura 6</b> – Mecanismo catalítico das subunidades $\beta$ com intervenção de moléculas de água.	13
<b>Figura 7</b> – Mecanismo de ação das subunidades $\beta$ sem intervenção de moléculas de água.	14
<b>Figura 8</b> – Alguns exemplos de inibidores da partícula nuclear 20S.	20
<b>Figura 9</b> – Ubistatina A, um inibidor da partícula reguladora 19S.	20
<b>Figura 10</b> – O processo de descoberta e desenvolvimento de novos fármacos.	23
<b>Figura 11</b> – Sobreposição do modelo de homologia humano da subunidade CT-L do proteassoma com as subunidades de levedura iniciais utilizadas.	49
<b>Figura 12</b> – <i>Docking</i> do bortezomib na Thr1 no modelo de homologia na presença de uma molécula de água e na ausência de moléculas de água.	51
<b>Figura 13</b> – <i>Docking</i> do carfilzomib na Thr1 no modelo de homologia na presença de uma molécula de água e na ausência de moléculas de água.	51
<b>Figura 14</b> – O MF desenvolvido por Lei <i>et al.</i> alinhado com o bortezomib (MG341) e as três moléculas que apresentaram $IC_{50}$ inferior ao bortezomib	53
<b>Figura 15</b> – Alguns exemplos de inibidores da subunidade $\beta 5c$ utilizados para a construção da base de dados de “controlo positivo”.	54
<b>Figura 16</b> – Alguns exemplos dos compostos utilizados para a construção da base de dados de “controlo negativo”.	55
<b>Figura 17</b> – O modelo farmacofórico tridimensional efectuado com base na estrutura de inibidores da subunidade $\beta 5c$ do proteassoma (epoxicetonas) obtido após optimização e validação, MFL4.	66
<b>Figura 18</b> – Os raios (em Å) das características farmacofóricas presentes no MFL4.	67
<b>Figura 19</b> – As distâncias (em Å) entre as características farmacofóricas presentes no MFL4.	67
<b>Figura 20</b> – Os ângulos (em °) entre as características farmacofóricas presentes no	67

MFL4.	
<b>Figura 21</b> – Local activo CT-L do proteassoma identificado pela funcionalidade <i>Site Finder</i> .	70
<b>Figura 22</b> – Superfície molecular do recetor.	71
<b>Figura 23</b> – Mapa de preferências de contacto.	71
<b>Figura 24</b> – Refinamento dos “ <i>dummy atoms</i> ” resultantes dos centros alfa.	72
<b>Figura 25</b> – O primeiro MF obtido com base na estrutura tridimensional do recetor antes da refinamento do mesmo.	73
<b>Figura 26</b> – O MF final obtido com base na estrutura do local activo CT-L do proteassoma. Características farmacofóricas e raio de cada característica.	76
<b>Figura 27</b> – O MF final obtido com base na estrutura do local activo CT-L do proteassoma. Distâncias (em Å) entre as características farmacofóricas.	76
<b>Figura 28</b> – O MF final obtido com base na estrutura do local activo CT-L do proteassoma. Principais ângulos entre as características farmacofóricas.	76
<b>Figura 29</b> – O inibidor 2b obtido por Basse <i>et al.</i> através de VS.	77
<b>Figura 30</b> – VS com base na estrutura de ligandos.	79
<b>Figura 31</b> – VS com base na estrutura do recetor.	81
<b>Figura 32</b> – Algumas das moléculas do VS com base na estrutura do recetor.	82
<b>Figura 33</b> – Algumas das moléculas do VS com base na estrutura do recetor.	84
<b>Figura 34</b> – O composto com o código NSC722655 (cor-de-rosa) posicionado no modelo de homologia e sobreposto com o PR-957 (azul). A ThrI apresenta a cor verde.	89
<b>Figura 35</b> – Interações do composto com o código NSC722655 <i>dockado</i> no modelo de homologia	89
<b>Figura 36</b> – <i>Esquerda</i> : O composto com o código NSC113926 (cor-de-rosa) posicionado no modelo de homologia e sobreposto com o PR-957 (azul).	89
<b>Figura 37</b> – O composto com o código NSC617200 (cor-de-rosa) posicionado no modelo de homologia, encontrando-se o ligando cristalográfico PR-957 a azul e a ThrI a verde.	90
<b>Figura 38</b> – Interações do composto com o código NSC617200 <i>dockado</i> no modelo de homologia.	90
<b>Figura 39</b> – <i>Esquerda</i> : O composto com o código NSC83442 (cor-de-rosa) posicionado no modelo de homologia. A azul encontra-se o ligando cristalográfico PR-957 e a ThrI apresenta a cor verde. <i>Direita</i> : Não se verificaram interações entre	91

o composto com o código NSC83442 e a ThrI.

**Figura 40 (anexo 3)** – Alguns inibidores conhecidos *dockados* da subunidade catalítica  $\beta 5c$  na escala do nM e respetivos *scores* obtidos. 100

**Figura 41 (anexo 4)** – Sobreposição dos modelos de homologia dos *pockets* das subunidades catalíticas do proteassoma e do imunoproteassoma com as subunidades de levedura iniciais utilizadas. 101

**Figura 42 (anexo 5)** – Base de dados de controlo positivo utilizada na etapa de modelação farmacofórica. 103

**Figura 43 (anexo 5)** – Base de dados de controlo negativo utilizada na etapa de modelação farmacofórica. 110

## LISTA DE GRÁFICOS

<b>Gráfico 1</b> – Gráfico de Ramachandran do modelo de homologia humano da subunidade CT-L do proteassoma.	49
<b>Gráfico 2</b> – Número de moléculas detetadas nos controlos positivo e negativo durante o refinamento do MFL3.	66
<b>Gráfico 3</b> – Número de moléculas detetadas nos controlos positivo e negativo durante o refinamento do MF com base na estrutura do recetor.	75

## LISTA DE TABELAS

<b>Tabela 1</b> – As subunidades $\alpha$ e as suas diferentes designações.	6
<b>Tabela 2</b> – As subunidades $\beta$ e as suas diferentes designações.	8
<b>Tabela 3</b> – Principais classes de inibidores da partícula nuclear 20S.	19
<b>Tabela 4</b> – Os compostos de cada uma das classes de inibidores utilizadas para obter os modelos farmacofóricos iniciais	57
<b>Tabela 5</b> – Os MF iniciais obtidos para cada uma das cinco classes utilizadas de inibidores do proteassoma.	59
<b>Tabela 6</b> – Modelos farmacofóricos obtidos para as epoxicetonas antes de ser efetuado o seu refinamento.	61
<b>Tabela 7</b> – Modelos farmacofóricos obtidos para as vinilsulfonas antes de ser efetuado o seu refinamento.	62
<b>Tabela 8</b> – Refinamento final do MFL3.	65
<b>Tabela 9</b> – Refinamento final do modelo farmacofórico com base na estrutura do recetor.	74
<b>Tabela 10 (anexo 1)</b> – Exemplos de estruturas cristalográficas do proteassoma.	96
<b>Tabela 11 (anexo 2)</b> – Percentagem de identidade das subunidades do proteassoma/imunoproteassoma entre diversos organismos.	98
<b>Tabela 12 (anexo 4)</b> – Resumo dos códigos utilizados para efetuar cada modelo de homologia.	101
<b>Tabela 13 (anexo 6)</b> – Refinamento das características farmacofóricas dos melhores modelos farmacofóricos iniciais obtidos com base na estrutura de epoxicetonas e vinilsulfonas e respetivos resultados na deteção de compostos das bases de dados do controlo positivo e negativo.	113
<b>Tabela 14 (anexo 7)</b> – Refinamento das características farmacofóricas do primeiro modelo farmacofórico obtido com base na estrutura tridimensional do recetor.	115
<b>Tabela 15 (anexo 8)</b> – Compostos selecionados no VS com base na estrutura de ligandos.	117
<b>Tabela 16 (anexo 9)</b> – Compostos selecionados no VS com base na estrutura do recetor.	119

I – O PROTEASSOMA

## **I.1 – CONSIDERAÇÕES GERAIS**

A relevância da degradação proteolítica nas células e o papel da ubiquitina em vias proteolíticas foi reconhecido em 2004 através da atribuição do Prémio Nobel da Química a Aaron Ciechanover, Avram Hershko e Irwin Rose.<sup>1,2</sup>

O sistema proteolítico intracelular citosólico mais relevante é o proteassoma, encontrando-se este presente em diversos organismos, desde *archae* e bactérias, bem como em plantas, leveduras e animais.<sup>1,3,4</sup>

Nas células eucarióticas, o proteassoma localiza-se no citoplasma e no núcleo, podendo ainda estar ligado ao retículo endoplasmático e associado ao citoesqueleto.<sup>3,5,6,7</sup>

Geralmente os subcomponentes do sistema proteassómico são referidos tendo como base o coeficiente de sedimentação de Svedberg (S).<sup>2</sup> O svedberg é uma unidade que não pertence ao Sistema Internacional (SI) e que está relacionada como a velocidade de sedimentação (particularmente ao nível da centrifugação). Isto reflete a velocidade em que uma molécula sedimenta sob a força centrífuga, proporcionando uma medida do tamanho da partícula com base na sua velocidade de sedimentação.<sup>8</sup>

Nas células eucarióticas, o proteassoma 26S tem um papel fundamental na degradação proteica dependente do ATP.<sup>9,10,11</sup> O proteassoma 26S tem cerca de 2500 kilodaltons (kDa) em termos de massa molecular e contém uma partícula nuclear 20S e duas partículas reguladoras 19S.<sup>6,12</sup>

No entanto, para além da forma livre 20S e da forma 26S, o proteassoma pode existir na forma de imunoproteassoma, proteassoma híbrido e timoproteassoma.<sup>13</sup>

O imunoproteassoma é particularmente importante na medida em que a expressão deste se encontra aumentada devido aos níveis elevados de citocinas pró-inflamatórias em diversas patologias, nomeadamente em doenças autoimunes e neurodegenerativas.<sup>14</sup>

## **I.2 – O SISTEMA UBIQUITINA-PROTEASSOMA**

A existência de um *turnover* altamente regulado das proteínas celulares contribui para a homeostase celular. Assim, as proteínas que estão desnaturadas, danificadas (devido, por exemplo, ao stress oxidativo) ou que já não são necessárias devem ser reconhecidas e removidas através de degradação proteolítica catalisada por proteases que clivam ligações peptídicas.<sup>1,3,15</sup>

Nas células eucarióticas, existem duas vias principais de degradação proteica intracelular: a via lisossomal e a via ubiquitina-proteassoma. As proteínas exógenas (que habitualmente entram para a célula por endocitose) e as proteínas membranares são geralmente degradadas através de lisossomas, enquanto que a maior parte das proteínas citoplasmáticas são degradadas pela via ubiquitina-proteassoma.<sup>12,16</sup>

O sistema ubiquitina-proteassoma (*Ubiquitin Proteasome System – UPS*) é uma via relevante para a degradação controlada de diversas proteínas, não só a nível do ciclo celular e apoptose, mas também a nível da resposta inflamatória, da apresentação de antígenos, do stress oxidativo e da carcinogénese. No entanto, para além da função proteolítica referida anteriormente, o proteassoma também pode estar envolvido em diversas atividades celulares não proteolíticas tais como reparação do ADN e remodelação da cromatina.<sup>10,17,18</sup>

Existe uma enorme diversidade de substratos do proteassoma, tais como moléculas sinalizadoras, supressores de tumores, reguladores do ciclo celular, fatores de transcrição (tais como c-Jun e E2F1), moléculas inibidoras, proteínas anti-apoptose (das quais é exemplo o Bcl-2), entre outros. Quando se verifica uma interrupção ao nível da degradação destes substratos proteicos as consequências que daí advêm são enormes, nomeadamente ao nível de células que se dividem rapidamente, como é o caso das células neoplásicas que requerem uma enorme disponibilidade de proteínas promotoras do crescimento de modo a que a mitose acelerada e não controlada (característica deste tipo de células) seja possível.<sup>15,19</sup>

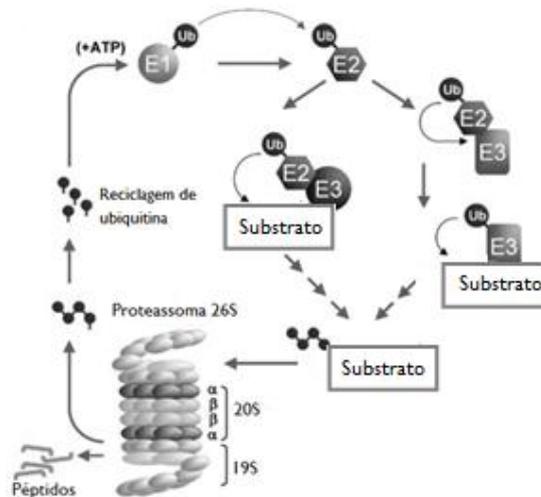
Deficiências ao nível da homeostase proteica podem conduzir a patologias cardiovasculares, neurodegenerativas, oncológicas e metabólicas. Como tal, a atividade desta protease pode ter como controlo-chave a regulação da entrada do substrato na superfície interna do proteassoma.<sup>20</sup>

Regra geral, para que uma proteína seja reconhecida pelo proteassoma, esta é marcada com ubiquitina, sendo a ubiquitina um pequeno péptido composto por 76 resíduos de aminoácidos de 8500 Da de peso molecular que tem diversos resíduos de lisina (Lys) que possibilitam a formação de ligações isopeptídicas de modo a que se forme uma cadeia poliubiquitinada.<sup>19,21,22</sup> No entanto, alguns substratos são degradados pelo proteassoma sem haver necessidade de serem previamente marcados com moléculas de ubiquitina.<sup>12,23</sup> Tal acontece no caso de partículas reguladoras com tempo de semi-vida curto, não se encontrando estas necessariamente danificadas ou desnaturadas e, como tal, não têm de ser marcadas com um fragmento hidrofóbico externo sob a forma de uma cadeia poliubiquitinada. É então necessário que o ATP seja hidrolizado para que ocorra o desdobramento desses substratos de forma a que consigam entrar no núcleo proteolítico do

proteassoma. Para além do que foi mencionado, as proteínas danificadas oxidativamente apresentam resíduos hidrofóbicos expostos, facto este que permite que não seja necessário conjugar com ubiquitina as proteínas oxidadas de modo a torná-las mais hidrofóbicas, não sendo também necessário recorrer à energia proveniente da hidrólise do ATP visto que a proteína já se encontra desdobrada.<sup>13</sup>

O processo de ubiquitinação ocorre mediante a ação de uma cascata enzimática que, de um modo genérico, ativa a ubiquitina que se encontra livre e transporta-a até à proteína alvo. As enzimas intervenientes denominam-se E1 (enzima ativante da ubiquitina), E2 (enzima de conjugação da ubiquitina) e E3 (ligase ubiquitina-proteína – é a enzima que determina a especificidade do substrato).<sup>1,4,19</sup>

Todo o processo tem início com a ativação da enzima E1 (esta enzima utiliza ATP para formar uma ligação tioéster altamente energética com a ubiquitina) que vai então transferir a ubiquitina para a proteína transportadora E2. E2 vai transferir a ubiquitina a E3 que, por sua vez, se vai ligar à proteína alvo e interagir com E2 de modo a ligar covalentemente a ubiquitina. Este processo é efetuado diversas vezes, o que origina uma cadeia poliubiquitinada que sinaliza a proteína para destruição pelo proteassoma (figura 1).<sup>21,24</sup>



**Figura 1** – O sistema ubiquitina-proteassoma. Adaptado de Zolko *et al.*<sup>25</sup>

### 1.3 – A PARTÍCULA NUCLEAR 20S

A partícula nuclear 20S (muitas vezes designada por “proteassoma 20S” ou, por vezes, apenas por “proteassoma”) representa a fração catalítica do sistema proteassómico e tem cerca de 700 kDa.<sup>26</sup>

Atualmente, conhecem-se duas formas básicas do proteassoma 20S: uma forma mais ancestral que está presente em *archaea* (sendo o *Thermoplasma acidophilum* um exemplo) e uma outra forma mais evoluída que se encontra em leveduras, plantas e animais.<sup>3,15</sup>

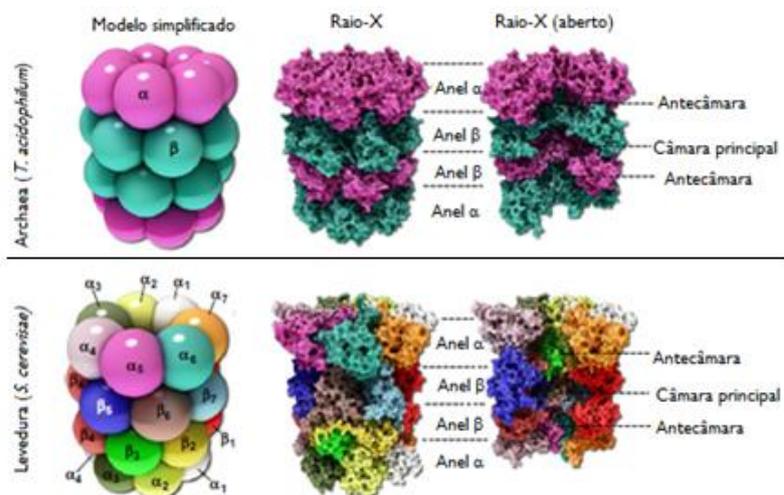
A estrutura tridimensional do proteassoma de diversos organismos tem sido investigada através de cristalografia de raio-X, tendo-se verificado que têm a mesma arquitetura básica (figura 2).<sup>4</sup>

O proteassoma na sua forma mais ancestral contém quatro anéis heptaméricos [dois anéis alfa ( $\alpha$ ) e dois anéis beta ( $\beta$ ) organizados na forma  $\alpha$ - $\beta$ - $\beta$ - $\alpha$ ] dispostos cilíndricamente e que são constituídos por sete subunidades iguais.<sup>4,7,22</sup>

Estruturalmente, o proteassoma dos mamíferos apresenta também uma estrutura cilíndrica, tendo cerca de 100x160 Å.<sup>3</sup> Esta estrutura é constituída por quatro anéis (dois anéis  $\alpha$  e dois anéis  $\beta$  organizados na forma  $\alpha$ - $\beta$ - $\beta$ - $\alpha$ ), sendo cada anel constituído por sete subunidades diferentes.<sup>15,18</sup> Assim, a forma mais evoluída do proteassoma 20S compreende catorze subunidades diferentes ( $\alpha$ 1 –  $\alpha$ 7 e  $\beta$ 1 –  $\beta$ 7) com massas moleculares compreendidas entre 20 e 30 kDa.<sup>8,19,27,28</sup>

Há ainda a considerar a subdivisão do interior do proteassoma em duas antecâmaras (entre os anéis  $\alpha$  e  $\beta$ ) e uma “câmara principal” que se situa entre os dois anéis  $\beta$ .

O centro catalítico do proteassoma encontra-se ao nível dos anéis mais internos (os anéis  $\beta$ ). Por sua vez, os anéis  $\alpha$  conseguem alterar quer a atividade quer a especificidade do proteassoma visto que são responsáveis pelo reconhecimento e ligação dos substratos, assim como pela regulação da entrada dos substratos na câmara proteolítica.<sup>9,11</sup>



**Figura 2** – Estrutura dos proteassomas 20S de *archaea* e de eucariotas (aqui representados por uma levedura). Adaptado de Jung *et al* <sup>3</sup>.

Em termos da atividade deste sistema proteolítico altamente organizado, existem diversos reguladores que se encontram ligados ao proteassoma 20S de modo a modular a sua atividade.<sup>3</sup> Estas partículas reguladoras têm em comum o facto de se ligarem às subunidades  $\alpha$  dos anéis externos.<sup>4</sup>

### 1.3.1 – AS SUBUNIDADES $\alpha$

Os proteassomas dos *archae* possuem seis subunidades  $\alpha$  idênticas. O anéis  $\alpha$  das leveduras e dos mamíferos são compostos por sete subunidades diferentes ( $\alpha 1$ –  $\alpha 7$ ), podendo estas subunidades surgir com diferentes designações (tabela 1).<sup>19</sup>

Nome sistemático da subunidade	Outras designações para as subunidades nas leveduras (exemplo: <i>Saccharomyces cerevisiae</i> )	Outras designações para as subunidades nos mamíferos (exemplo: <i>Homo sapiens</i> )
$\alpha 1$	C7/Prs2	HsPROS27/Hslota
$\alpha 2$	Y7	HsC3
$\alpha 3$	Y13	HsC9
$\alpha 4$	Pre6	HsC6/XAPC7
$\alpha 5$	Pup2	HsZeta
$\alpha 6$	Pre5	HsC2/HsPROS30
$\alpha 7$	C1/ Prs1	Prs1 HsC8

**Tabela 1** – As subunidades  $\alpha$  e as suas diferentes designações. Adaptado de Jung et al.<sup>19</sup>

Geralmente, o proteassoma 20S encontra-se sob a forma inativa. No entanto, pode ser ativado, por exemplo, por diversas partículas reguladoras e proteínas desnaturadas.<sup>3</sup>

Os anéis  $\alpha$  participam essencialmente nas funções de reconhecimento do substrato e na regulação do seu acesso à câmara proteolítica interna. A entrada de substratos no proteassoma pode ser modulada através da ligação aos anéis  $\alpha$  de uma partícula reguladora (como, por exemplo, I9S, I1S ou PA200).<sup>3,4</sup>

A ligação dos anéis  $\alpha$  a aminoácidos hidrofóbicos (que, em condições normais, se encontram no interior de proteínas globulares e solúveis em água; no entanto, em proteínas danificadas, ocorre exposição de aminoácidos hidrofóbicos) parece mediar a ativação do proteassoma.<sup>3</sup>

Assim, após contactar com um determinado substrato, ocorre uma alteração conformacional ao nível dos anéis  $\alpha$ , abrindo a porta por eles criada, o que vai permitir a entrada do substrato na câmara proteolítica interna. Esta porta é constituída pelos N-terminais das três subunidades  $\alpha 2$ ,  $\alpha 3$  e  $\alpha 4$ .<sup>9,19</sup>

Outro aspeto a ter em consideração é a distinção entre a atividade de peptidase e a atividade de protease do proteassoma. Relativamente à atividade de peptidase, esta está relacionada com a degradação de pequenos fragmentos de aminoácidos, sendo geralmente independente do *status* da porta; por sua vez, a atividade de protease diz respeito à degradação de proteínas desdobradas, sendo dependente do *status* da porta. Pelo que foi mencionado, tudo leva a crer que os anéis  $\alpha$  exercem pequenas interações com pequenos péptidos, mas apresentam um papel importante na degradação de grandes péptidos ou proteínas.<sup>3,19</sup>

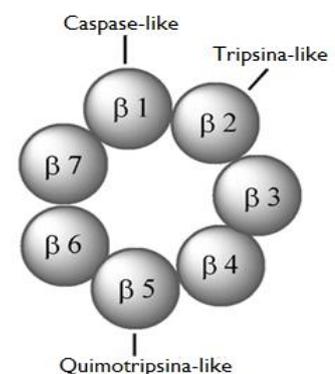
### 1.3.2 – AS SUBUNIDADES $\beta$

A principal função das subunidades  $\beta$  do proteassoma consiste na atividade proteolítica propriamente dita, atividade esta que origina péptidos de 3 a 30 aminoácidos, predominando os fragmentos de 6 a 8 aminoácidos. Muitos destes péptidos resultantes são posteriormente degradados em aminoácidos individuais através da intervenção de aminopeptidasas.<sup>16,19,29</sup>

Tal como já foi referido, o proteassoma dos *archaea* é composto por sete subunidades  $\beta$  idênticas num anel e, como tal, tem sete centros proteolíticos por anel. Por sua vez, os proteassomas de leveduras mais desenvolvidas e de mamíferos apenas têm três centros proteolíticos distintos por anel  $\beta$  (nas subunidades  $\beta 1$ ,  $\beta 2$  e  $\beta 5$ ). Tal deve-se ao facto de as subunidades  $\beta$  estarem geneticamente codificadas como precursores, sendo processadas durante a maturação da partícula nuclear por autólise, ocorrendo assim a clivagem de uma ligação amida, libertando o N terminal da treonina 1 (Thr1).<sup>30,31</sup>

Assim sendo, a atividade enzimática das subunidades  $\beta$  está associada com os resíduos N-terminal da Thr1, nos quais o grupo hidroxilo  $\gamma$  (Thr1O $\gamma$ ) atua como nucleófilo na reação de hidrólise da ligação peptídica, o que permite classificar o proteassoma como membro do grupo da superfamília das Ntn hidrolases (“N-terminal nucleophilic”).<sup>32,33</sup>

No caso dos mamíferos e leveduras mais evoluídas, cada uma das três subunidades proteolíticas apresenta aspetos distintos na ligação a substratos e atividade exercida: a subunidade  $\beta 1$  apresenta atividade “caspase-like” (C-L) ou “post acidic” (PA) e cliva após aminoácidos ácidos; a



**Figura 3** – Os centros catalíticos do proteassoma. Adaptado de Rentsch et al.<sup>40</sup>

subunidade  $\beta 2$  exerce atividade “tripsina-like” (T-L), sendo a clivagem efetuada após aminoácidos básicos; a subunidade  $\beta 5$  tem atividade “quimotripsina-like” (CT-L), atua após aminoácidos neutros (figura 3).<sup>11,27</sup>

Como é possível constatar através da tabela 2, tal como nas subunidades  $\alpha$  também as subunidades  $\beta$  podem ter diferentes designações.

Nome sistemático da subunidade	Outras designações para as subunidades nas leveduras (exemplo: <i>Saccharomyces cerevisiae</i> )	Outras designações para as subunidades nos mamíferos (exemplo: <i>Homo sapiens</i> )
$\beta 1$	Pre3	HsDelta/Y
$\beta 2$	Pup1	Z
$\beta 3$	Pup3	HsC10-II
$\beta 4$	C11/Pre1	HsC7-I
$\beta 5$	Pre2	X/MB1
$\beta 6$	C5/Prs3	HsC5
$\beta 7$	Pre4	HsN3/HsBPROS26

**Tabela 2** – As subunidades  $\beta$  e as suas diferentes designações. Adaptado Jung et al.<sup>19</sup>

#### 1.4 – REGULAÇÃO DA PARTÍCULA NUCLEAR 20S

O reconhecimento e a degradação de substratos têm de ser devidamente regulados de modo a que a degradação proteolítica ocorra de forma controlada.<sup>3</sup>

Os anéis  $\alpha$  regulam a entrada de substratos na câmara proteolítica do proteassoma 20S, podendo esta entrada ser ativada através de componentes hidrofóbicos que se encontram expostos à superfície em proteínas oxidativamente danificadas. No entanto, existem diferentes reguladores que podem mudar a atividade e a especificidade da partícula nuclear 20S.<sup>3,15</sup> As proteínas são marcadas com ubiquitina de modo a serem alvo de degradação proteossômica, sendo que o regulador mais importante que participa no reconhecimento de proteínas ubiquitinadas é o chamado “complexo regulador 19S”.<sup>3,18</sup>

O complexo regulador 19S (também denominado por “PA700”, “ativador do proteassoma 700kDa” e, no caso dos *archae*, designa-se PAN) tem como função regular a entrada de substratos no proteassoma 20S, participar no reconhecimento dos substratos proteicos ubiquitinados, na libertação das moléculas de ubiquitina que se encontram ligadas aos substratos e no desdobraimento das proteínas.<sup>4,7,22</sup>

Se ocorrer a ligação do regulador 19S em cada um dos anéis  $\alpha$  das extremidades do proteassoma 20S, ocorre a formação do chamado “proteassoma 26S”, uma partícula com cerca de 2MDa (figura 4).<sup>3</sup>

De uma forma geral, o regulador 19S é composto por duas estruturas: uma base em forma de anel (através da qual se liga aos anéis  $\alpha$  do proteassoma 20S) e ainda uma tampa que reconhece e liga proteínas poliubiquitinadas. Como a ligação entre a base e a tampa é relativamente fraca, pode ocorrer que, sob determinadas condições bioquímicas, se formem proteassomas sem tampa devido à separação desta relativamente à base.<sup>3,20</sup>

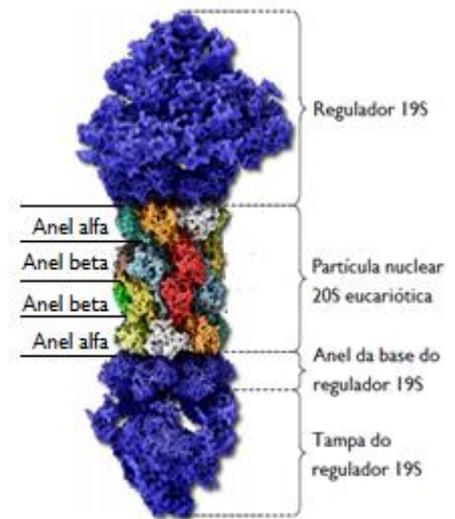
As cadeias de ubiquitina são reconhecidas por *Ubiquitin-binding domain* (UBD), uma estrutura que se encontra em várias proteínas celulares.<sup>4,18</sup>

Antes de uma proteína ser completamente degradada e desdobrada, a cadeia de moléculas de ubiquitina tem de ser removida através de um processo altamente cooperativo e regulado, começando pela clivagem da cadeia em moléculas individuais de ubiquitina de modo a que estas possam ser recicladas (sendo esta clivagem catalisada por enzimas desubiquitinantes, *Deubiquitinating enzymes* – DUB's, que se localizam na partícula reguladora 19S).<sup>3</sup>

Existem ainda outros reguladores, nomeadamente a partícula reguladora IIS, o regulador nuclear PA200 e os reguladores PR39 e PI31.

A partícula reguladora IIS (também conhecida por PA28 ou REG) é constituída por três subunidades diferentes (PA28 $\alpha$ , PA28 $\beta$  e PA28 $\gamma$ ).<sup>3,19</sup> Regra geral, os reguladores IIS ligam-se aos anéis  $\alpha$  e alteram as propriedades de degradação do substrato (aumentam a velocidade máxima e reduzem a constante de Michaelis-Menten), sendo esta degradação de substratos pelo complexo IIS-20S independente do ATP, o que sugere que apenas são utilizadas proteínas desnaturadas como substratos.<sup>3,19</sup>

O ativador do proteassoma conhecido por PA200 encontra-se apenas no núcleo de células de mamíferos e altera a sua estrutura de modo a reforçar o acesso do substrato à partícula nuclear, intervindo ao nível da espermatogénese e da reparação do ADN. Este regulador apresenta três isoformas diferentes: PA200i, PA200ii e PA200iii (no entanto, apenas a isoforma PA200i interage com o proteassoma). Trata-se de uma estrutura monomérica que está em contacto com praticamente todas as subunidades  $\alpha$ , com exceção da subunidade  $\alpha 7$ ; em contrapartida, a forma Blm10 (que se encontra presente em leveduras) liga-se à totalidade das subunidades  $\alpha$ .<sup>3</sup>



**Figura 4** – Estrutura do proteassoma 26S eucariótico. Adaptado de Jung e Grune.<sup>3</sup>

Os reguladores PR39 e PI3I atuam como inibidores celulares do proteassoma. O PR39 é composto por 39 aminoácidos e trata-se de um inibidor não-competitivo que está presente em leveduras e em mamíferos. O seu mecanismo de inibição consiste na sua ligação à subunidade  $\alpha 7$  do proteassoma, o que origina uma alteração alostérica na estrutura do proteassoma que conduz a uma diminuição da atividade proteolítica e interfere na ligação ao regulador 19S. Por sua vez, a proteína PI3I presente em mamíferos compete com as formas  $\alpha$  e  $\beta$  do PA28 para a ligação ao proteassoma.<sup>3</sup>

## **1.5 – OUTROS TIPOS DE PROTEASSOMA**

### **1.5.1 – O IMUNOPROTEASSOMA**

Nos proteassomas de mamíferos, pode ser induzida uma forma especial do proteassoma constitutivo 20S denominada imunoproteassoma (i20S), encontrando-se no local das subunidades constitutivas ( $\beta 1$ ,  $\beta 2$  e  $\beta 5$ ) os seus equivalentes indutivos ( $\beta 1i$ ,  $\beta 2i$  e  $\beta 5i$ ).<sup>17</sup>

Podemos encontrar na literatura outras designações para estas subunidades indutivas. Assim, a subunidade  $\beta 1i$  pode surgir mencionada como  $\beta 9$  ou “*low-molecular-weight protein 2*” (LMP2),  $\beta 2i$  pode ser designada como  $\beta 10$  ou “*multicatalytic endopeptidase-like-complex-1*” (MECL1) e, por fim,  $\beta 5i$  pode representar-se por  $\beta 8$  ou LMP7.<sup>3,27</sup>

O imunoproteassoma, o qual é expresso em células de origem hematopoiética, é uma forma alternativa do proteassoma constitutivo e pode também ser induzido em células não hematopoiéticas após exposição a citocinas anti-inflamatórias, tais como o interferão  $\gamma$  (INF- $\gamma$ ) e o fator de necrose tumoral  $\alpha$ .<sup>34</sup>

Este processo não se trata de uma substituição, mas sim de uma síntese *de novo*.<sup>19</sup>

Os produtos do imunoproteassoma são geralmente pequenos fragmentos proteicos com cerca de oito a dez aminoácidos.<sup>4</sup>

O imunoproteassoma exerce um papel importante devido à produção de péptidos antigénicos que podem ser apresentados pelo complexo de histocompatibilidade major classe I (MHC-I) à superfície da célula participando, desta forma, na iniciação da resposta imune.<sup>17</sup>

### **1.5.2 – O PROTEASSOMA HÍBRIDO**

O proteassoma híbrido contém ambas as estruturas reguladoras 19S (PA700) e 11S (PA28), estando cada um destes reguladores ligado a cada uma das extremidades do proteassoma 20S.<sup>3,19</sup>

Relativamente às funções desempenhadas pelos proteassomas híbridos, estas ainda não se encontram totalmente esclarecidas, apenas existem algumas suposições. Assim, crê-se que as especificidades proteolíticas no núcleo proteassómico ligado a 19S sejam alteradas aquando da ligação de um regulador PA28 adicional, originando diferentes produtos oligopeptídicos. No entanto, a atividade proteolítica da forma 26S é superior à das estruturas híbridas.<sup>3</sup>

Presume-se que haja cooperação entre o imunoproteassoma e as formas híbridas do proteassoma no processamento de antígenos devido ao facto de as formas referidas poderem ser induzidas pelo interferão  $\gamma$ . Tal sugere que a maior parte da proteólise necessária para apresentação de antígenos ao MHC-I ocorre via proteassoma 26S ou através de proteassomas híbridos.<sup>3</sup>

### **1.5.3 – O PROTEASSOMA ESPECÍFICO DO TIMO (TIMOPROTEASSOMA)**

A designação “timoproteassoma” foi sugerida para a configuração  $\beta 1i$ ,  $\beta 2i$  e  $\beta 5t$  do proteassoma 20S devido à existência da subunidade  $\beta 5t$ .<sup>19</sup>

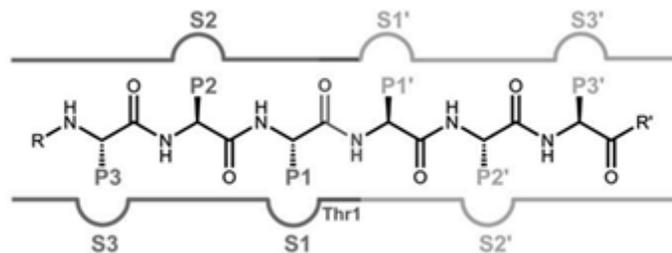
Este proteassoma encontra-se nas células epiteliais do timo cortical e é responsável pela formação de pequenos oligopéptidos antigénicos na superfície celular, seleccionando positivamente células T CD8<sup>+</sup>.<sup>19</sup>

### **1.6 – TOPOLOGIA E MECANISMO DAS SUBUNIDADES CATALÍTICAS DO PROTEASSOMA**

Segundo a nomenclatura de Schechter e Berger<sup>35</sup>, as interações de uma protease com o substrato podem ser definidas denominando por “P1, P2, P3, ..., Pn” os resíduos de aminoácidos do substrato a partir da ligação peptídica hidrolisada até ao grupo amina terminal (são os chamados “*nonprimed P sites*”), sendo os resíduos de aminoácidos que se encontram na direção do grupo carboxilo terminal designados por “P1’, P2’, P3’, ..., Pn’ ” (“*primed sites*”). A cada um dos resíduos referidos corresponde um determinado *pocket* no

local ativo da enzima, denominando-se estes *pockets* por “S1, S2, S3, ..., Sn” (“*non-primed pockets*”) e “S1’, S2’, S3’, ..., Sn’ ” (“*primed pockets*”). Cada *pocket* é determinante para a especificidade do substrato, sendo os *pockets* responsáveis pela estabilização do ligando.<sup>17,36</sup>

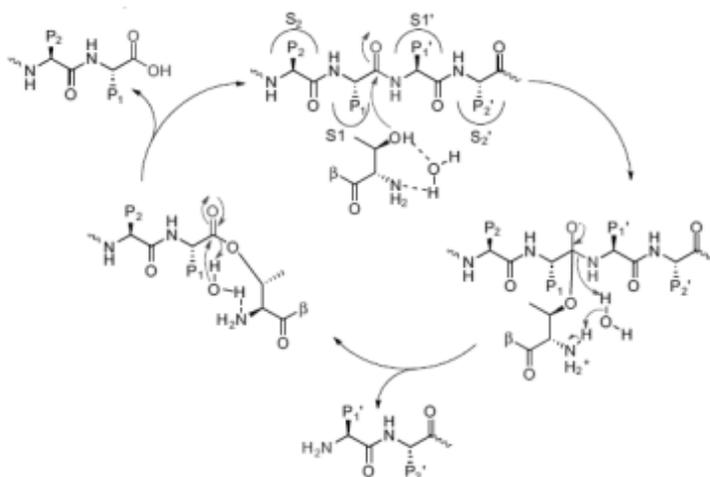
Um aspeto importante a ter em conta é o facto de cada local ativo ser formado por mais do que uma subunidade  $\beta$ . Apesar de ser ao nível das subunidades  $\beta 1$ ,  $\beta 2$  e  $\beta 5$  (no caso do proteassoma constitutivo) e das subunidades subunidades  $\beta 1i$ ,  $\beta 2i$  e  $\beta 5i$  (no caso do proteassoma indutivo) que ocorre a atividade catalítica propriamente dita, para a formação dos *pockets* e posicionamento dos substratos é importante o auxílio de outras subunidades. Assim, no caso do proteassoma constitutivo, o local ativo C-L envolve as subunidades  $\beta 1$  e  $\beta 2$ , o local ativo T-L compreende as subunidades  $\beta 2$  e  $\beta 3$  e o local ativo CT-L é composto pelas subunidades  $\beta 5$  e  $\beta 6$ . No caso do proteassoma indutivo, as subunidades constitutivas são substituídas pelos equivalentes indutivos ( $\beta 1i$ ,  $\beta 2i$  e  $\beta 5i$ ).<sup>17,14,37,38</sup>



**Figura 5** – Os *pockets* de especificidade do substrato e a respectiva nomenclatura dos resíduos do substrato. Adaptado de Huber et al.<sup>14</sup>

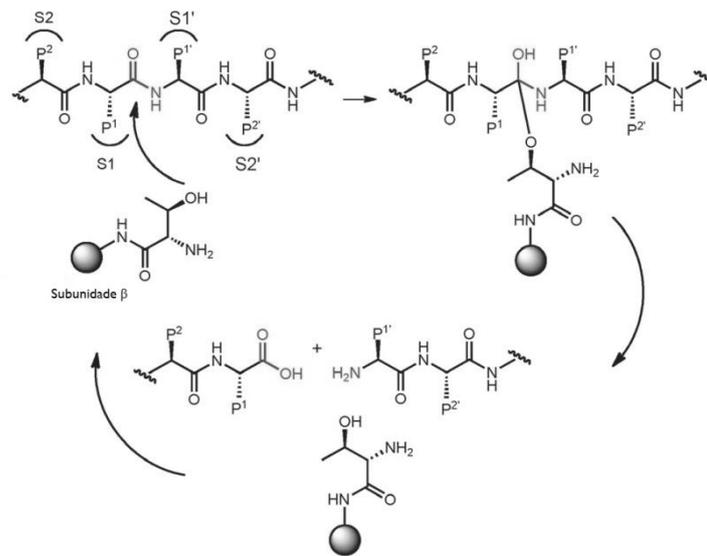
Relativamente aos resíduos de aminoácidos importantes para cada um dos locais ativos, estes são referidos por Huber et al.<sup>14</sup>. Ao nível das subunidades catalíticas dos seres eucariotas (*S. cerevisiae*, *Bos taurus*, *Mus musculus* e *Homo sapiens* – sendo que, com exceção da levedura que só tem subunidades constitutivas, existem subunidades catalíticas constitutivas e indutivas), os aminoácidos importantes em cada local ativo são a treonina 1 (Thr1), o aspartato 17 (Asp17), a lisina 33 (Lys33), a serina 129 (Ser129), o aspartato 166 (Asp166) e a serina 169 (Ser169); no *T. acidophilum*, todos os aminoácidos são iguais com exceção do aminoácido da posição 17 que é o ácido glutâmico (Glu) em vez de um aspartato. A Thr1, o Asp17 e a Lys33 são os resíduos mais importantes no mecanismo proteolítico, enquanto que os restantes resíduos referidos (Ser129, Asp166 e a Ser169) contribuem não só para a catálise, mas também para a integridade estrutural do local ativo. Um aspecto a salientar é o facto de a numeração atribuída aos aminoácidos ter como base o alinhamento das sequências dos seres eucarióticos com a subunidade  $\beta$  do *T. acidophilum*.<sup>14,17,39</sup>

Embora exista uma enorme diversidade de substratos proteicos do proteassoma, o mecanismo catalítico genérico ao nível das subunidades  $\beta$  apresenta etapas semelhantes. No mecanismo utilizado com maior frequência (figura 6), uma molécula de água auxilia na transferência do protão entre o grupo hidroxilo da Thr1 (Thr1 O $\gamma$ ) e o N da Thr1. Assim, o grupo hidroxilo da Thr1 N-terminal do proteassoma atua como nucleófilo e ataca o carbono do grupo carbonilo da ligação peptídica, originando um intermediário acil-enzima e um fragmento peptídico com um novo N-terminal. Posteriormente, ocorre a hidrólise da acil-enzima por adição da molécula de água nucleofílica, sendo o local ativo restaurado e sendo a reação completada pela libertação do C-terminal do produto de clivagem N-terminal.<sup>24,26,28,40</sup>



**Figura 6** – Mecanismo catalítico das subunidades  $\beta$  com intervenção de moléculas de água. Adaptado de Rentsch *et al.*<sup>40</sup>

No entanto, em 2012 foi proposto por Wei *et al.*<sup>41</sup> um novo mecanismo (figura 7) no qual a água não efetua a mediação do processo de ativação, podendo o O $\gamma$  da Thr1 ser ativado diretamente pelo grupo amina do N-terminal (Thr1-N<sup>2</sup>H<sub>2</sub>), baseando-se este estudo em cálculos de mecânica quântica/mecânica molecular (MQ/MM), perturbação da energia livre (*Free Energy Perturbation* – FEP) e simulações de dinâmica molecular (DM).



**Figura 7** – Mecanismo de ação das subunidades  $\beta$  sem intervenção de moléculas de água.  
 Adaptado de Kisselev *et al.* <sup>24</sup>

## II – INIBIDORES DO PROTEASSOMA

## 2.1 – CONSIDERAÇÕES GERAIS

O proteassoma exerce um papel relevante na degradação controlada de diversas proteínas, nomeadamente reguladores celulares que controlam o ciclo celular e a apoptose.<sup>10</sup>

Sendo o proteassoma 26S constituído pelo núcleo proteolítico 20S e por duas partículas reguladoras 19S, trata-se de um complexo proteolítico que apresenta diversos alvos para a intervenção terapêutica, facto este que torna o design de inibidores do proteassoma um desafio.<sup>27,28</sup>

A primeira consequência da inibição do proteassoma verifica-se na diminuição das taxas de degradação proteica nas células, facto este que tem como consequência a rápida acumulação de proteínas de semi-vida curta conjugadas com ubiquitina. Para além disso, a inibição do proteassoma também promove a acumulação de proteínas danificadas ou desnaturadas, sendo que esta acumulação de polipéptidos desnaturados conduz à expressão de “*heat shock proteins*” e proteínas de stress oxidativo do retículo endoplasmático expressas como mecanismos de protecção contra diversas condições tóxicas.<sup>19</sup>

A maioria dos inibidores do proteassoma é de natureza peptídica, sendo a especificidade de ligação destes inibidores aos locais ativos do proteassoma determinada quer pela sequência peptídica quer pela natureza química do farmacóforo.<sup>42</sup> Para que ocorra uma diminuição da proteólise, não é necessário que o inibidor atue em todos os locais ativos.<sup>19</sup> O bloqueio da atividade CT-L é mais eficaz na inibição do crescimento celular *in vitro*. No entanto, estudos recentes mostram que a inibição da atividade CT-L pelos inibidores do proteassoma pode não ser suficiente para obter o máximo de citotoxicidade, sendo habitualmente requerida a co-inibição de outras subunidades catalíticas para se alcançar o máximo de citotoxicidade possível.<sup>42</sup> Apesar de estas conclusões serem importantes para o desenvolvimento de inibidores mais específicos, não nos podemos esquecer que estes estudos foram efetuados em linhas celulares hematológicas neoplásicas, as quais também expressam o imunoproteassoma. Como tal, não está totalmente esclarecido quais as subunidades específicas que devem ser tidas em conta como alvos em tumores sólidos que não expressem o imunoproteassoma.<sup>43,44</sup>

A maioria dos inibidores de locais CT-L são moléculas muito hidrofóbicas e, como tal, muito mais permeáveis a nível celular que os inibidores dos locais T-L ou C-L, os quais contêm resíduos carregados. O local ativo CT-L do proteassoma cliva principalmente após resíduos hidrofóbicos de maiores dimensões. Como tal, pode ser difícil conseguir uma

elevada seletividade na inibição do proteassoma por compostos baseados em péptidos simplesmente através da manipulação da fração peptídica do inibidor. Em alternativa, é necessário utilizar um farmacóforo com preferência pelo N-terminal da treonina.<sup>19</sup> Ou seja, a especificidade dos locais ativos também é influenciada pela natureza química do farmacóforo.<sup>42,45,46</sup>

Em termos de aplicações terapêuticas, os inibidores do proteassoma podem atuar ao nível do tratamento de neoplasias, no transplante de órgãos, em doenças autoimunes (miastenia gravis, esclerose múltipla, artrite reumatóide, lúpus), como tendo atividade anti-inflamatória (asma, psoríase, colite), em patologias neurodegenerativas (doença de Parkinson, doença de Alzheimer) em terapêuticas anti-infecciosas (malária), no tratamento da lesão de reperfusão após AVC e na estimulação do crescimento ósseo e do cabelo.<sup>12,19,24,47</sup>

Nas duas últimas décadas, foram descobertos diversos inibidores estruturais dos três locais ativos que estão presentes na partícula nuclear 20S.<sup>27</sup> Um desses inibidores é o bortezomib (Velcade<sup>®</sup>, PS-341), aprovado pela *Food and Drug Administration* (FDA) em 2003 como terapia de terceira linha para o mieloma múltiplo recidivante e refractário; no entanto, atualmente já é utilizado como tratamento de primeira linha para esta patologia.<sup>28</sup> O bortezomib pode ainda ser utilizado em outras situações clínicas, tais como no linfoma de células do manto e na rejeição aguda do enxerto.<sup>24,47,48</sup> Este fármaco inibe a atividade quimotripsina-like do proteassoma, o que leva à degradação ineficiente de proteínas poliubiquitinadas, ativando a apoptose celular.<sup>49</sup>

O último inibidor a ser aprovado pela FDA foi o carfilzomib (Kyprolis<sup>®</sup>, PR-171), tendo tal ocorrido há pouco mais de um ano (a 20 de Julho de 2012).

Consultando a base de dados do U.S. National Institutes of Health (disponível em <http://www.clinicaltrials.gov/>)<sup>50</sup>, podemos verificar que atualmente se encontram em ensaios clínicos diversos inibidores do proteassoma, isoladamente ou em combinação com outros fármacos (regra geral, fármacos antineoplásicos). Assim, existem diversos ensaios clínicos a decorrer com bortezomib, carfilzomib e com os inibidores de segunda geração ixazomib (MLN-9708), marizomib (salinosporamida A, NPI-0052), delanzomib (CEP-18770) e oprozomib (ONX-0912, PR-047).

O marizomib é, até à data, o mais potente de todos os inibidores do proteassoma que se encontra em ensaios clínicos, destinando-se este ao tratamento do mieloma múltiplo, leucemia, linfomas e tumores sólidos.<sup>24</sup>

## 2.2 – PRINCIPAIS CLASSES DE INIBIDORES DA PARTÍCULA NUCLEAR 20S

A nível estrutural, os inibidores da partícula nuclear 20S são muito diversos, podendo ser divididos em dois grandes grupos dependendo do facto de formarem ou não ligações covalentes com o O $\gamma$  da Thr1 do local ativo. Por sua vez, cada um destes dois grupos pode ser subdividido em diversas classes estruturais. Existem ainda inibidores alostéricos (ou seja, que não interagem diretamente com os locais ativos), assim como inibidores não específicos e inibidores específicos de um local ativo.<sup>19,24,38</sup>

As moléculas pertencentes aos inibidores não covalentes apresentam a vantagem de serem desprovidas de grupos funcionais propensos ao ataque nucleofílico dado que grupos reativos estão geralmente associados a reatividade excessiva, falta de especificidade e instabilidade.<sup>19</sup>

Na tabela 3 encontra-se um resumo das principais classes de inibidores da partícula nuclear 20S, assim como exemplos de compostos pertencentes a cada classe. Esta tabela teve como base a classificação utilizada por Kisselev *et al.*<sup>24</sup>. Na figura 8 estão representadas as estruturas de alguns destes inibidores, sendo possível observar que podem possuir estruturas muito diversificadas.

Grande parte destes inibidores é de origem sintética, baseando-se estes em péptidos e resultam da combinação de uma fração peptídica com um farmacóforo ativo.<sup>19</sup> No entanto, também existem diversos inibidores provenientes de origem natural, nomeadamente a felutamida B (originalmente isolada a partir do *Penicillium fellutanum*), a epoxomicina, a eponemicina, o omuralide, o marizomib (proveniente do microorganismo marinho *Salinispora tropica*), a belactosina A, a glidobactina e a siringolina A isolada a partir da *Pseudomonas syringae*.<sup>24,51,52</sup>

Alguns dos inibidores forem identificados através do varrimento de bibliotecas químicas, nomeadamente as hidroxureias e o PI-083.<sup>24</sup>

No corrente ano, Kikuchi, J. *et al.*<sup>53</sup> publicaram um artigo sobre derivados homopiperazina que inibem o proteassoma e apresentam uma boa biodisponibilidade. No entanto, necessitam ainda de mais estudos, sobretudo ao nível da toxicidade.

Moreira *et al.*<sup>54</sup> publicaram também em 2013 um artigo sobre triterpenóides pentacíclicos tipo oleanano com um grupo boronato em C3 cujo mecanismo de ação inclui a indução da apoptose e a inibição da atividade do proteassoma de células tumorais.

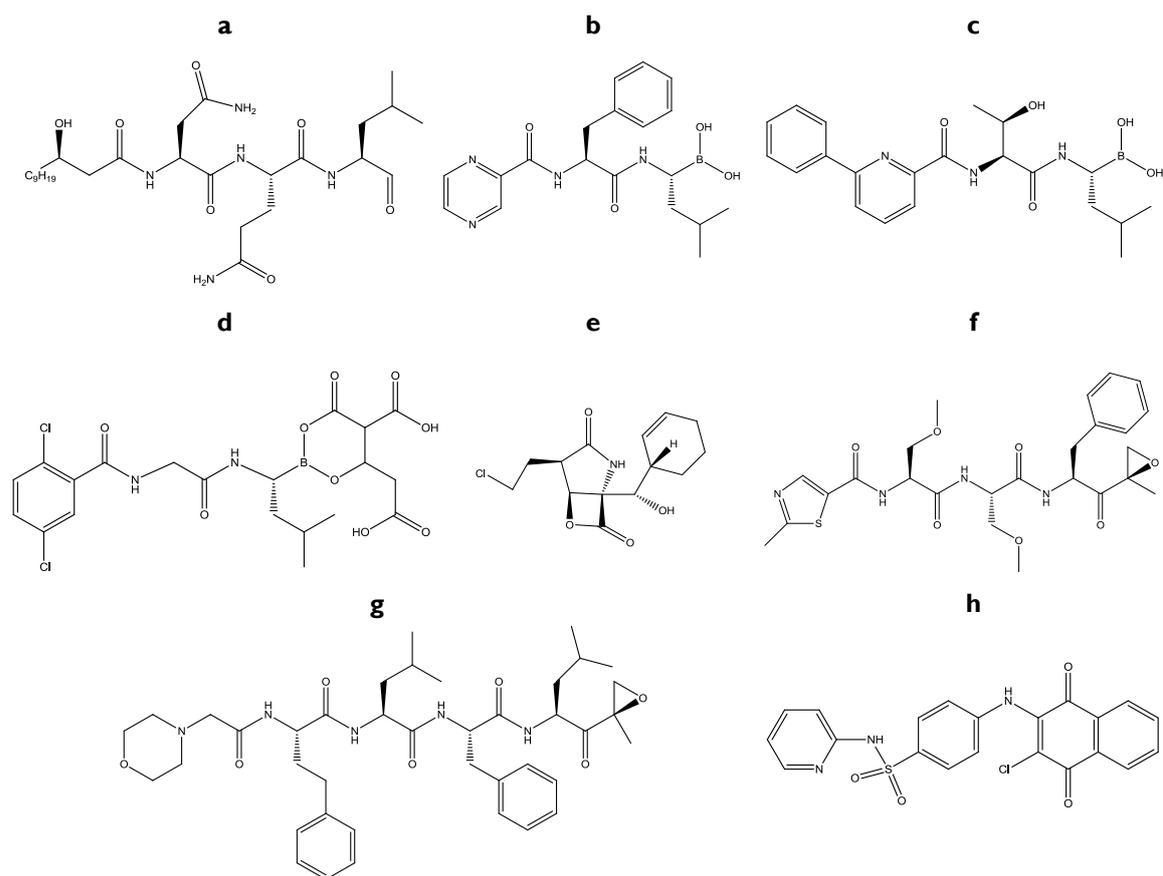
Em termos de envolvimento das diferentes subunidades catalíticas do proteassoma, existem inibidores que podem atuar nas três subunidades catalíticas (felutamida B,

bortezomib, MLN2238, CEP-18770, carfilzomib, siringolina A, TMC-95A, PI-083, K-7174) e inibidores que atuam em apenas uma subunidade catalítica (existem diversas moléculas que exercem apenas atividade CT-L, tais como as hidroxíureias, os péptidos não cíclicos e isósteros peptídicos, o NC-005, o NC-005-VS, YU-101, LU-005; há compostos que exercem apenas atividade C-L, como o YU-102, NC-001; e, por fim, compostos com atividade T-L, como o NC-022). Há ainda a salientar o facto de o inibidor alostérico PR-39 ligar-se à subunidade  $\alpha 7$ .<sup>30,48,51,55,56,57,58</sup>

Existem ainda inibidores específicos de subunidades catalíticas do imunoproteassoma. Por exemplo, o PR-957 atua na subunidade  $\beta 5i$  e o composto IPSI-001 atua na subunidade  $\beta 1i$ .<sup>24</sup>

Tipo de inibidor	Classe estrutural	Exemplos
<b>Inibidores covalentes</b>	Aldeídos	MG-132, PSI, felutamida B
	Boronatos	Bortezomib (MG-341, PS-341), delanzomib (CEP-18770), MLN2238, ixazomib (MLN9708)
	$\alpha',\beta'$ -epoxicetonas	Epoxomicina, eponemicina, carfilzomib (PR-171), oprozomib (ONX-0912, PR-047)
	Cetoaldeídos	
	$\beta$ -lactonas	Omuralide, PS-519, marizomib (NPI-0052, salinosporamida A), belactosina A
	Sulfonas vinílicas	<sup>125</sup> I-NIP-L <sub>3</sub> VS, MV151
	<i>Syrbactins</i>	Syringolin A, glidobactina A
<b>Inibidores não-covalentes</b>	Oxatiazol-2-onas (específicos de proteassomas bacterianos)	HT1171, GL5
	Péptidos cíclicos	TMC 95A, argirina A, <i>scytenomid A</i>
	Péptidos não cíclicos e isósteros peptídicos	Ritonavir, benzilastatina, CVT-659
	Inibidores não peptídicos	PI-083, hidroxíureias
<b>Inibidores não específicos</b>	Hidroxíureias	
		<i>Thiostrepton</i> , siomicina A
<b>Inibidores alostéricos</b>		PR-39, 5-amina-8-hidroxiquinolina
<b>Inibidores site-specific</b>		NC-005, NC-005-VS, YU-101, LU-005, YU-102, NC-001, NC-022

**Tabela 3** – Principais classes de inibidores da partícula nuclear 20S.

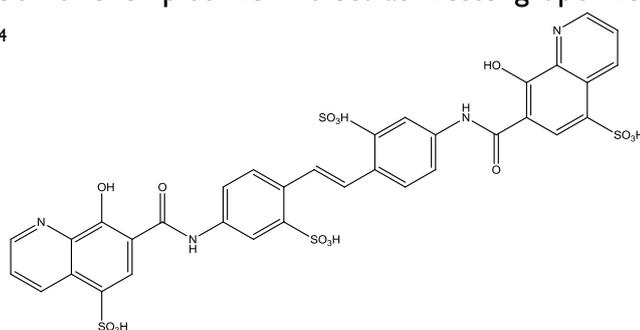


**Figura 8** – Alguns exemplos de inibidores da partícula nuclear 20S.

**a** – fellutamide B, **b** – bortezomib, **c** – CEP-18770, **d** – MLN9708, **e** – marizomib, **f** – ONX-0912, **g** – carfilzomib, **h** – PI-083.

### 2.3 – INIBIDORES DA PARTÍCULA REGULADORA 19S

O desenvolvimento de inibidores da partícula reguladora 19S do proteassoma encontra-se ainda em fase embrionária relativamente ao desenvolvimento de inibidores que atuam na partícula nuclear 20S, sendo uma área que futuramente poderá ser um bom desafio. Assim, podemos pensar na existência de diversos alvos potenciais na partícula reguladora 19S, nomeadamente ATPases, recetores de ubiquitina e enzimas que promovem a desubiquitinação. Como exemplos de moléculas deste grupo de inibidores, temos as ubistatinas (figura 9).<sup>24</sup>



**Figura 9** – Ubistatina A, um inibidor da partícula reguladora 19S.

### III – O DESENVOLVIMENTO DE FÁRMACOS ASSISTIDO POR COMPUTADOR

### 3.1 – O PROCESSO DE DESCOBERTA DE NOVOS FÁRMACOS E O DESENHO DE FÁRMACOS ASSISTIDO POR COMPUTADOR

O desenho de fármacos assistido por computador (*Computer-Aided Drug Design – CADD*) diz respeito à utilização de métodos e ferramentas computacionais para auxiliar na descoberta, desenho e otimização de compostos biologicamente ativos uma vez que fornece, por exemplo, dados relativos às estruturas químicas, interações e propriedades físico-químicas.<sup>45,59,60</sup> O CADD deve ser sinergicamente integrado com a química medicinal e sintética tradicional, assim como com a biologia e a farmacologia de modo a maximizar o seu impacto.<sup>61</sup>

Cada vez mais é necessário encontrar medicamentos mais seguros e eficazes que satisfaçam as exigências cada vez mais rigorosas das autoridades reguladoras. Em média, são necessários cerca de 10 a 15 anos e um enorme investimento financeiro de modo a introduzir no mercado um novo medicamento. A indústria farmacêutica, na tentativa de acelerar este processo, recorre cada vez mais às metodologias computacionais no desenho de fármacos.<sup>62,63</sup>

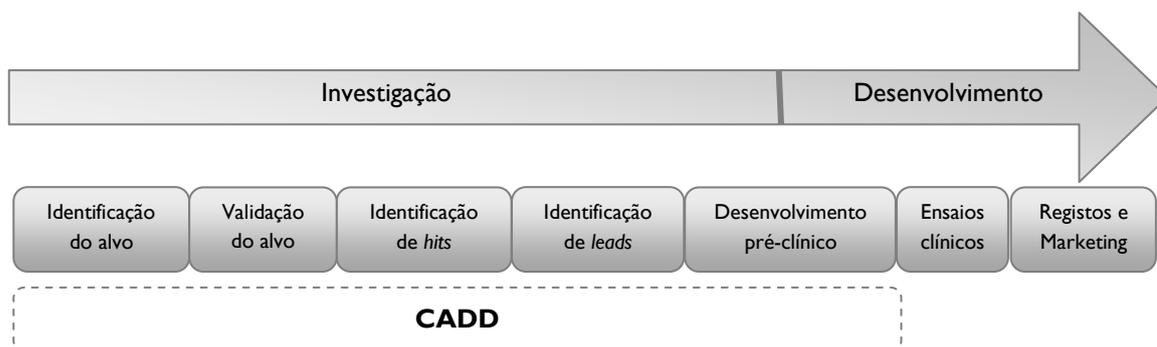
Em termos esquemáticos (figura 10), assim que um alvo é identificado, a etapa seguinte consiste na validação desse mesmo alvo e no desenvolvimento de metodologias que permitam determinar se um dado composto é ativo no alvo em questão. São então selecionados compostos *hit*, ou seja, moléculas que apresentam atividade reprodutível num ensaio biológico.<sup>64</sup> Regra geral, utiliza-se o *High Throughput Screening* (HTS, *screening* de elevado rendimento) para identificar compostos *hit*, consistindo este na automação de bioensaios que são efetuados num elevado número de compostos.<sup>61</sup> A otimização de compostos *hit* é o processo de síntese de compostos que derivam de um composto *hit* de modo a melhorar uma ou mais propriedades (por exemplo, a potência, a solubilidade ou a seletividade).<sup>62</sup> Os compostos *hit* que possuírem as características desejadas passam a compostos *lead* que posteriormente serão otimizados, sendo um composto *lead* um representante de uma série de compostos que possui potencial suficiente para seguir para um programa completo de desenvolvimento.<sup>64</sup>

O HTS e o CADD complementam-se frequentemente, sendo que os métodos computacionais por vezes substituem o HTS devido ao facto de serem mais rápidos, mais económicos e mais fáceis de configurar.<sup>63</sup> Para além disso, o CADD permite tentar otimizar compostos de modo a apresentarem maior afinidade de ligação e boa seletividade, assim

como propriedades farmacocinéticas aceitáveis, não estando este último parâmetro usualmente associado ao HTS.<sup>63</sup>

Como muitas das técnicas utilizadas no CADD são geralmente menos dispendiosas monetariamente e mais rápidas que a maioria dos métodos de ensaio experimentais, são geralmente testadas *in silico* bases de dados de milhares ou milhões de compostos antes de partes destas serem submetidas a testes *in vitro*, o que contribui para melhorar a eficiência de cada processo.<sup>63, 65</sup>

Geralmente, os projetos de desenvolvimento de fármacos têm início com centenas de milhares de compostos ou até mesmo milhões de compostos, podendo estes compostos ser provenientes de compilações corporativas, de catálogos de compostos comercialmente disponíveis para *screening* ou de bibliotecas de compostos virtuais. Assim, o *virtual screening* (VS – *virtual screening*) – também designado por *in silico* screening – é um recurso muito importante, consistindo este, de um modo muito geral, na pesquisa de moléculas em enormes bases de dados virtuais. Após esta abordagem *in silico*, parte desta base de dados é submetida a testes *in vitro*.<sup>63</sup>



**Figura 10** – O processo de descoberta e desenvolvimento de novos fármacos.

Um conceito importante no desenvolvimento de fármacos é o conceito de *drug-likeness*, ou seja, o facto de determinadas moléculas possuírem determinados grupos funcionais e/ou propriedades físico-químicas que se encontram presentes na maioria dos fármacos conhecidos, sendo por exemplo a absorção oral um elemento chave.<sup>64,66</sup>

A base de muitas das regras que tentam prever este comportamento tem em linha de conta a chamada “*Lipinsky rule of 5*”, segundo a qual os fármacos com maior biodisponibilidade oral devem apresentar idealmente as seguintes características: massa molecular inferior a 500 Da, menos de cinco doadores de ligações de hidrogénio, menos de dez aceptadores de ligações de hidrogénio e log P inferior a 5.<sup>58,62,67,68</sup>

O coeficiente de partição (log P) é uma medida experimental importante da lipofilia de uma molécula, estando relacionado com a solubilidade dessa molécula em água e numa fase orgânica (geralmente o octanol). No entanto, deve ter-se em conta que o log P apenas reflete uma propriedade geral, sendo necessários estudos mais pormenorizados das interações moleculares entre ligandos e macromoléculas.<sup>69</sup>

Foi demonstrado que a “*Lipinsky rule of 5*” pode ser aplicada a moléculas neutras ou carregadas positivamente, estando os aniões sujeitos a regras diferentes, podendo o tamanho da sua superfície polar ter algum poder preditivo.<sup>70</sup>

Assim, têm sido desenvolvidos outros meios de estimar computacionalmente a *drug-likeness*. Estes algoritmos tentam prever propriedades físico-químicas, toxicidade, estabilidade metabólica, entre outros exemplos. Estas abordagens são habitualmente denominadas como “filtros gerais” visto que apenas necessitam da estrutura molecular do composto para a realização do cálculo, sendo muito importantes na filtração das bases de dados.<sup>71</sup>

A otimização de compostos *hit* e *lead* segue abordagens semelhantes. As maiores diferenças devem-se ao facto de existirem mais dados na fase de otimização dos compostos *lead* e maior acesso a informação bioestrutural.<sup>45</sup>

O CADD utiliza diversas abordagens de desenvolvimento e otimização baseadas na estrutura de ligandos (tendo como base, por exemplo, a utilização de farmacóforos), abordagens baseadas na estrutura do recetor (das quais é exemplo o *docking* molecular) e ainda pode recorrer ao cálculo da relação estrutura-atividade quantitativa (QSAR – *quantitative structure-activity relationship*), podendo estas metodologias ser utilizadas em simultâneo.<sup>65</sup>

O CADD utiliza vários métodos computacionais para calcular e otimizar a energia de compostos. Métodos como a MM e a MQ são muitas vezes utilizados. Numa breve descrição, a MM é uma tentativa de formular um campo de força que possa servir como modelo computacional para avaliar o potencial energético para todos os graus de liberdade de uma molécula, sendo frequentemente utilizada nas minimizações energéticas iniciais. As alterações energéticas devido a alongamento de ligações, alteração de ângulos de ligação, energias de torção e interações eletrostáticas e de van der Waals são calculadas utilizando parâmetros como comprimentos de ligação, ângulos, torções e cargas. Assim, um campo de forças pode ser definido como um conjunto de equações e parâmetros que, quando avaliados por um sistema molecular, produz determinada energia em função das coordenadas atómicas. Esta energia expressa o desvio estrutural relativamente aos valores ideais.<sup>72</sup>

Uma vez que os cálculos de MM são rápidos, este é o método mais frequentemente utilizado na otimização inicial na estrutura de moléculas. Existem diversos campos de força que podem ser utilizados. Por exemplo, o AMBER99 que se encontra parametrizado para o estudo de proteínas e o Merck Molecular Force Field (MMFF), sendo este último mais utilizado para pequenas moléculas orgânicas.<sup>72</sup>

À medida que o estudo evolui, é muitas vezes necessário efetuar cálculos mais rigorosos através de MQ visto que esta permite descrever corretamente o comportamento dos eletrões em termos matemáticos.<sup>45</sup>

A MQ pode dividir-se em dois métodos: *ab initio* e semi-empírico. O primeiro método consiste em técnicas *ab initio* que descrevem o comportamento químico tendo em linha de conta todos os eletrões. São técnicas mais rigorosas e que não requerem parâmetros pré-definidos mas que, no entanto, dependem de funções que descrevem as orbitais num sistema. O segundo método denomina-se “semi-empírico”, sendo este mais rápido e apropriado a compostos maiores do que os métodos *ab initio* uma vez que se utilizam parâmetros pré-definidos; para além disso, este método apenas considera os eletrões das orbitais de valência para a construção das orbitais atômicas.<sup>58,61,69,72</sup>

Denomina-se por conformação bioativa a conformação de menor energia adotada por um dado ligando de modo a que haja complementaridade com o *pocket* de ligação. Em termos das conformações adoptadas pelas moléculas, estas podem ser estudadas através de métodos estatísticos como a dinâmica molecular (DM) e as simulações Monte Carlo (MC).

A DM é uma técnica computacional que integra a lei do movimento de Newton ( $F = ma$ ) e utiliza campos de força para gerar diversas conformações através da simulação do movimento browniano e vibracional, ou seja, simula-se um ambiente de temperatura mais elevada de modo a que sejam superadas barreiras energéticas. Trata-se de um auxílio para detetar as conformações dos mínimos locais e do mínimo global, resultando numa trajetória que descreve a forma como as posições e velocidades das partículas variam em função do tempo.<sup>72</sup> A DM pode ser aplicada em diversas áreas do desenho de fármacos, nomeadamente na previsão da seletividade do alvo, na otimização de modelos de homologia, na pós-análise de múltiplas conformações no *docking* e no desenvolvimento de farmacóforos.<sup>45</sup>

Os métodos MC geram conformações aleatórias do composto antes de minimizarem as suas energias e reter a estrutura (se for a estrutura de menor energia). À partida, esta abordagem aleatória consegue efetuar uma amostragem de todas as conformações possíveis, mas tal só

acontece se estiver a decorrer durante tempo suficiente (depende da flexibilidade do ligando, mas tal não pode ser definido antes da experiência ter início).<sup>62,72</sup>

Em termos do design de novos compostos, a relação estrutura-atividade (SAR, *structure-activity relationship*) apresenta-se como uma informação de extrema importância, permitindo esta sugerir modificações químicas que possam originar compostos com as características desejadas.<sup>45,62</sup>

O QSAR tenta fazer a previsão estatística do efeito das modificações estruturais efetuadas num determinado esqueleto químico (*scaffold*) relativamente à atividade biológica medida experimentalmente, podendo os modelos matemáticos obtidos ser utilizados para prever a atividade de outros compostos.<sup>61</sup>

A combinação do QSAR com informação estrutural tridimensional denomina-se relação estrutura-atividade quantitativa tridimensional (3D-QSAR) e inclui abordagens que utilizam métodos estatísticos para correlacionar a variação na atividade química ou biológica com informação na estrutura tridimensional de uma série de compostos.<sup>69</sup>

Em termos de aplicações do QSAR, é de salientar a sua importância no estudo da farmacocinética (absorção, distribuição, metabolização, excreção – ADME) e a previsão de diversos aspetos toxicológicos, nomeadamente ao nível da toxicidade aguda, mutagenicidade e carcinogenicidade.<sup>61</sup>

No CADD são ainda muito importantes métodos como a modelação por homologia, o *docking* molecular, o *virtual screening* e a construção de modelos farmacofóricos. Estes métodos vão ser descritos de seguida com pormenor uma vez que são utilizados na parte prática deste trabalho.

### **3.2 – MODELAÇÃO POR HOMOLOGIA**

Experimentalmente, a elucidação estrutural de proteínas pode ser obtida através de cristalografia de raio-X e de ressonância magnética nuclear (RMN).<sup>45</sup> Apesar da aplicação crescente da RMN, a cristalografia de raio-X continua a ser o método mais relevante para obter a estrutura de complexos proteína-ligando por com base na estrutura do recetor. Basicamente, na cristalografia de raio-X são gerados cristais do complexo proteína-ligando por precipitação lenta de uma solução utilizando, por exemplo, solventes orgânicos ou detergentes. É então dirigido um feixe de raios-X ao cristal, ocorrendo difração deste feixe de acordo com o arranjo dos átomos do cristal, podendo então produzir mapas de difração

de raios-X. É possível construir uma matriz tridimensional rodando o cristal no feixe de raios-X em intervalos regulares. A RMN é uma técnica que permite determinar propriedades de uma substância através da correlação da energia absorvida relativamente à frequência. Baseia-se na transição entre os diversos níveis energéticos produzida pela absorção de energia de alta frequência (radiofrequência) que tem lugar quando um núcleo atómico com *spin* (momento angular intrínseco) não nulo é submetido a um campo magnético estacionário de intensidade própria.<sup>45,62</sup>

A RMN tem duas vantagens relativamente à cristalografia de raio-X: a proteína não precisa de ser cristalizada e, para além disso, a RMN fornece informação dinâmica em tempo real (enquanto que as estruturas cristalinas fornecem estruturas termodinamicamente estáveis e estáticas). No entanto, a RMN tem como maior desvantagem o facto de o tamanho molecular limite rondar os 35 000 Da.<sup>58,61</sup>

As coordenadas atómicas que têm como base dados obtidos por difração de raios-X e que fornecem informação estrutural tridimensional de macromoléculas biológicas são geralmente colocadas em bases de dados (tais como o *Protein Data Bank* – PDB, <http://www.rcsb.org/pdb/home/home.do>)<sup>73</sup> e podem servir como *input* inicial para cálculos de análise conformacional teórica, assim como para *design* com base na estrutura do recetor.<sup>72,74</sup>

No entanto, existe uma enorme diferença em termos da quantidade de sequências de aminoácidos conhecidas e estruturas tridimensionais disponíveis (atualmente, Setembro de 2013, existem 93624 estruturas de proteínas no PDB). Como tal, quando não existem estruturas das proteínas obtidas experimentalmente, é necessário utilizar modelos computacionais para gerar as estruturas tridimensionais do alvo em estudo, baseando-se estes modelos estruturas cristalográficas conhecidas de estruturas homólogas.<sup>45,62,73</sup>

Segundo Vyas *et al.*<sup>75</sup>, “a modelação por homologia consiste na representação da similaridade dos resíduos circundantes nas posições topológicas correspondentes à proteína de referência”. A modelação por homologia tem como objetivo prever a estrutura tridimensional de proteínas tendo como base o facto de que proteínas com sequências de aminoácidos homólogas apresentam estruturas similares.<sup>76</sup> Para além disso, sabe-se que pequenas alterações na sequência de aminoácidos de uma proteína originam pequenas variações na estrutura tridimensional dado que a conformação estrutural de uma proteína encontra-se mais conservada do que a sua sequência de aminoácidos.<sup>75</sup>

O alinhamento das sequências de aminoácidos é extremamente importante para a obtenção de um bom modelo de homologia. Para tal, recorre-se à utilização de métodos de alinhamento das sequências de forma a determinar quais as regiões conservadas da proteína

de referência e da proteína cuja sequência de aminoácidos é desconhecida, sendo também uma abordagem importante para procurar sequências homólogas.<sup>72</sup>

Assim, tendo acesso à estrutura obtida experimentalmente de uma dada proteína (*template*), podem ser obtidos modelos de homologia para uma sequência homóloga (alvo) que apresente pelo menos 30% de percentagem de identidade ao nível da sequência de aminoácidos.<sup>74,75</sup>

Podemos então considerar que os modelos que apresentam mais de 50% de identidade em termos de sequência são suficientemente precisos para utilizar no processo de descoberta de fármacos.<sup>74,75</sup>

Segundo Cavasotto e Phatak<sup>74</sup>, de uma forma geral, podemos considerar as seguintes etapas na construção de modelos de homologia:

1. Identificação e seleção da estrutura tridimensional da proteína que vai ser utilizada como *template* [por exemplo, a obtenção de uma lista das estruturas armazenadas no PDB que melhor correspondem à sequência para a qual queremos obter o modelo de homologia pode ser pesquisadas no servidor *Basic Local Alignment Search Tool* (BLAST – <http://blast.ncbi.nlm.nih.gov/BLAST.cgi>)<sup>77</sup> ou utilizando alguns *softwares* – como o *Molecular Operating Environment* (MOE)<sup>78</sup> ou o *Modeller*<sup>79</sup>, podendo as estruturas cristalográficas ser posteriormente obtidas no PDB];
2. Alinhamento das sequências de aminoácidos das proteínas alvo e *template* (a sequência de aminoácidos da proteína alvo pode ser obtida, por exemplo, no Universal Protein Resource – UniProt (<http://www.uniprot.org>)<sup>80</sup> em formato FASTA);
3. Construção do modelo tridimensional da proteína alvo tendo como base a estrutura tridimensional do *template* e o alinhamento das sequências;
4. Refinamento e avaliação/validação do modelo de homologia.

Após a obtenção do modelo de homologia, este é otimizado/refinado através de uma etapa de minimização energética, utilizando um campo de forças de mecânica molecular podendo ainda, de seguida, serem aplicadas técnicas de DM, MC e algoritmos genéticos.

No entanto, há que não esquecer que os modelos podem conter erros. Tendo em conta a percentagem de identidade entre as sequências de aminoácidos do alvo e do *template*, assim como a qualidade do alinhamento, os modelos de homologia gerados podem diferir 1 a 2 Å em termos dos carbonos alfa ( $C_{\alpha}$ ) correspondentes.<sup>74</sup>

A ocorrência de erros na obtenção de um modelo de homologia pode ser avaliada através das seguintes formas:<sup>75</sup>

1. Sobrepôr o modelo obtido com a estrutura nativa e calcular o RMSD dos  $C_{\alpha}$ ;
2. Obter um *score* que meça a significância estatística entre as estruturas cristalográficas que podem ser utilizadas para construir o modelo;
3. Desenvolvimento de uma função de *scoring* que permita classificar qualitativamente os modelos de homologia obtidos em “bons” e “maus”.

Por sua vez, as metodologias de validação enquadram-se geralmente em dois grupos distintos: análise da estereoquímica da proteína (simetria, geometria – quiralidade, comprimentos de ligação, ângulos de ligação, ângulos de torção) e da qualidade estrutural; análise da correspondência entre a forma como a sequência e a estrutura encaixam uma na outra (*fitness*) e atribuição de um *score* para cada resíduo que se posiciona corretamente.<sup>75</sup>

Os gráficos de Ramachandran são extremamente importantes na avaliação da qualidade da estrutura tridimensional proteica. Os ângulos de torção da cadeia polipeptídica (também denominados por ângulos de Ramachandran) constituem os parâmetros estruturais locais mais importantes que controlam o *folding* das proteínas, estando relacionados com as rotações das cadeias polipeptídicas em torno das ligações entre o N do  $C_{\alpha}$  (denominado Phi,  $\phi$ ) e o C do  $C_{\alpha}$  (denominado Psi,  $\psi$ ). Assim, o gráfico de Ramachandran permite analisar a distribuição dos ângulos de torção de uma estrutura proteica, assim como as regiões permitidas ou não pelos valores dos ângulos de torção.<sup>75,81</sup>

A modelação por homologia encontra aplicações tão diversas como o design de fármacos, estudo da interação do ligando com o recetor, estudo da especificidade de substratos e pode ainda auxiliar na elucidação de estruturas de cristalografia de raio-X, RMN e microscopia eletrónica.<sup>75</sup>

Uma vez que a determinação da estrutura tridimensional de uma proteína através de métodos experimentais (tais como RMN ou cristalografia de raio-X) requer imenso tempo e não é eficiente com todas as proteínas (sobretudo se forem proteínas membranares), a construção de modelos de homologia *in silico* constitui uma alternativa viável e apelativa em termos de tempo e de investimento financeiro, originando comprovadamente modelos rigorosos e muito úteis na descoberta de novos fármacos. No entanto, é importante não esquecer que a qualidade dos modelos obtidos por modelação por homologia influencia muito o VS e a obtenção de bons resultados de *docking*.<sup>75</sup>

Em termos dos servidores e softwares disponíveis para a modelação por homologia, podemos destacar, por exemplo, o SWISS-MODEL<sup>82</sup>, o Protein Model Portal<sup>83</sup> o Modbase<sup>84</sup>, o MOE<sup>78</sup> e o Modeller<sup>79</sup>.

### 3.3 – MODELAÇÃO FARMACOFÓRICA

#### 3.3.1 – CONSIDERAÇÕES GERAIS

A definição de farmacóforo sofreu alguma evolução ao longo dos tempos. A primeira definição de farmacóforo foi sugerida por Paul Ehrlich em 1909 como sendo “a estrutura molecular que carrega as características essenciais (*phoros*) responsáveis pela atividade biológica de um fármaco (*pharmacon*)”. Em 1977, Peter Gund atualizou a definição de farmacóforo para “conjunto de características estruturais de uma molécula que é reconhecido pelo recetor e que é responsável pela atividade biológica da molécula”. Em 1998, a *International Union of Pure and Applied Chemistry* (IUPAC) definiu farmacóforo como sendo “o conjunto de características estéricas e eletrónicas necessárias para assegurar as interações supramoleculares com um determinado alvo biológico de forma a desencadear (ou inibir) a sua resposta biológica”.<sup>45,85</sup> Assim, numa proteína alvo podem atuar diferentes moléculas mas que se ligam da mesma forma dado que apresentam o mesmo farmacóforo.<sup>65</sup>

Em termos de construção do farmacóforo, os grupos funcionais não são considerados num nível atómico (por exemplo, -OH), são sim consideradas as suas características de grupo, por exemplo aceitador de ligações de hidrogénio, dador de ligações de hidrogénio, carga positiva, carga negativa, regiões hidrofóbicas e aromáticas.<sup>62</sup> Podemos ainda construir farmacóforos tridimensionais, sendo estes um conjunto de propriedades e arranjos tridimensionais que o composto possui para ser ativo num dado teste biológico.<sup>58,61</sup>

A modelação farmacofórica é um dos métodos mais rápidos e eficazes para a descoberta de um novo *scaffold*, sendo o objetivo do farmacóforo ser utilizado como modelo de forma a identificar moléculas que possuam características semelhantes às suas. Os farmacóforos podem utilizar-se para proceder a uma pesquisa rápida em bases de dados de milhares de compostos, podendo chegar-se a compostos bastante diferentes dos iniciais. Como tal, trata-se de uma ferramenta muito importante na procura de compostos *lead*.<sup>58,61</sup>

A modelação farmacofórica trata-se de uma técnica que pode ser efetuada tendo como base a estrutura de ligandos, a estrutura do local ativo das proteínas ou ainda, numa

abordagem mais recente, a formação do complexo ligando-recetor devido à interação do(s) ligando(s) com o local ativo da proteína.<sup>65,86</sup>

Nos métodos de geração de farmacóforos com base na estrutura de ligandos podem ser utilizadas duas abordagens: obtenção do farmacóforo através de características compartilhadas por diversos ligandos e farmacóforos com base no QSAR.<sup>65</sup> Na primeira abordagem, é efetuada a construção de modelos farmacofóricos tendo como base características químicas comuns a um determinado conjunto de ligandos ativos no recetor alvo em questão. Para tal, é utilizado um algoritmo para proceder ao alinhamento molecular das estruturas dos ligandos para, de seguida, determinar e selecionar quais as características presentes. No método de geração de farmacóforos com base na QSAR, podem-se utilizar os valores da concentração inibitória a 50% ( $IC_{50}$ ) ou da constante de inibição ( $K_i$ ) para tentar correlacionar a disposição tridimensional das características químicas com as atividades biológicas de um determinado conjunto de moléculas. No entanto, a elucidação farmacofórica com base em ligandos pode apresentar algumas limitações, nomeadamente a determinação da conformação bioativa.<sup>65,85</sup>

Em termos dos programas que podem ser utilizados para gerar farmacóforos com base em ligandos, podemos referir o DS Catalyst da Accelrys<sup>87</sup> (que inclui o algoritmo Catalyst/HipHop para a abordagem com base em características comuns aos ligandos, destinando-se o algoritmo Catalyst/Hypogen à geração de farmacóforos com base na QSAR), o Phase da Schrodinger<sup>88</sup> (para farmacóforos com base na QSAR) e o MOE<sup>78</sup> (que utiliza a “elucidação farmacofórica” para gerar hipóteses farmacofóricas e alinhar as moléculas dos diversos ligandos tendo em conta todas as combinações possíveis, existindo ainda a funcionalidade “consensus” que sugere hipóteses farmacofóricas tendo como base um conjunto de compostos ativos alinhados, sendo os cálculos efetuados tendo um raio de tolerância).<sup>65</sup>

Nos métodos de modelação farmacofórica com base na estrutura do recetor, os farmacóforos são obtidos tendo como base as estruturas tridimensionais do recetor, representando os farmacóforos obtidos as possíveis interações do ligando com os aminoácidos do local ativo do recetor. De um modo geral, estes modelos farmacofóricos tridimensionais têm em linha de conta cerca de seis tipos de características químicas (aceitador/doador de ligações de hidrogénio, anião/catião, anel aromático, grupo hidrofóbico) e restrições de volume. Os farmacóforos com base no recetor podem ser utilizados para identificar o modo como os ligandos interagem com a proteína, para diferenciar compostos e para obter compostos *lead* no *screening in silico*.<sup>65</sup>

Os *softwares* disponíveis para a abordagem farmacofórica com base na estrutura incluem, por exemplo, o DS Structure Based Pharmacophore da Accelrys<sup>89</sup>, o MOE<sup>78</sup> e o Phase<sup>88</sup>.

Nos métodos de geração de farmacóforos com base na interação do ligando com o recetor utiliza-se uma combinação das características farmacofóricas do ligando com a forma do recetor, podendo ser obtida uma hipótese tridimensional bastante seletiva capaz de selecionar muito menos compostos *hit* do que as buscas convencionais.<sup>65</sup>

O DS Catalyst<sup>87</sup> e o MOE<sup>78</sup> são exemplos de softwares que podem ser utilizados nesta abordagem de modelação farmacofórica.<sup>65</sup>

Através da combinação das características farmacofóricas com restrições de forma e volumes de exclusão é possível otimizar o farmacóforo de modo a diminuir o número de moléculas que possuem uma estrutura tridimensional adequada mas cuja forma não é a mais adequada, prevenindo-se assim que encaixem no local ativo da proteína alvo.<sup>65</sup> No entanto, as hipóteses farmacofóricas têm de ser avaliadas e validadas.<sup>86</sup> Relativamente à validação de farmacóforos, habitualmente utiliza-se um conjunto de compostos teste (que consiste num conjunto de compostos ativos e outro de compostos inativos) para verificar se o modelo farmacofórico é suficientemente robusto e seletivo de modo a detetar os compostos ativos e não detetar os compostos inativos. Estando disponíveis valores de IC<sub>50</sub> observados e previstos, estes podem ser correlacionados.<sup>90,91,92,93,94</sup>

Assim que temos disponíveis modelos farmacofóricos tridimensionais, podemos filtrar bases de dados com o objetivo de selecionar compostos que apresentem as características presentes no modelo farmacofórico. A complexidade do farmacóforo aumenta à medida que se adicionam múltiplas características e volumes de exclusão (fornecem informação sobre regiões do local de ligação que não se encontram acessíveis aos ligandos); como tal, o tempo de cálculo também aumenta.<sup>85</sup> Uma possível solução para este tempo de pesquisa eventualmente excessivo baseia-se em realizar a busca em “camadas”, começando por usar um filtro para excluir os compostos que não têm os grupos funcionais necessários e depois selecionar apenas os compostos que passaram neste teste para comparar com o farmacóforo.<sup>45,62</sup> Para além da inspeção visual, podemos recorrer a outros parâmetros como propriedades físico-químicas de *drug-likeness* ou *lead-likeness* para reduzir a lista de compostos selecionados.<sup>85</sup>

Os farmacóforos são representações importantes que podem posteriormente ser aplicadas no VS com base em farmacóforos, no *de novo design*, na obtenção de compostos *lead* e na otimização de compostos.

Embora de extrema importância e utilidade, há que ter a noção de que o farmacóforo é uma representação que não consegue explicar tudo. Como tal, é extremamente importante ter sempre presente as limitações deste conceito.

### 3.3.2 – DEFINIÇÃO E REPRESENTAÇÃO DAS CARACTERÍSTICAS QUÍMICAS

Na interação ligando-recetor, as principais forças envolvidas na ligação seletiva do ligando ao recetor macromolecular são pontes de hidrogénio, interações eletrostáticas e interações hidrofóbicas. A representação das características de um farmacóforo envolve dois passos. No primeiro, a molécula deve ser dividida num conjunto de elementos ou características, cada um capaz de um tipo particular de interação intermolecular com a proteína. Em segundo lugar, cada característica deve ser representada por um ou mais pontos, que podem ser utilizados para correlacionar diferentes moléculas através do método dos mínimos quadrados.<sup>95</sup>

De modo a identificar ligandos biologicamente ativos, as características representam complementaridades químicas entre o recetor e o espaço tridimensional, correspondendo a diferentes grupos químicos que apresentam similaridade em termos das suas propriedades.<sup>65</sup>

O principal objetivo do uso de características químicas consiste em permitir que as moléculas sejam sobrepostas de forma ponderada. A qualidade de uma sobreposição pode ser avaliada através do cálculo do desvio quadrático médio (RMSD), ou seja, é avaliada pelo desvio resultante entre os pontos das características que coincidem. Há que fazer um balanço da vantagem/desvantagem de incluir todas as características possíveis e ter informação suficiente para originar um modelo farmacofórico significativo. Por exemplo, se predominarem as características hidrofóbicas, estas podem dominar o alinhamento. Embora assim se possa obter um melhor modelo de acordo com o *score*, o modelo em si pode ser menos útil para trabalho posterior uma vez que irá ser menos discriminatório e seletivo.<sup>96</sup>

Para uma boa relação estrutura-atividade não é suficiente identificar quais as características que estão presentes em cada molécula de ligando, é também necessário determinar a localização associada a cada uma das características que serão posteriormente sobrepostas no farmacóforo resultante. A maioria dos programas usados hoje em dia associa características doadoras e aceitadoras, não só com os átomos do ligando correspondente, mas também com a presumível localização do(s) átomo(s) complementare(s) da proteína envolvido(s) na interação. Esta estratégia tem a vantagem de permitir a sobreposição de dois

ligandos que estabeleçam ligações de hidrogénio com o mesmo átomo da proteína, mas a partir de diferentes localizações e direções.<sup>96</sup>

### 3.3.2.1 – GRUPOS DOADORES DE LIGAÇÕES DE HIDROGÉNIO

Podemos considerar como doadores de ligações de hidrogénio o grupo hidroxilo, hidrogénios ligados a azoto, tióis e grupos -CH acetilénicos. No entanto, os dois últimos são controversos e vários investigadores excluem-nos por serem doadores de ligações de hidrogénio muito fracas.<sup>95</sup>

Outro aspeto muito importante a ter em conta é a ionização. Por exemplo, o hidroxilo do COOH por vezes não é considerado doador dado que, a pH fisiológico, é provável que o ácido se encontre ionizado. Em contrapartida, podem incluir-se neste grupo aminas básicas (por exemplo,  $RCH_2NMe_2$ ) visto que a pH fisiológico estão protonadas. Assim, as incertezas relativas aos estados de ionização e tautoméricos dificultam a análise de um farmacóforo dado que podem inverter a natureza de uma característica.<sup>95</sup>

### 3.3.2.2 – GRUPOS ACEITADORES DE LIGAÇÕES DE HIDROGÉNIO

Regra geral, é considerado aceitador de ligações de hidrogénio qualquer átomo de azoto, oxigénio ou enxofre que tenha, pelo menos, um par de eletrões disponível (não deslocalizado). No entanto, existem forte indícios de que alguns tipos de átomos de oxigénio devem ser excluídos deste grupo nomeadamente os que se encontram em anéis de furano ou oxazol, cujas evidências teóricas e cristalográficas indicam que são aceitadores muito fracos.<sup>97</sup>

### 3.3.2.3 – GRUPOS CARREGADOS POSITIVAMENTE E NEGATIVAMENTE

As características positivas ou negativas são definidas por átomos que possuem cargas formais (tendo em conta os prováveis estados de ionização *in vivo*), a menos que estejam ligados a um átomo com carga oposta. Alguns autores têm o cuidado de referir especificamente características ionizáveis positivas e negativas enfatizando, por exemplo, que o grupo  $RNH_3^+$  é considerado uma característica positiva, mas não o  $RNMe_3^+$ . Isto faz todo o sentido porque as interações intermoleculares destes dois grupos são muito diferentes, o

primeiro pode estabelecer ligações de hidrogénio muito fortes e o último, na melhor das hipóteses, apenas consegue estabelecer ligações de hidrogénio muito fracas.<sup>96</sup>

### 3.3.2.4 – GRUPOS HIDROFÓBICOS

Decidir quais os átomos ou grupos que devem ser considerados hidrofóbicos tais como anéis aromáticos, grupos isopropilo ou cadeias alifáticas, não é uma tarefa simples. Existem algoritmos que procuram primeiro anéis (atribuindo um valor de hidrofobicidade ao centro do anel, ponderado pela hidrofobicidade dos átomos que o constituem) e, em seguida, avaliam grupos como o CF<sub>3</sub> e depois cadeias alifáticas. No entanto, há que salientar que não existe uma forma restrita de atribuir a localização de uma determinada característica dado que as regiões hidrofóbicas de diferentes moléculas podem não ser sobreponíveis. No geral, as interações hidrofóbicas não se encontram tão bem definidas geometricamente como as ligações de hidrogénio e, como tal, são difíceis de localizar numa pequena região da molécula.<sup>96</sup>

## 3.4 – DOCKING MOLECULAR

### 3.4.1 – CONSIDERAÇÕES GERAIS

O *docking* molecular define-se como sendo o processo que envolve a previsão da conformação e orientação de um ligando num sítio alvo de uma proteína da forma energeticamente mais favorável. Trata-se de um procedimento que pode ser um desafio devido à enorme flexibilidade conformacional, sobretudo ao nível do recetor macromolecular.<sup>59,60</sup>

Podemos considerar três tipos de *docking*: o *docking* rígido (o ligando e a proteína são tratados como entidades rígidas, podendo o ligando ter movimentos de rotação e translação no local ativo de ligação do alvo), o *docking* semi-flexível (apenas o ligando é tratado de forma flexível) e o *docking* totalmente flexível (tanto o ligando como a proteína são tratados de forma flexível).<sup>45,62</sup>

Podemos ter em conta, pelo menos, três critérios para avaliar a exatidão dos métodos de *docking*: a precisão na previsão da pose do ligando, a estimativa da energia livre

do complexo proteína-ligando e a capacidade de distinguir entre compostos ativos e inativos.<sup>98</sup>

O *docking* requer a existência de algoritmos de procura que consigam prever o modo como cada composto se liga, assim como a utilização de funções de *scoring* para prever a afinidade recetor-ligando.<sup>58,61</sup>

Muitas vezes é necessário utilizar diversas técnicas para tentar prever o modo como os compostos de interesse se ligam. A primeira etapa consiste em identificar o local de ligação (o *binding pocket*) na proteína alvo; no entanto há que ter em conta que o ligando pode ligar-se num local distante do local de ligação clássico, denominando-se então os ligandos de alostéricos. Assim que o local de ligação é identificado, este deve ser caracterizado em termos de polaridade, eventuais áreas com carga, quais os resíduos que estão disponíveis para interagir com os ligandos, se existem ou não grupos metálicos a interagir, qual a dimensão do *pocket*, entre outros exemplos.<sup>45,58,61</sup>

Outro aspeto extremamente importante num cálculo de *docking* molecular está relacionado com o facto de se incluírem ou não moléculas de água no local de ligação. Na maioria dos casos, a água está envolvida na interação do ligando com a proteína. No caso do software GOLD (Genetic Optimisation for Ligand Docking), as moléculas de água podem ser adicionadas ou retiradas, podendo ter também alguma mobilidade de modo a otimizar a interação com o ligando *dockado*.<sup>70</sup>

Uma das desvantagens do *docking* molecular deve-se ao facto de ser necessário a existência de uma base de dados de compostos, o que significa que o *docking* não gera compostos completamente novos, apenas pode prever a possibilidade de um composto existente ter uma dada atividade biológica.<sup>62</sup>

Alguns dos programas de *docking* mais conhecidos e utilizados incluem o FlexX<sup>99</sup>, o GOLD<sup>100</sup>, o Glide<sup>101</sup>, entre outros.<sup>102</sup>

### 3.4.2 – ALGORITMOS DE PROCURA

Os *softwares* de *docking* molecular utilizam diversos métodos de procura (algoritmos) de modo a permitirem gerar múltiplas conformações do ligando *dockado* e, desta forma, ter em conta a flexibilidade conformacional do ligando. Estes algoritmos podem ser classificados em três categorias básicas: métodos sistemáticos, métodos estocásticos ou aleatórios e métodos de simulação.<sup>71</sup>

É de salientar que os algoritmos dependem da existência de uma função de *scoring* robusta e precisa de modo a que possa ser identificado corretamente o modo e a afinidade de ligação.<sup>71</sup>

#### **3.4.2.1 – MÉTODOS SISTEMÁTICOS**

Como exemplos deste grupo podemos incluir a procura conformacional, a construção incremental de uma pose ótima de ligação (presente nos *softwares* FlexX<sup>99</sup> e Surflex<sup>103</sup>), o algoritmo de Hammerhead e bases de dados que exploram todos os graus de liberdade de uma molécula de modo a colocar os ligandos (fragmentos moleculares) no centro ativo das proteínas.<sup>98</sup>

#### **3.4.2.2 – MÉTODOS ESTOCÁSTICOS OU ALEATÓRIOS**

Os programas que efetuam procuras estocásticas utilizam algoritmos Monte Carlo (o caso dos *softwares* ICM<sup>104</sup> e GlamDock<sup>105</sup>) ou algoritmos genéticos (no caso do GOLD<sup>100</sup>, Moldock<sup>106</sup>) para avaliar as conformações flexíveis do ligando no local de ligação. Os métodos de Monte Carlo geram diversas conformações do ligando no centro ativo da proteína e, posteriormente, fazem o *scoring* dessas conformações; os algoritmos genéticos lidam com princípios de competição biológica e de dinâmica populacional para colocar o ligando no centro ativo, ocorrendo a manipulação de um conjunto de estruturas denominadas cromossomas, sendo que cada uma destas estruturas representa uma possível orientação do ligando relativamente ao local de ligação proteico, sendo-lhe atribuído um *score*.<sup>98,107,108</sup>

#### **3.4.2.3 – MÉTODOS DE SIMULAÇÃO**

Os métodos de simulação, tais como a dinâmica molecular e a minimização de energia, descrevem o processo de *docking* molecular com mais pormenor. A dinâmica molecular utiliza as forças atómicas e as massas para determinar as posições atómicas sobre uma série de passos temporais muito pequenos. Por sua vez, os procedimentos de minimização energética são geralmente empregues juntamente com outros métodos de procura, já que é somente atingido um mínimo de energia local.<sup>72</sup>

### 3.4.3 – FUNÇÕES DE SCORING

As funções de *scoring* para as interações ligando-proteína são extremamente importantes no CADD. A função de *scoring* ideal deve ser capaz de reconhecer a pose de ligação nativa de um ligando na superfície proteica relativamente a poses que não são as corretas, devendo ainda ser capaz de prever com rigor a afinidade de ligação (ou a energia livre de ligação) de modo a que as moléculas ativas possam ser distinguidas das não ativas.<sup>109</sup>

O *scoring* consiste na previsão teórica da afinidade ligando-proteína. A maioria das funções de *scoring* avalia interações polares (tais como ligações de hidrogénio) e interações apolares (tais como a liberdade rotacional).<sup>72</sup>

Com as funções de *scoring* podemos ordenar energeticamente cada composto relativamente aos restantes compostos testados, podendo tal ser efetuado durante o *docking* (de modo a otimizar a colocação do ligando) e após o *docking* (para classificar o complexo formado).<sup>45,58</sup> No entanto, diferentes funções de *scoring* podem funcionar melhor em diferentes complexos proteína-ligando.<sup>71</sup>

Todos os programas de *docking* e funções de *scoring* podem gerar um número significativo de falsos positivos. Para minimizar esta questão, pode-se combinar a filtração farmacofórica com métodos de *docking*, assim como utilizar diversas funções de *scoring*.<sup>71</sup>

Podemos considerar três grupos de funções de *scoring* quando pretendemos fazer o screening de centenas ou milhares de compostos: funções com base em campos de forças, funções empíricas e funções com base no conhecimento.<sup>107</sup>

#### 3.4.3.1 – FUNÇÕES COM BASE EM CAMPOS DE FORÇAS

As funções de *scoring* que têm como base campos de forças (das quais é exemplo a função GOLDScore do *software* GOLD<sup>100</sup> que é utilizado neste trabalho) utilizam campos de força de mecânica molecular para quantificar a soma das energias relacionadas com as energias de interação (tais como interações eletrostáticas e de Van der Waals, interações ligando-recetor e as energias internas do ligando).<sup>98</sup>

Para que a afinidade possa ser prevista com precisão, é necessário flexibilidade e adaptabilidade da proteína aquando da ligação do ligando. No entanto, a atual tecnologia não consegue satisfazer este requisito. Para além disso, a ausência de termos entrópicos e de solvatação é o lado fraco das tradicionais funções de *scoring* com base em campo de forças.<sup>107</sup>

### 3.4.3.2 – FUNÇÕES EMPÍRICAS

As funções de *scoring* empíricas (das quais a ChemScore do software GOLD<sup>100</sup> é um exemplo) aproximam a afinidade de ligação do complexo proteína-ligando como sendo uma soma das interações ponderadas. Este tipo de funções habitualmente tem em linha de conta ligações de hidrogénio, interações iónicas, interações hidrofóbicas, termos de dessolvatação e entropia de ligação.<sup>71,98</sup>

Este tipo de funções permite obter dados experimentais, tais como energias de ligação, como a soma de várias funções parametrizadas.<sup>71</sup>

### 3.4.3.3 – FUNÇÕES KNOWLEDGE-BASED

Este tipo de funções tem como base a análise estatística da informação estrutural de estruturas cristalográficas disponível de complexos proteína-ligando.<sup>71</sup>

Estas funções de *scoring* baseiam-se na hipótese de que uma amostra de dados suficientemente grande pode servir para derivar regras e princípios gerais inerentemente armazenados nessa base de conhecimento. Cada composto é comparado com os outros, sendo favorecidos os contactos preferenciais e penalizadas as interações repulsivas. Trata-se de um método que tenta sobretudo gerar poses corretas e não tanto calcular energias de ligação. Como exemplo de funções *knowledge-based* temos a função Drugscore do software FlexX.<sup>71,99,107</sup>

## 3.5 – VIRTUAL SCREENING

O VS é um método que consiste na seleção de compostos a partir de bases de dados virtuais, sendo uma técnica cada vez mais utilizada para identificar compostos *hit* no processo de desenvolvimento de fármacos.<sup>62,85</sup>

A seleção de compostos por VS pode seguir uma abordagem baseada na estrutura de ligandos (*ligand-based*) ou na estrutura do recetor (*structure-based*, também conhecida por *target-based*).

O VS baseado na estrutura de ligandos recorre a um conjunto de compostos ativos conhecidos como ponto de partida para construir modelos (por exemplo, farmacóforos),

podendo ainda ser utilizados procedimentos de procura baseados na semelhança e na subestrutura.<sup>45,58,61</sup>

O VS baseado na estrutura do recetor utiliza o *docking* molecular ou farmacóforos construídos com base no recetor para pesquisar compostos que se prevê que sejam complementares do local de ligação (*binding pocket*) da proteína alvo devido a eventuais interações que possam ocorrer.<sup>45,58,61</sup> O objetivo desta abordagem consiste em prever os modos de ligação e as afinidades dos compostos quando interagem com a proteína de modo a identificar as moléculas que melhor encaixam no local ativo e que têm afinidades mais favoráveis devido a interações específicas com o recetor.<sup>85,110</sup>

Habitualmente, preparam-se as estruturas cristalográficas proteicas antes de ser efetuado o *docking*, estando incluídas nesta etapa, por exemplo, a correta protonação e a otimização das ligações de hidrogénio. No caso dos ligandos, estes podem ser preparados em termos de minimização da estrutura tridimensional e dos estados de ionização antes de se efetuar o VS embora seja um processo computacionalmente dispendioso.<sup>110,111</sup>

A vantagem dos métodos baseados na estrutura do recetor deve-se ao facto de conseguirem incorporar informação do tamanho e forma do local ativo sob a forma de volumes de exclusão (previnem eventuais interações estéreas adversas).<sup>62</sup>

Tanto a abordagem baseada na estrutura do ligando como a abordagem baseada na estrutura do recetor permitem identificar conjuntos de compostos com potencial atividade que posteriormente devem ser testados experimentalmente em ensaios *in vitro*.<sup>45,62</sup>

Qualquer que seja a estratégia de VS utilizada, pode ser muito útil ir aumentando gradualmente a complexidade da procura de modo a maximizar o tempo e os recursos disponíveis. De facto, podem ser utilizados farmacóforos 3D para filtrar grandes bases de dados antes de o *docking* ser efetuado e mesmo após o *docking*. Apesar de os modelos farmacofóricos serem menos representativos do reconhecimento molecular, estes apresentam a vantagem de derivarem diretamente de características-chave do local ativo e/ou de ligandos conhecidos.<sup>70,85</sup> Filtrar compostos indesejáveis devido a fraca *drug-likeness* é também uma estratégia muito útil quando se utilizam métodos intensivos de procura, como o *docking* molecular.<sup>45</sup>

Ao contrário do *docking*, o VS efetuado com base em farmacóforos reduz os problemas decorrentes da insuficiente consideração da flexibilidade da proteína, ou da utilização de funções de *scoring* insuficientemente otimizadas, visto que existe um raio de tolerância para cada característica farmacofórica. No entanto, o VS com base em farmacóforos pode ter como limitação uma elevada taxa de falsos positivos e/ou falsos

negativos dado que apenas uma pequena percentagem dos compostos *hit* virtuais é realmente bioativa. Esta elevada taxa de falsos positivos pode dever-se ao facto de a restrição estérea do alvo macromolecular não ser suficientemente considerada e também devido ao facto de a flexibilidade ser tratada através da introdução de um raio de tolerância para cada característica farmacofórica, o que impossibilita uma completa flexibilidade macromolecular. Para além disso, a maioria das interações entre ligando e proteína são sensíveis à distância (sobretudo interações de curto alcance, como é o caso das interações eletrostáticas), sendo estas dificilmente incorporadas num modelo farmacofórico.<sup>86</sup>

O VS com base em farmacóforos conjugado com o *docking* tem-se revelado extremamente importante na descoberta de fármacos e no aumento da eficiência na otimização de compostos *lead*.<sup>86</sup>

Podemos então efetuar a comparação entre o VS e o HTS. Em termos da disponibilidade dos compostos, o HTS necessita que os compostos a testar estejam de facto disponíveis, enquanto que o VS pode aplicar-se a compostos virtuais que podem ser comprados ou sintetizados se de facto houver interesse nisso, o que representa uma enorme vantagem económica.<sup>45,62</sup>

Dependendo do método utilizado, o VS pode ser aplicado a milhões de compostos numa escala de tempo relativamente curta, sendo muito mais rápido do que o HTS.<sup>61</sup>

Relativamente às desvantagens do VS, estas incluem o facto de ser necessário que os compostos sejam testados, havendo a necessidade de efetuar ensaios laboratoriais. Para além disso, se o VS for utilizado com restrições muito rígidas, tal poderá limitar a possibilidade de encontrar novos compostos. Uma outra questão que se coloca é a fiabilidade dos métodos aplicados; mas, comparando como o HTS, a taxa de sucesso é comparável e, muitas vezes, até melhor.<sup>45,62</sup>

Ao contrário do HTS, no VS é necessário haver conhecimento sobre os fatores espaciais e energéticos que são responsáveis pela interação do ligando com o recetor. Como tal, deve estar disponível a conformação tridimensional da estrutura do alvo macromolecular (através de análises cristalográficas, RMN ou modelos de homologia) ou, pelo menos, um ligando de referência com uma conformação bioativa rígida conhecida.<sup>70</sup>

No entanto, o verdadeiro valor do VS deve-se à sua complementaridade com o HTS. A utilização do VS permite libertar recursos para testar outros compostos que não estão na coleção de *screening*, aumentando a qualidade do HTS.<sup>61,70</sup>

Alguns exemplos de casos de sucesso em que se aplicou o VS incluem a descoberta de novos inibidores da lipoxigenase por Kenyon *et al.*<sup>112</sup> e a identificação de novos inibidores da falciparina-2 (relevante para a malária) por Li *et al.*<sup>113</sup>

### 3.5.1 – O SOFTWARE GOLD

O GOLD é um programa para efetuar cálculo de *docking* molecular que possibilita realizar o *docking* flexível de ligandos nos locais ativos das proteínas, tendo sido este o *software* utilizado na parte prática da tese, motivo pelo qual se encontra descrito mais pormenorizadamente.<sup>114</sup>

Este *software* permite que a proteína apresente uma flexibilidade parcial e que o ligando tenha total flexibilidade, podendo considerar-se que este programa consiste em três partes principais: uma função de *scoring*, um mecanismo para colocar o ligando no local ativo e um algoritmo de procura.<sup>115,110</sup>

Existem quatro funções de *scoring* disponíveis no GOLD: Goldscore, ChemScore, CHEMPLP e ASP (*Astex Statistical Potential*), sendo ainda possível definir uma nova função de *scoring* se tal for necessário.<sup>114</sup> Como a função utilizada na parte prática deste trabalho foi o GoldScore, iremos dar-lhe maior ênfase.

A função de *scoring* GoldScore baseia-se numa expressão com quatro componentes:

- Energia das ligações de hidrogénio proteína-ligando (ligações de hidrogénio externas – *external H-bond*);
- Energia de van der Waals proteína-ligando (*external vdw*);
- Energia interna de van der Waals do ligando (*internal vdw*);
- Energia das ligações de hidrogénio intramoleculares do ligando (ligações de hidrogénio internas – *internal H-bond*).<sup>114,115</sup>

Quando o *score* é calculado, a *external vdw* é multiplicada por um factor de 1.375, sendo esta uma correcção empírica para correção do cálculo do contacto hidrofóbico proteína-ligando.<sup>115</sup>

Em termos da expressão que permite calcular o *score* na função de *fitness* Goldscore, esta é a seguinte:

$$\text{GoldScore} = \text{S}(\text{hb\_ext}) + \text{S}(\text{vdw\_ext}) + \text{S}(\text{hb\_int}) + \text{S}(\text{vdw\_int})$$

$S(\text{hb\_ext})$  representa a contribuição das ligações de hidrogénio proteína-ligando,  $S(\text{vdw\_ext})$  é a contribuição das interações de van der Waals,  $S(\text{hb\_int})$  é a contribuição das ligações de hidrogénio intramoleculares e  $S(\text{vdw\_int})$  é a contribuição da tensão intramolecular no ligando.<sup>115,110</sup>

Relativamente ao mecanismo de posicionamento do ligando no local de ligação, o GOLD utiliza um método que se baseia em pontos de ajuste (*fitting points*), método este que adiciona pontos de ajuste aos grupos que estabelecem ligações de hidrogénio na proteína e no ligando, traça pontos aceitadores no ligando e pontos aceitadores na proteína e vice-versa. Para além disso, o GOLD origina pontos de ajuste hidrofóbicos na cavidade proteica na qual os grupos CH do ligando são colocados.<sup>110</sup>

O algoritmo de procura para explorar os modos de ligação possíveis é um algoritmo genético (*genetic algorithm – GA*), permitindo otimizar os seguintes parâmetros: os diedros das rotações das ligações do ligando; a geometria dos anéis do ligando; diedros dos grupos OH ou  $\text{NH}_3^+$  da proteína; mapeamento dos pontos de ajuste, ou seja, qual a posição do ligando no local de ligação.<sup>110</sup>

As definições do GA influenciam diretamente a duração da corrida de *docking* e a probabilidade de encontrar as melhores poses. A duração da corrida de *docking* e a precisão deste são sobretudo afectados pelo número de *dockings* e pelo número de operações do GA em cada cavidade de *docking*. Se reduzirmos o número de corridas de *docking*, o processo torna-se mais rápido mas, no entanto, há que ter em linha de conta que um maior número de corridas aumenta o espaço de procura conformacional e, conseqüentemente, a probabilidade de obter os resultados corretos.<sup>110</sup>

A eficiência de procura pode ser utilizada para controlar a velocidade do *docking* e prever a exatidão (a confiança) dos resultados. Se a eficiência de procura for de 100%, O GOLD tenta aplicar as configurações ideais para cada ligando. Se a eficiência de procura for de 50%, o *docking* duplica a velocidade mas o espaço de procura será menos explorado. Em contrapartida, se a eficiência de procura for superior a 100%, a procura será mais exaustiva, mas mais lenta.<sup>114</sup>

## IV. RESULTADOS E DISCUSSÃO

#### 4.1 – DESENVOLVIMENTO DE MODELOS DE HOMOLOGIA DAS SUBUNIDADES CATALÍTICAS DO PROTEASSOMA

Sendo o objetivo deste trabalho identificar novos inibidores do proteassoma a partir do VS de uma base de dados de compostos, para a etapa de *docking* molecular é necessário conhecer a estrutura da proteína alvo.

Pela consulta do Protein Data Bank, verificamos que, até este momento, não existe disponível nenhuma estrutura cristalográfica do proteassoma humano. As diversas estruturas cristalográficas do proteassoma/imunoproteassoma encontradas no PDB são provenientes de *archae*, bactérias (*Mycobacterium tuberculosis* – 3KRD, 2FHH; *Rhodococcus* – 1Q5Q), leveduras (*Saccharomyces cerevisiae* – por exemplo, 3UN4, 3UN8) e, no que diz respeito a mamíferos, *Bos Taurus* (código PDB: 1IRU) e *Mus musculus* (códigos PDB: 3UNH, 3UNE, 3UNF, 3UNB) – anexo I.

Assim, devido à ausência de estruturas experimentais do proteassoma humano, é necessário recorrer a métodos computacionais para efetuar a modelação por homologia de modo a obter a estrutura das subunidades pretendidas de modo a que seja possível chegar à estrutura tridimensional que posteriormente será utilizada nos cálculos de *virtual screening* de uma base de dados.

Tal como já foi referido anteriormente, a modelação por homologia é uma técnica que permite prever a estrutura tridimensional de proteínas tendo como base o facto de que proteínas que apresentam uma sequência de aminoácidos similar possuem estruturas similares. A partir de uma estrutura proteica homóloga obtida experimentalmente (*template*), é possível construir modelos para uma sequência homóloga da proteína alvo que apresente similaridade em termos da sequência de aminoácidos. A qualidade do modelo de homologia depende da qualidade do alinhamento da sequência (percentagem de homologia) e da qualidade da estrutura do *template* (resolução).<sup>34</sup>

Ao nível do proteassoma, apenas Lei *et al.*<sup>34</sup> desenvolveram um modelo de homologia (com o servidor SWISS-MODEL) para as subunidades  $\beta 1i$  e  $\beta 2i$  humanas, utilizando para tal as sequências de aminoácidos das subunidades em questão (retiradas da base de dados Uniprot).

Através do algoritmo “Basic Local Alignment Search Tool” (BLAST – <http://blast.ncbi.nlm.nih.gov/Blast.cgi>)<sup>77</sup> é possível encontrar regiões que apresentam similaridade entre sequências. Assim, este algoritmo compara as sequências de proteínas ou de nucleótidos com sequências armazenadas em bases de dados e calcula a significância

estatística das comparações efetuadas.<sup>77</sup> Utilizando o BLAST no UniProt (<http://www.uniprot.org/>)<sup>80</sup> e o MOE versão 2010.10, verificámos que o *Mus musculus* e o *Bos taurus* são as espécies que possuem cadeias cujas sequências de aminoácidos apresentam maior percentagem de identidade com as estruturas correspondentes do proteassoma humano (por exemplo, no caso da subunidade  $\beta 5c$  murina a percentagem é de 95% e na subunidade  $\beta 5c$  do bovino a percentagem é de 96%). No caso da levedura, as percentagens variam um pouco consoante a subunidade em questão, sendo que na subunidade  $\beta 5c$  a percentagem de identidade relativamente à estrutura humana correspondente é de 67%. No anexo 2 podemos encontrar todos os valores calculados para a percentagem de identidade das diversas subunidades  $\beta$  do proteassoma de diversos mamíferos (homem, rato e bovino) e de levedura.

Através da informação contida nos diversos códigos PDB (que nos indicam quais as cadeias proteicas que constituem cada subunidade e ainda qual o código Uniprot correspondente a cada subunidade) e nos códigos Uniprot (através dos quais é possível obter a sequência de aminoácidos de cada subunidade) e utilizando o *software* MOE 2010.10 que nos permite visualizar as cadeias e respetivas sequências de aminoácidos, foi possível verificar quais as cadeias envolvidas na interação de um dado ligando com o proteassoma (anexo I).

Pela análise das estruturas cristalográficas disponíveis para o proteassoma, nomeadamente das interações que os inibidores estabelecem com este, verificou-se que não basta utilizar as cadeias das subunidades catalíticas individualmente para construir os modelos de homologia. Uma vez que o *pocket* de ligação é constituído por duas cadeias (sendo a cadeia não catalítica importante para posicionar o substrato), há que ter em linha de conta que, no caso do proteassoma constitutivo, o local ativo C-L envolve as subunidades  $\beta 1c$  e  $\beta 2c$ , o local ativo T-L compreende as subunidades  $\beta 2c$  e  $\beta 3c$  e o local ativo CT-L é composto pelas subunidades  $\beta 5c$  e  $\beta 6c$ . Relativamente ao proteassoma indutivo, as subunidades constitutivas são substituídas pelos equivalentes indutivos ( $\beta 1i$ ,  $\beta 2i$  e  $\beta 5i$ ).<sup>14,17</sup>

Embora todas as subunidades catalíticas do proteassoma sejam importantes, por uma questão prática optámos por centrar o nosso estudo no local ativo CT-L, o qual é constituído pelas subunidades  $\beta 5c$  e  $\beta 6c$ . Apesar de existirem estudos que indiquem que a inibição da atividade CT-L pelos inibidores do proteassoma pode não ser suficiente para obter o máximo de citotoxicidade, sendo habitualmente requerida a co-inibição de outras subunidades catalíticas para se alcançar o máximo de citotoxicidade possível, sabe-se que o bloqueio da atividade CT-L é mais eficaz na inibição do crescimento celular *in vitro*.<sup>42</sup> Para

além disso, o estudo pormenorizado dos restantes locais ativos do proteassoma exigia mais tempo de pesquisa disponível de modo a efetuar um trabalho igualmente minucioso como o que foi desenvolvido para a atividade catalítica da subunidade  $\beta 5c$ .

Analisando a literatura disponível, verificámos que não foram ainda desenvolvidos modelos para as subunidades  $\beta 5c$  e  $\beta 6c$  do proteassoma humano.

Apesar de o *Mus musculus* apresentar a sequência de aminoácidos com elevada percentagem de identidade relativamente ao *Homo sapiens*, as estruturas cristalográficas murinas disponíveis (3UNB, 3UNE, 3UNF e 3UNH) não puderam ser utilizadas para a modelação por homologia das subunidades  $\beta 5c$  e  $\beta 6c$ ; para além das estruturas cristalográficas do *Mus musculus*, o *Bos taurus* é o único mamífero com estrutura cristalográfica disponível até ao momento (código PDB: IIRU). A justificação para a não utilização das estruturas cristalográficas mencionadas deve-se ao facto de o posicionamento da subunidade  $\beta 6c$  não estar totalmente de acordo com os dados da bibliografia, tendo-se verificado que a subunidade em questão nas estruturas cristalográficas corresponde subunidade  $\beta 1c/\beta 1i$ . Para além desta questão dos *pockets*, a estrutura cristalográfica disponível do *Bos taurus* não se encontra complexada com um ligando, pelo que a estrutura cristalográfica em questão pode não se encontrar na forma ativa.

Assim, tendo em conta as condicionantes anteriores referidas, optou-se por efetuar a modelação por homologia das subunidades das subunidades que constituem o local ativo CT-L a partir de uma estrutura cristalográfica de levedura (*Saccharomyces cerevisiae*) dado que as estruturas de levedura têm as subunidades posicionadas de acordo com os dados encontrados na bibliografia, possuem também uma percentagem de identidade elevada (67% na subunidade  $\beta 5c$  e 50% na subunidade  $\beta 6c$ ), as estruturas do proteassoma encontram-se complexadas com diversos inibidores e possuem ainda uma resolução bastante razoável (2.70 Å).

Em termos de etapas, o processo de modelação por homologia foi efetuado da seguinte forma:

#### **I. Identificação e seleção da estrutura tridimensional da proteína que vai ser utilizada como *template***

Foi selecionado o código PDB 3UN8 [“Proteassoma 20S de levedura em complexo com PR-957 (epóxido)”, resolução: 2.70 Å], apresentando-se este complexado com o ligando PR-957, o que é útil em termos da posterior validação do modelo de homologia. Não se optou pelo código PDB 3UN4 [“Proteassoma 20S de

levedura em complexo com PR-957 (morfolina)”, resolução: 3.40 Å] devido ao facto de este apresentar uma pior resolução.

No caso da atividade CT-L (cujo local ativo é formado pelas subunidades  $\beta$ 5c e  $\beta$ 6c), começou-se por selecionar para *template* as cadeias K e L do código PDB 3UN8 [visto que a subunidade  $\beta$ 5c (PRE2) é constituída pelas cadeias K e Y, sendo estas cadeias iguais, pelo que basta utilizar uma; e a subunidade  $\beta$ 6c (C5) é constituída pelas cadeias L e Z, das quais basta utilizar apenas uma pelo mesmo motivo de  $\beta$ 5c].

## **2. Alinhamento das sequências de aminoácidos das proteínas alvo e *template***

A sequência das subunidades alvo foi obtida no UniProt (<http://www.uniprot.org>)<sup>80</sup> em formato FASTA. Para a subunidade  $\beta$ 5c de *Homo sapiens* o código Uniprot é o P28074 e para a subunidade  $\beta$ 6c o código Uniprot de *Homo sapiens* é o P28072. No software MOE 2010.10 utilizou-se a funcionalidade “Align” para alinhar as sequências alvo e *template*.

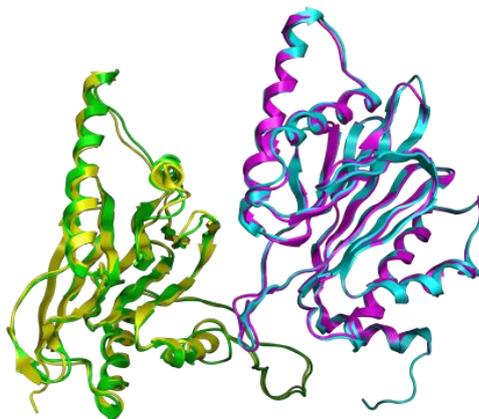
## **3. Construção do modelo tridimensional da proteína alvo tendo como base a estrutura tridimensional do *template* e o alinhamento das sequências**

A modelação por homologia foi efetuada utilizando a funcionalidade “Homology model” do software MOE 2010.10 e utilizando o campo de forças Amber99 (mais apropriado para proteínas). O modelo de homologia de cada subunidade foi efetuado individualmente, tendo as duas subunidades que constituem o *pocket* sido unidas após a obtenção dos modelos de homologia humanos.

## **4. Refinamento e avaliação/validação do modelo de homologia**

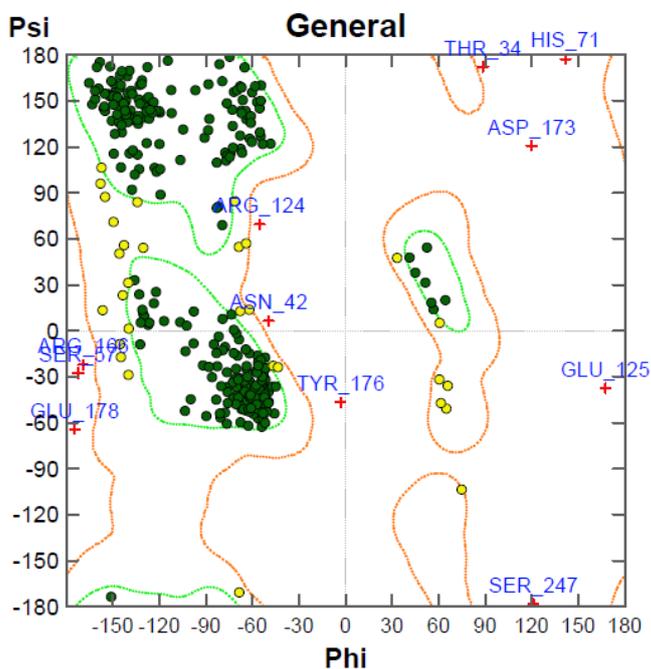
Em termos do refinamento do modelo de homologia obtido, este foi realizado no MOE 2010.10 com o campo de forças Amber99: para a protonação recorreu-se à funcionalidade “Protonate 3D” (tendo esta sido efetuada a pH 7.4 e à temperatura de 300K), tendo a minimização energética sido realizada através da funcionalidade “Energy minimize”. Em termos de avaliação do modelo de homologia obtido, comparando a estrutura final relativamente ao PDB inicial (3UN8), o RMSD obtido a partir da funcionalidade “Protein Superpose” do MOE 2010.10 apresenta um valor de 0.74 Å, sendo este um valor que reflete um bom grau de semelhança estrutural entre as estruturas utilizadas como *template* e o

modelo de homologia obtido dado que o RMSD é inferior a 2.0 Å (regra geral, é este o valor limite de RMSD).<sup>110</sup> Na figura 11, encontra-se a sobreposição do modelo de homologia obtido para o local ativo CT-L (constituído pelas subunidades  $\beta 5c$  e  $\beta 6c$ ) com as subunidades correspondentes do código PDB 3UN8.



**Figura 11** – Sobreposição do modelo de homologia humano da subunidade CT-L do proteassoma com as subunidades de levedura iniciais utilizadas.  
 Legenda de cores: azul – subunidade catalítica da levedura ( $\beta 5c$ ); rosa – subunidade catalítica homóloga humana ( $\beta 5c$ ); amarelo – subunidade complementar do *pocket* da levedura ( $\beta 6c$ ); verde – subunidade complementar homóloga humana ( $\beta 6c$ ).

Na funcionalidade do MOE 2010.10 “*Measure – protein geometry*” é possível obter o gráfico Ramachandran do modelo de homologia obtido (gráfico 1); podemos então observar que a maioria dos resíduos de aminoácidos se encontra em zonas nucleares (“*core*”) (resíduos representados a verde), existindo também resíduos que se encontram em zonas permitidas (representados a amarelo), sendo os resíduos *outlier* (cruzes vermelhas) muito poucos e, para além disso, nestes não se encontra nenhum resíduo importante para a atividade catalítica.



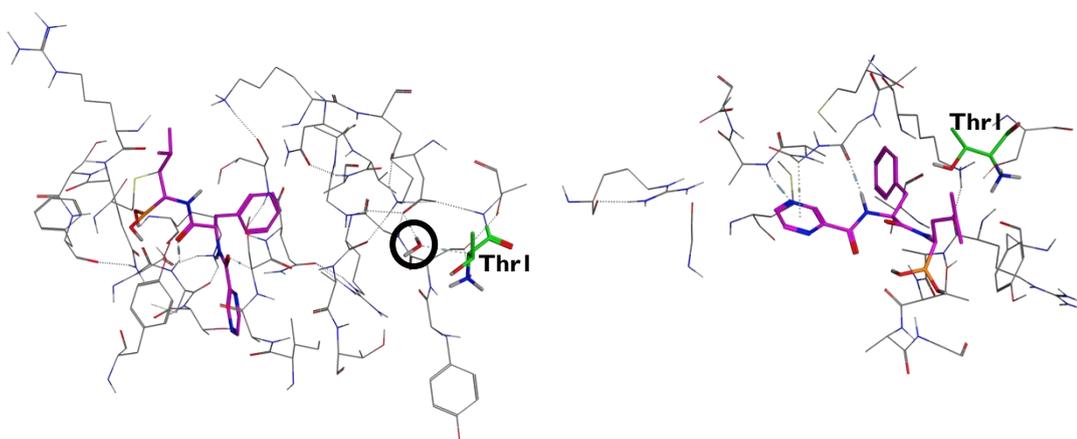
**Gráfico 1** – Gráfico de Ramachandran do modelo de homologia humano da subunidade CT-L do proteassoma. Verde – região nuclear (*core*), amarelo – região permitida, cruces vermelhas – resíduos *outlier*.

Uma questão bastante relevante analisada durante todo este processo deve-se ao facto de serem ou não incluídas moléculas de água nos modelos de homologia. Tal como foi referido na componente teórica desta tese, existem dois mecanismos possíveis ao nível das subunidades catalíticas: um mecanismo em que a água auxilia na transferência do protão entre o grupo hidroxilo da ThrI (ThrI O $\gamma$ ) e o N da ThrI, e um outro mecanismo em que a água não efetua a mediação do processo de ativação, podendo o O $\gamma$  da ThrI ser ativado diretamente pelo grupo amina do N-terminal.<sup>24,41</sup>

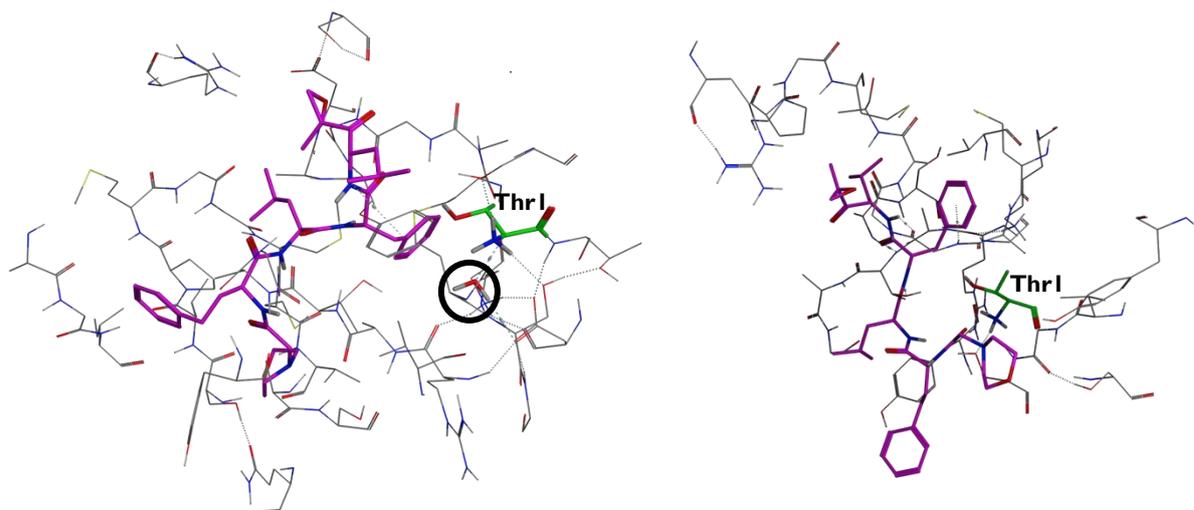
Em termos das estruturas cristalográficas disponíveis, verificou-se que algumas possuem moléculas de água perto da ThrI das subunidades catalíticas (por exemplo, nos códigos PDB 3UN8, 3UNF, IIRU, 2F16), enquanto que outras estruturas cristalográficas não possuem moléculas de água nas subunidades catalíticas (por exemplo, nos códigos PDB 3MG0, 3SHJ, 2ZCY). Para testar a importância e validade da inclusão de águas no modelo decidimos colocar uma molécula de água a cerca de 3 Å da ThrI da subunidade  $\beta$ 5c no modelo de homologia deste *pocket* e *dockar* alguns inibidores bem conhecidos do proteassoma (bortezomib, carfilzomib, marizomib e PI-083) que atuam sobretudo ao nível desta subunidade catalítica (na escala do nM), tendo também sido efetuado o *docking* destes compostos num modelo de homologia na ausência de moléculas de água. A colocação da molécula de água a cerca de 3 Å teve como base a informação retirada da visualização das estruturas cristalográficas que contêm moléculas de água próximas da ThrI (nomeadamente, os códigos PDB 3UN8, 3UNF, 3KRD e IJD2). As estruturas dos compostos foram construídas e otimizadas através do *software* MOE 2010.10, tendo-se atribuído as cargas atómicas parciais e efetuado a protonação de todas as estruturas através da funcionalidade “Protonate 3D” a pH 7.4 e à temperatura de 300K. De seguida, efetuou-se a minimização energética utilizando o campo de forças MMFF94x. Cada *docking* não covalente foi efetuado com o *software* GOLD 5.1.0, utilizando a função de *scoring* Goldscore (a função de *scoring* Goldscore tinha sido inicialmente validada com o bortezomib e apresentou os melhores resultados), *dockando* cada composto junto à ThrI, com um raio de 10 Å, efetuando 500 corridas para cada ligando e guardando as 10 melhores poses de *docking* de cada ligando.

Pela análise dos resultados obtidos, verificou-se que o bortezomib e o PI-083 não se encontram posicionados corretamente no local ativo quando a molécula de água está presente, ficando estes inibidores bastante longe do local pretendido (figura 12). De facto, mesmo o carfilzomib cuja pose mais favorável se encontra no local ativo parece não estabelecer qualquer interação tanto com a ThrI como com a água, sugerindo que não é favorável a presença desta nos cálculos (figura 13). No caso do marizomib, constatou-se nas

interações deste com o recetor que ficou posicionado próximo da Thr1, estando a molécula de água entre o ligando e a Thr1.



**Figura 12** – Docking do bortezomib na Thr1 no modelo de homologia na presença de uma molécula de água (imagem da esquerda) e na ausência de moléculas de água (imagem da direita).  
Legenda de cores: bortezomib – rosa, Thr1 – verde, água – círculo preto.



**Figura 13** – Docking do carfilzomib na Thr1 no modelo de homologia na presença de uma molécula de água (imagem da esquerda) e na ausência de moléculas de água (imagem da direita).  
Legenda de cores: carfilzomib – rosa, Thr1 – verde, água – círculo preto.

Tendo em conta que quando se incluiu água no modelo de homologia do *pocket*  $\beta 5c/\beta 6c$  nem todos os inibidores foram posicionados corretamente no local ativo e que, no caso dos inibidores que se conseguiram posicionar corretamente, os scores obtidos na presença e ausência de água são muito semelhantes e somente o marizomib interagiu com a água próximo da Thr1, optámos por não incluir água no modelo de homologia para os cálculos futuros.

Para além do bortezomib, carfilzomib, marizomib e PI-083 que foram *dockados* para decidir se seria ou não considerada a presença de moléculas de água junto à Thr1 do modelo de homologia do local ativo CT-L, foram ainda *dockados* neste modelo outros inibidores conhecidos da subunidade catalítica  $\beta 5c$  na escala do nM (CEP-18770, fellutamide B, MG-132, NLVS, ritonavir, TMC-95A) tendo as condições do *docking* efetuado sido as mesmas que foram referidas para os primeiros inibidores mencionados. Verificou-se que as interações que se estabeleciam entre os inibidores conhecidos e o proteassoma eram as esperadas e que os *scores* obtidos demonstravam a capacidade inibitória destes compostos. No anexo 3 encontram-se os *scores* para cada um dos inibidores *dockados*.

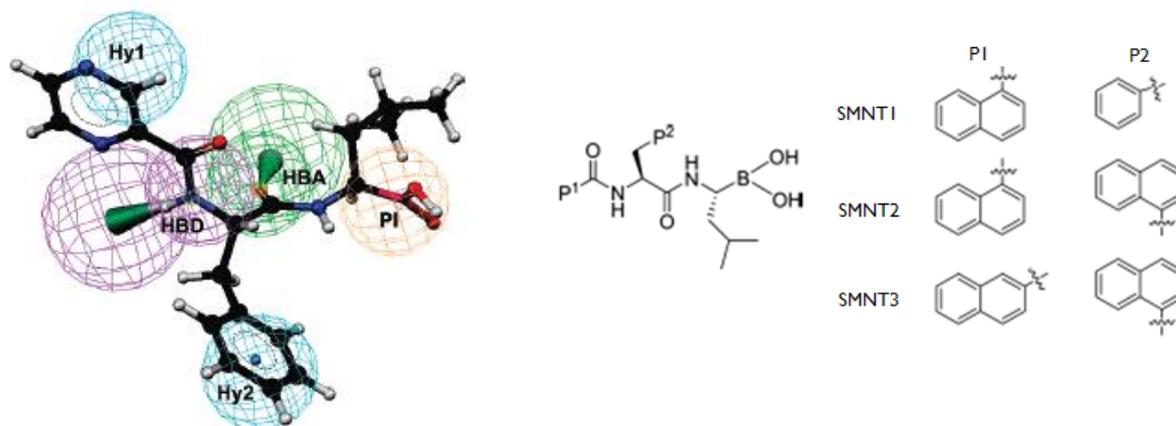
Apesar de o nosso estudo incidir na atividade catalítica da subunidade  $\beta 5c$ , foram efetuados modelos de homologia das restantes subunidades catalíticas constitutivas e indutivas para utilizar em estudos futuros. No entanto, apenas apresentamos nesta secção de resultados práticos o modelo de homologia do local ativo CT-L que posteriormente foi utilizado nos cálculos de *docking*, encontrando-se os dados relativos à construção dos restantes modelos de homologia obtidos ( $\beta 1c$ ,  $\beta 2c$ ,  $\beta 1i$ ,  $\beta 2i$ ,  $\beta 5i$ ), assim como os próprios modelos, incluídos no anexo 4. Para as restantes subunidades catalíticas constitutivas ( $\beta 1c$  e  $\beta 2c$ ) e indutivas ( $\beta 1i$ ,  $\beta 2i$  e  $\beta 5i$ ), o procedimento seguido foi semelhante ao descrito para a subunidade  $\beta 5c$ .

## 4.2 – DESENVOLVIMENTO DE MODELOS FARMACOFÓRICOS TRIDIMENSIONAIS

Sendo a modelação farmacofórica um dos métodos mais rápidos e eficazes para a descoberta de um novo *scaffold*, trata-se de uma técnica muito importante no CADD. O farmacóforo é um modelo que auxilia na identificação de moléculas que possuem características semelhantes ao farmacóforo, revelando-se bastante importante na procura de potenciais compostos *lead*.<sup>61</sup>

Lei *et al.*<sup>116</sup> desenvolveram previamente um modelo farmacofórico (figura 14 para inibidores do proteassoma tendo como base inibidores dipeptídicos com átomos de boro na sua constituição. Para tal, utilizaram o *software* Catalyst sendo o modelo obtido constituído por cinco características: uma característica aceitadora de ligações de hidrogénio, uma característica doadora de ligações de hidrogénio, uma característica positiva ionizável e duas características hidrofóbicas. Este farmacóforo foi validado utilizando um conjunto de 26 moléculas (entre as quais se inclui o bortezomib – MG341) e permitiu identificar novos

compostos *lead*, tendo sido efetuada a síntese e a caracterização biológica de três desses compostos que demonstraram ser mais potentes que o bortezomib.



**Figura 14** – O MF desenvolvido por Lei *et al.* alinhado com o bortezomib (MG341) e as três moléculas que apresentaram  $IC_{50}$  inferior ao bortezomib.<sup>116</sup>

Para além do exemplo dos inibidores do proteassoma, alguns exemplos de estudos que recorreram à utilização de farmacóforos incluem a descoberta de inibidores não esteróides da  $5\alpha$ -redutase, inibidores da tirosina cinase, inibidores seletivos da ciclo-oxigenase 2 (COX-2) e inibidores da aldose redutase humana, entre muitos outros.<sup>65</sup>

A obtenção de farmacóforos pode ser efetuada recorrendo às seguintes metodologias: tendo como base a estrutura de ligandos, tendo como base a estrutura do local ativo das proteínas ou ainda, numa abordagem mais recente, a formação do complexo ligando-recetor devido à interação do(s) ligando(s) com o local ativo da proteína.<sup>65,86</sup>

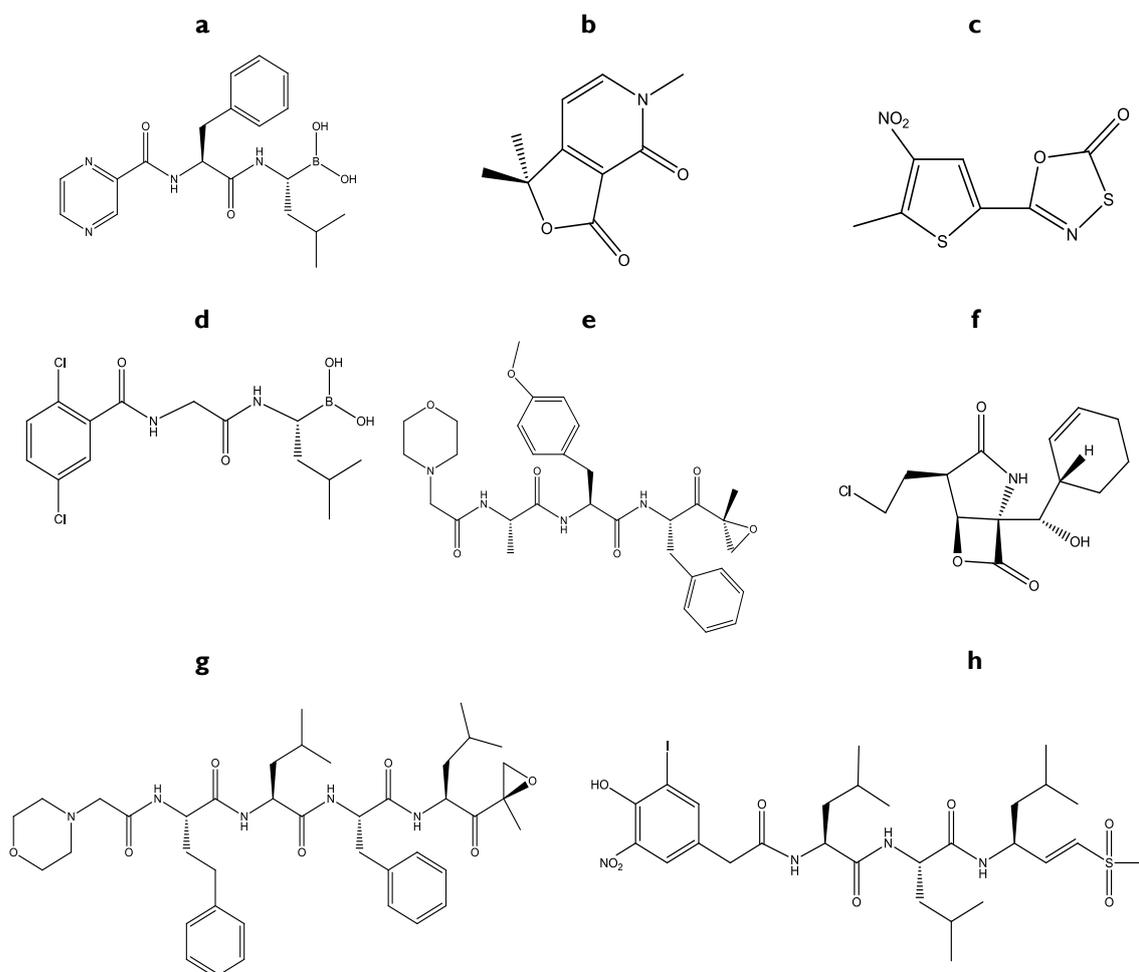
Neste trabalho, optámos por desenvolver e validar dois farmacóforos que permitem diferenciar possíveis inibidores de moléculas não ativas: um primeiro farmacóforo desenvolvido com base na estrutura de inibidores conhecidos da subunidade catalítica  $\beta 5c$  do proteassoma e um segundo farmacóforo desenvolvido com base na estrutura tridimensional do local ativo CT-L (o qual é formado pelas subunidades  $\beta 5c$  e  $\beta 6c$ ).

Uma etapa bastante importante em todo este processo passou pela criação de duas bases de dados: uma base de dados com 142 compostos com atividade inibidora comprovada na subunidade CT-L, sendo que alguns dos compostos em questão se encontram comercializados, em ensaios clínicos ou publicados – “Controlo positivo” (figura 15); e uma base de dados com 58 compostos inativos na subunidade  $\beta 5c$ , tendo estes compostos sido selecionados a partir da literatura consultada – “Controlo negativo” (figura

16). No anexo 5 encontram-se as bases de dados de “controlo positivo” e de “controlo negativo”.

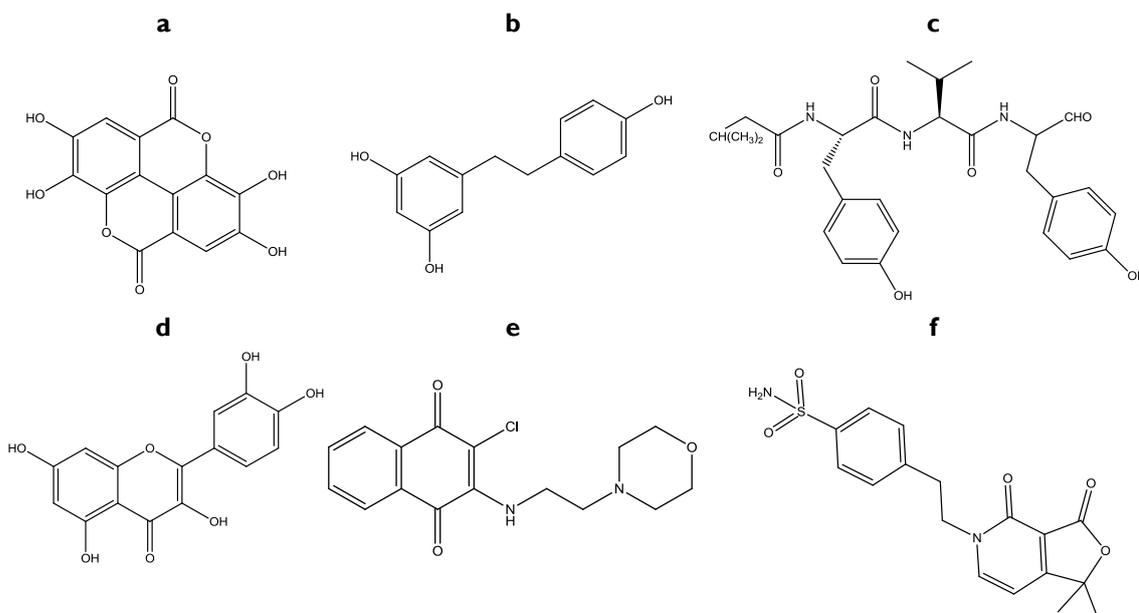
A existência destas duas bases de dados é importante para a validação dos modelos farmacofóricos. Nenhuma das moléculas presentes na base de dados do controlo negativo foi utilizada para o desenvolvimento de farmacóforos, somente foram utilizadas para refinamento e validação dos modelos farmacofóricos (MF).

As estruturas de todos os compostos foram construídas e otimizadas através do software MOE 2010.10. Começou-se por atribuir as cargas atômicas parciais e por efetuar a protonação de todas as estruturas através da funcionalidade “Protonate 3D”, tendo tal sido efetuado a pH 7.4 e à temperatura de 300K. De seguida, efetuou-se a minimização energética utilizando o campo de forças MMFF94x. Após obtenção das estruturas otimizadas, foram criadas no MOE as bases de dados anteriormente referidas.



**Figura 15** – Alguns exemplos de inibidores da subunidade  $\beta 5c$  utilizados para a construção da base de dados de “controlo positivo”.

**a** – bortezomib, **b** – cerpegina, **c** – HT1171, **d** – MLN2238, **e** – ONX0914, **f** – marizomib, **g** – carfilzomib, **h** – NLVS.



**Figura 16** – Alguns exemplos dos compostos utilizados para a construção da base de dados de “controlo negativo”.

**a** – ácido elágico, **b** – resveratrol, **c** – tiropeptina A, **d** – quercetina, **e** – composto 12<sup>117</sup>, **f** – composto 24<sup>37</sup>.

De uma forma geral, a criação de um MF ou de uma hipótese farmacofórica no MOE envolve as seguintes etapas:<sup>78</sup>

1. Preparar os dados;
2. Gerar os pontos de anotação do farmacóforo;
3. Selecionar características para incluir na hipótese farmacofórica;
4. Refinar a hipótese farmacofórica;
5. Validar a hipótese farmacofórica.

A preparação dos dados depende do tipo de farmacóforo que se pretende obter. Por exemplo, no caso de um farmacóforo efetuado com base na estrutura de ligandos é necessário gerar um conjunto de conformações para as moléculas que constituem a base de dados. No caso de um farmacóforo obtido com base na estrutura do recetor, é necessário conhecer a estrutura tridimensional do recetor de modo a localizar corretamente os *pockets* de ligação.

Podemos definir “anotação” como sendo o processo de identificar regiões de importância farmacofórica no espaço em torno de uma conformação molecular e associar tipos de características farmacofóricas a essas regiões.<sup>78</sup> De uma forma geral, os pontos de anotação podem ser divididos em quatro categorias: centrados em átomos (aceitadores de ligações de hidrogénio, Acc; doadores de ligações de hidrogénio, Don; aniões, Ani; catiões, Cat; ligando metálico, ML; átomo hidrofóbico, HydA), projetados (doador projetado, Don2;

aceitador projetado, Acc2; ligando metálico projetado, ML2; anel normal, PiN), centróides (aromáticos, Aro; anel  $\pi$ , PiR; hidrofóbico, Hyd) e bioisósteros (NCN+, CN2; COO<sup>-</sup>, O2).<sup>78</sup>

O MOE possui diversos esquemas de anotação farmacofórica, sendo que cada um destes esquemas apresenta um determinado conjunto de tipos de anotação. Por exemplo, o esquema pré-definido é o *Unified*, inclui os tipos de anotação dos outros esquemas e sem perda de eficiência.<sup>78</sup>

Após a obtenção dos pontos de anotação, segue-se a transformação destes em características farmacofóricas com auxílio do *Pharmacophore query editor*. No entanto, não nos podemos esquecer que os pontos de anotação são apenas sugestões do tipo de características que podem ser posteriormente consideradas.<sup>78</sup>

Para refinar a hipótese farmacofórica, podem utilizar-se diversas estratégias, sendo estas selecionadas caso a caso. Assim, algumas das estratégias utilizadas nesta etapa incluem a utilização de restrições de volume (volumes de exclusão, volumes de inclusão e volumes exteriores) e de forma, a adição ou exclusão de características e volumes, a determinação de características através da análise do *pocket* (por exemplo, em termos de contactos preferenciais e de superfície eletrostática) e ainda o ajuste dos raios e posição de cada característica.<sup>78</sup>

Para validar a hipótese farmacofórica, esta pode ser avaliada através de uma procura farmacofórica utilizando esta funcionalidade do MOE (*Pharmacophore search*). Assim, pode correr-se o MF nas bases de dados de compostos ativos e inativos de modo a verificar a percentagem de compostos de cada uma das bases de dados que é detetada.<sup>78</sup>

#### **4.2.1 – DESENVOLVIMENTO DE MODELOS FARMACOFÓRICOS TRIDIMENSIONAIS COM BASE NA ESTRUTURA DE INIBIDORES DA SUBUNIDADE CATALÍTICA $\beta$ 5C DO PROTEASSOMA**

Na modelação farmacofórica com base em ligandos foram inicialmente desenvolvidos cinco modelos através do algoritmo “elucidação farmacofórica” (disponível no *software* MOE 2012.10) de modo a que, posteriormente, se refinasse e otimizasse o melhor farmacóforo.

Assim, cada um dos cinco modelos inicialmente desenvolvidos teve como base uma de cinco classes de inibidores do proteassoma que atuam na subunidade catalítica em questão – aldeídos, boronatos,  $\beta$ -lactonas, epoxicetonas e vinilsulfonas. Os compostos

utilizados na obtenção dos MF de cada uma destas classes podem ser consultados na tabela

4.

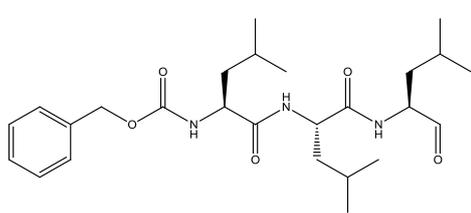
---

**Compostos de cada uma das classes de inibidores utilizadas para obter os modelos farmacofóricos iniciais**

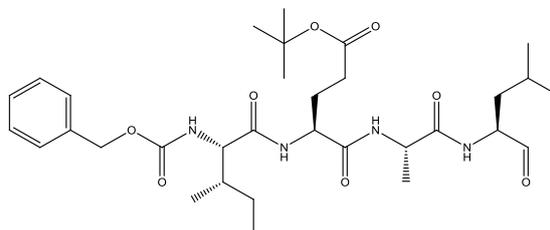
---

**Aldeídos**

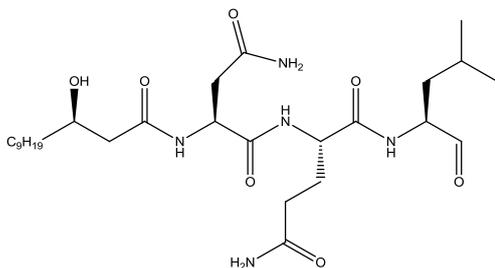
---



**MG-132**



**PSI**

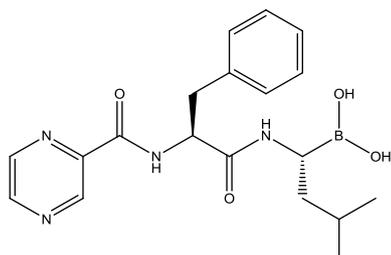


**Felutamida B**

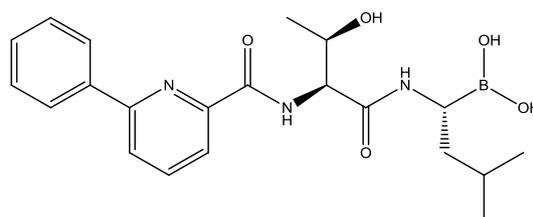
---

**Boronatos**

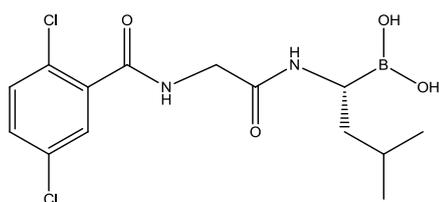
---



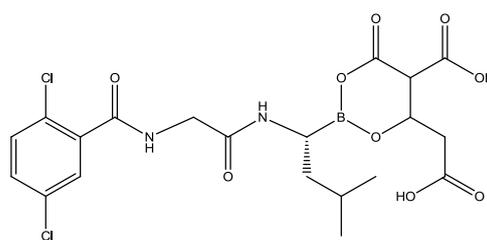
**Bortezomib**



**CEP-18770**



**MLN-2238**

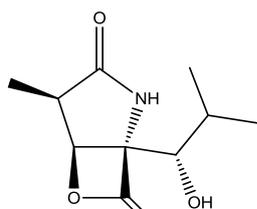


**MLN-9708**

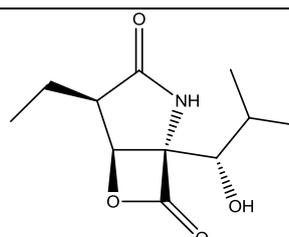
---

**$\beta$ -lactonas**

---

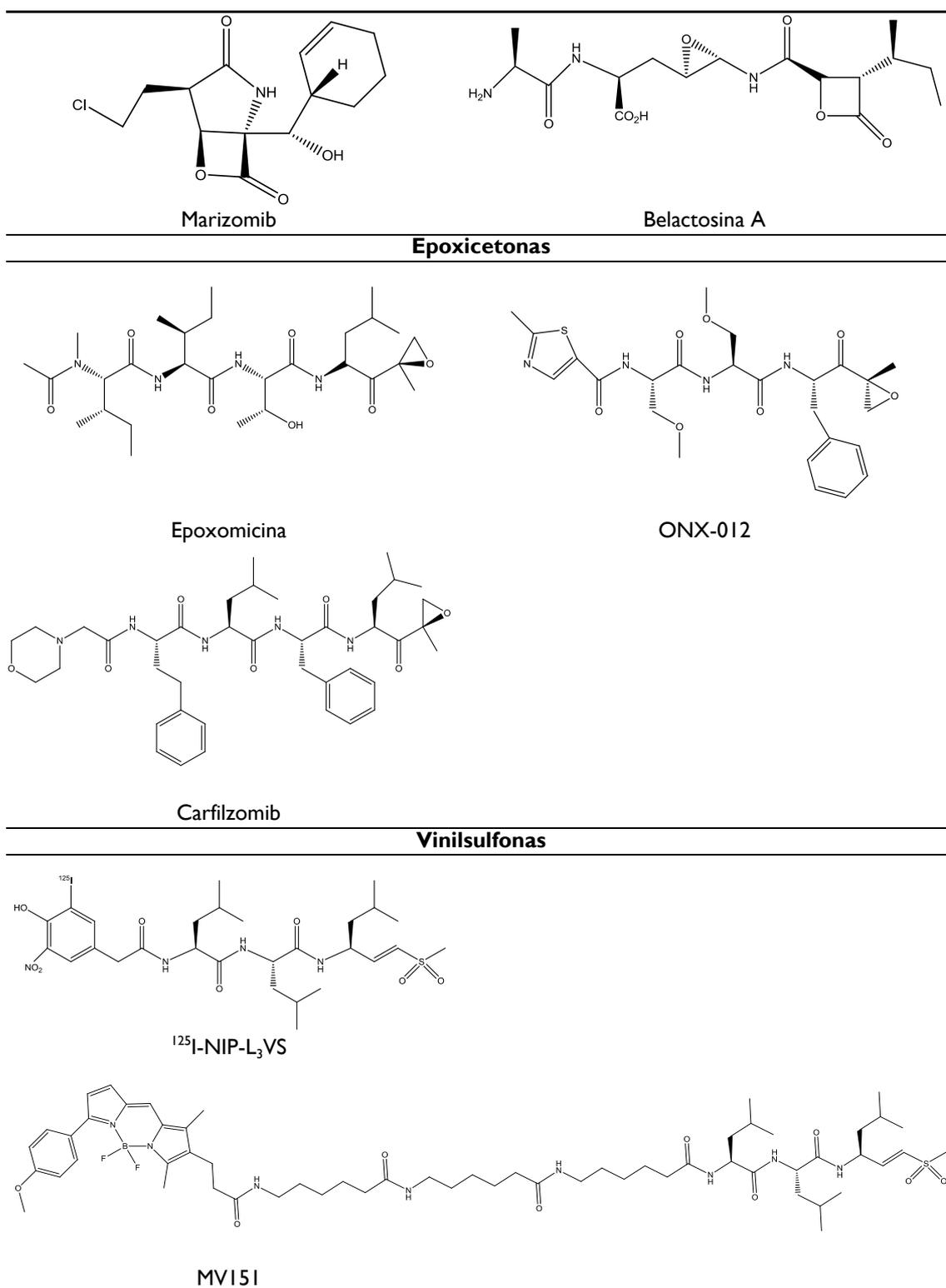


**Omuralide**



**PS-519**

---



**Tabela 4** – Os compostos de cada uma das classes de inibidores utilizadas para obter os modelos farmacofóricos iniciais.

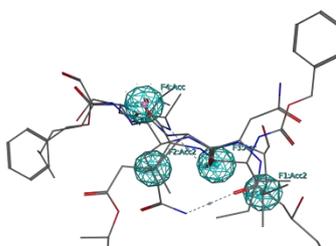
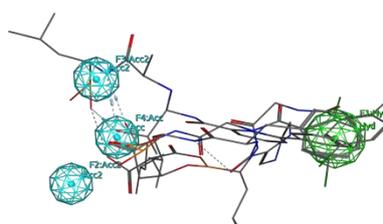
Os MF obtidos por elucidação farmacofórica foram efetuados de duas formas: dando ênfase no alinhamento das moléculas às características aromáticas (Ar)/aceitadoras (Acc)/doadoras (Don) e não dando ênfase no alinhamento a estas características. Estes MF

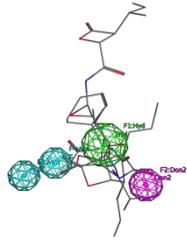
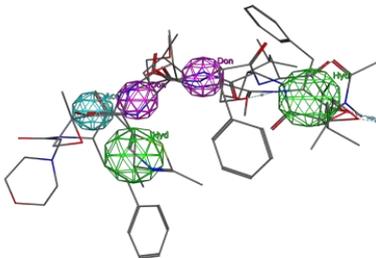
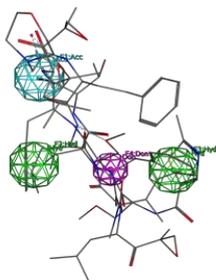
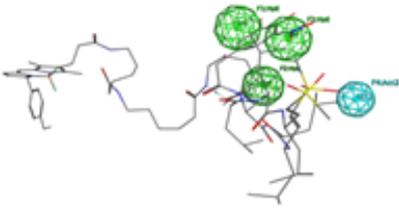
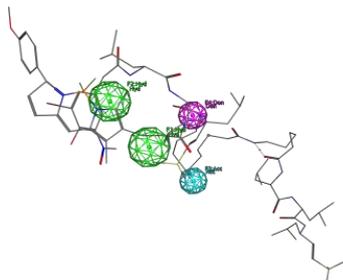
foram obtidos utilizando o esquema *Unified*, incluindo este esquema os seguintes tipos de anotação: anião (Ani), catião (Cat), CO2 centróide (O2), centro NCN+ (CN2), doador de ligações de hidrogénio (Don), doador de ligações de hidrogénio projetado (Don2), aceitador de ligações de hidrogénio (Acc), aceitador de ligações de hidrogénio projetado (Acc2), centro aromático (Aro), centro de anel  $\pi$  (PiR), anel aromático ou  $\pi$  normal (PiN), átomo hidrofóbico (Hyd).

Os resultados obtidos para cada um dos cinco MF iniciais encontram-se agrupados na tabela 5.

Antes de prosseguir para outras etapas, começou-se por averiguar se cada MF conseguia detetar as moléculas que lhes deram origem, tendo tal sido efetuado no MOE 2012.10 (*“Pharmacophore search: search database”*). Verificou-se que cada um dos MF conseguia detetar todas as moléculas que lhe deram origem.

A etapa seguinte consistiu em correr cada MF nas bases de dados dos controlos positivos e negativos para registar quantos compostos eram detetados em cada uma das bases de dados.

Base de dados	Características	Raio (Å)	Modelo farmacofórico e alinhamento molecular	Número de hits		
				Controlo Positivo	Negativo	
<b>Aldeídos</b> (a, b, d) (a, c, d)	2 Acc (F3, F4)	F1: 1		90 em 142 (63%)	24 em 58 (41%)	
	2 Acc2 (F1, F2)	F2: 1 F3: 1 F4: 1				
<b>Boronatos</b> (a, b, d) (a, c, d)	1 Acc (F4)	F1: 1.4			101 em 142 (71%)	31 em 58 (53%)
2 Acc2 (F2, F3)	F2: 1 F3: 1 F4: 1					
1 Hyd (F1)						

<b>Beta-lactonas</b> (a, b, d)	1 Acc (F4) 1 Acc2 (F3) 1 Don2 (F2) 1 Hyd (F1)	F1: 1.4 F2: 1 F3: 1 F4: 1		112 em 142 (79%)	41 em 58 (69%)
<b>Epoxicetonas</b> (a, b, d)	1 Acc (F3) 2 Don (F4, F5) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1 F5: 1		57 em 142 (42%)	6 em 58 (10%)
<b>Epoxicetonas</b> (a, c, d)	1 Acc (F1) 1 Don (F4) 2 Hyd (F2, F3)	F1: 1.4 F2: 1.4 F3: 1.4 F4: 1		98 em 142 (69%)	8 em 58 (14%)
<b>Vinilsulfonas</b> (a, b, d)	1 Acc2 (F4) 3 Hyd (F1, F2, F3)	F1: 1.4 F2: 1.4 F3: 1.4 F4: 1		95 em 142 (67%)	4 em 58 (7%)
<b>Vinilsulfonas</b> (a, c, d)	1 Acc (F3) 1 Don (F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1		74 em 142 (52%)	5 em 58 (8%)

a) Alinhamento obtido por elucidação farmacofórica; b) Alinhamento sem ênfase nas características Aro/Acc/Don; c) Alinhamento com ênfase nas características Aro/Acc/Don; d) Esquema *Unified*.

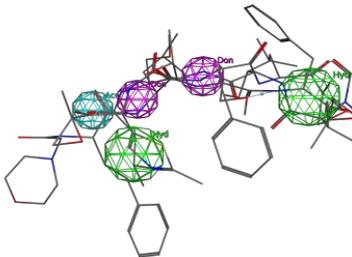
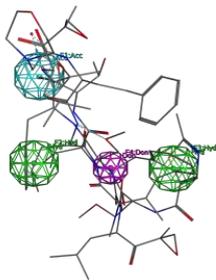
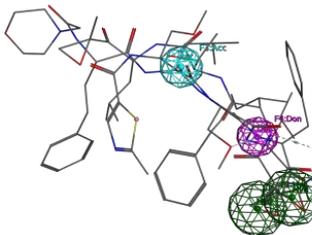
F – Característica farmacofórica

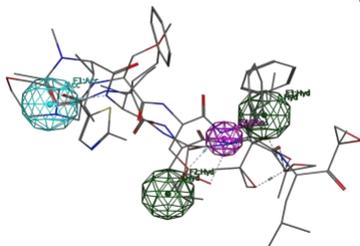
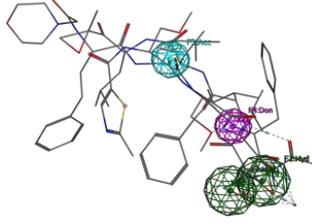
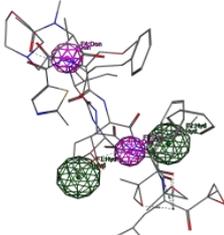
**Tabela 5** – Os MF iniciais obtidos para cada uma das cinco classes utilizadas de inibidores do proteassoma.

Como podemos constatar, os dois melhores MF obtidos são referentes às epoxicetonas (com ênfase nas características Ar/Acc/Don) e às vinilsulfonas (sem ênfase nas características Ar/Acc/Don), verificando-se que nestes dois casos a percentagem de *hits*

detetados no controlo positivo é bastante superior à percentagem de falsos positivos detetada no controlo negativo.

Adicionalmente, com o objetivo de aprofundar o estudo prévio dos melhores MF até ao momento, foram efetuados MF para as epoxicetonas e para as vinilsulfonas utilizando o esquema PCHD [que inclui as anotações Acc, Don, Cat, Ani, Aro, Hyd, ML, Don2, Acc2, centro projetado de anel aromático (Aro2) e ponto projetado de ligando metálico (ML2)] e PCH [que inclui as anotações Acc, Don, Cat, Ani, Aro, ML e V (*Volume constraint*)]. Não foi possível utilizar os esquemas CHD e PPCH porque estes não conseguiram gerar qualquer MF. Nas tabelas 6 e 7 encontram-se, respetivamente, os MF obtidos para as epoxicetonas e para as vinilsulfonas utilizando diferentes esquemas e antes de passarmos ao seu refinamento.

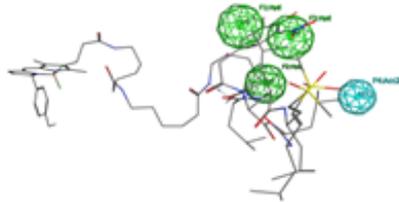
Base de dados	Características	Raio (Å)	Modelo farmacofórico e alinhamento molecular	Número de hits	
				Controlo Positivo	Controlo Negativo
<b>Epoxicetonas (a, b, d)</b>	1 Acc (F3) 2 Don (F4, F5) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1 F5: 1		57 em 142 (42%)	6 em 58 (10%)
<b>Epoxicetonas (a, c, d)</b>	1 Acc (F1) 1 Don (F4) 2 Hyd (F2, F3)	F1: 1.4 F2: 1.4 F3: 1.4 F4: 1		98 em 142 (69%)	8 em 58 (14%)
<b>Epoxicetonas (a, b, e)</b>	1 Acc (F3) 1 Don (F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1		89 em 142 (63%)	15 em 58 (25%)

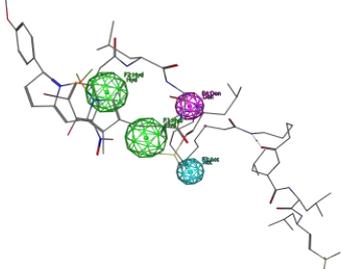
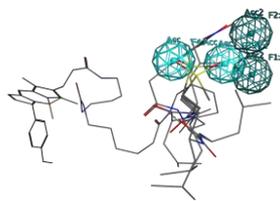
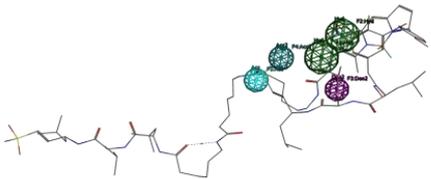
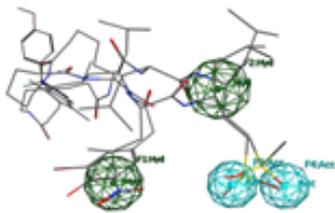
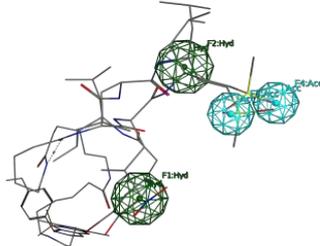
<b>Epoxicetonas (a, c, e)</b>	1 Acc (F1) 1 Don (F4) 2 Hyd (F2, F3)	F1: 1.4 F2: 1 F3: 1.4 F4: 1.4		47 em 142 (33%)	7 em 58 (12%)
<b>Epoxicetonas (a, b, f)</b>	1 Acc (F3) 1 Don (F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1		89 em 142 (63%)	15 em 58 (25%)
<b>Epoxicetonas (a, c, f)</b>	2 Don (F3, F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1		32 em 142 (23%)	4 em 58 (7%)

a) Alinhamento obtido por elucidação farmacofórica; b) Alinhamento sem ênfase nas características Aro/Acc/Don; c) Alinhamento com ênfase nas características Aro/Acc/Don; d) Esquema *Unified*; e) Esquema PCHD, f) esquema PCH.

F – Característica farmacofórica

**Tabela 6** – Modelos farmacofóricos obtidos para as epoxicetonas antes de ser efetuado o seu refinamento.

Base de dados	Características	Raio (Å)	Modelo farmacofórico e alinhamento molecular	Número de hits	
				Positivo	Negativo
<b>Vinilsulfonas (a, b, d)</b>	3 Hyd (F1, F2, F3) 1 Acc2 (F4)	F1: 1.4 F2: 1.4 F3: 1.4 F4: 1		95 em 142 (67%)	4 em 58 (7%)

<b>Vinilsulfonas (a, c, d)</b>	1 Acc (F3) 1 Don (F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1	74 em 142 (52%)	5 em 58 (8%)	
<b>Vinilsulfonas (a, b, e)</b>	2 Acc (F1, F2) 2 Acc2 (F3, F4)	F1: 1 F2: 1 F3: 1 F4: 1	102 em 142 (72%)	42 em 58 (71%)	
<b>Vinilsulfonas (a, c, e)</b>	1 Acc (F5) 1 Acc2 (F4) 1 Don2 (F3) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1 F5: 1	16 em 142 (11%)	2 em 58 (3%)	
<b>Vinilsulfonas (a, b, f)</b>	2 Acc (F3, F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1	50 em 142 (35%)	2 em 58 (3%)	
<b>Vinilsulfonas (a, b, f)</b>	2 Acc (F3, F4) 2 Hyd (F1, F2)	F1: 1.4 F2: 1.4 F3: 1 F4: 1	50 em 142 (35%)	2 em 58 (35%)	

a) Alinhamento obtido por elucidação farmacofórica; b) Alinhamento sem ênfase nas características Aro/Acc/Don; c) Alinhamento com ênfase nas características Aro/Acc/Don; d) Esquema *Unified*; e) Esquema PCHD, f) esquema PCH.

F – Característica farmacofórica.

**Tabela 7** – Modelos farmacofóricos obtidos para as vinilsulfonas antes de ser efetuado o seu refinamento.

De seguida, efetuou-se o refinamento dos dois melhores farmacóforos:

- **O MF referente às epoxicetonas**

- Obtido dando ênfase às características Ar/Acc/Don e utilizando o esquema *Unified*;
- Com as características F1 Acc (raio 1.4 Å), F2 Hyd (raio 1.4 Å), F3 Hyd (raio 1.4 Å) F4 Don (raio 1 Å).

- **O MF relativo às vinilsulfonas**

- Obtido sem dar ênfase às características Ar/Acc/Don e usando o esquema *Unified*;
- Com as características F1 Hyd (raio 1.4 Å), F2 Hyd (raio 1.4 Å), F3 Hyd (raio 1.4 Å) e F4 Acc2 (raio 1 Å).

Com o propósito de otimizar o MF obtido, foram sendo efetuadas alterações a cada característica, quer em termos de raio (aumento e diminuição) de modo a tentar definir o raio óptimo (cujo objetivo é identificar o máximo de moléculas ativas e rejeitar o máximo de moléculas não ativas), quer em termos de seleção das características a considerar essenciais, quer em termos de alteração das características presentes (transformar as características Hyd em Acc, Don ou Aro; transformar as características Acc em Don e vice-versa). No anexo 6 encontra-se um resumo das alterações efetuadas (mantendo as restantes características iniciais) e dos respetivos resultados na deteção de compostos das bases de dados do controlo positivo e negativo.

Tendo em conta os resultados obtidos com as alterações efetuadas aos MF iniciais, optou-se por selecionar o MF das epoxicetonas como a nossa base para a obtenção do MF final. Com o modelo farmacofórico preliminar MFL1 [F1 Acc (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Acc (raio 1)] no qual apenas foi alterada a característica F4 de Don para Acc, foram detetados 102 compostos do controlo positivo (72%) e 7 compostos do controlo negativo (12%), o que demonstra uma boa seletividade deste modelo.

Com o objetivo de otimizar o MF das epoxicetonas, efetuou-se em simultâneo um conjunto de modificações dos parâmetros de cada característica farmacofórica. Construiu-se então o seguinte MF tendo como base o MF inicial [F1 Acc (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Don (raio 1)]: F1 Acc (raio 1.6), F2 Hyd (raio 1.7), F3 Hyd (raio 1.4), F4 Acc (raio 1) – MFL2. Em termos do controlo positivo e negativo, foram detetados 110 *hits* ativos e 17 falsos *hits* (uma razão de 77%/29%).

No mesmo MF inicial das epoxicetonas, efetuaram-se as seguintes alterações: F1 Acc (raio 1.6), F2 Hyd (raio 1.7), F3 Hyd (raio 1.6), F4 Acc (raio 1). Com este MF foram detetados 114 *hits* ativos e 19 falsos *hits* (uma razão de 80%/32%). No entanto, este ligeiro

aumento na percentagem de compostos detetados no controlo positivo não compensa o aumento verificado também nos compostos do controlo negativo. Como tal, o aumento do raio em F3 de 1.4 para 1.6 não é assim tão vantajoso pois levaria a perda de seletividade.

Após verificação nas bases de dados de *output* dos controlos positivos de quais as moléculas que não foram selecionadas pelos MFL1 e MFL2, foi efetuado um MF com algumas das moléculas mais representativas dos compostos ativos não detetados( bortezomib, CVT-659, HT1171, MLN2238, marizomib, PDTC e *spirolactacystin*), sendo o objetivo deste MF identificar características dos compostos que pudessem ser adicionadas aos MF. Como resultado deste MF, foram detetados 106 *hits* ativos (75%) e 38 falsos positivos (64%). Dada a elevada percentagem detetada no controlo negativo, este MF foi excluído.

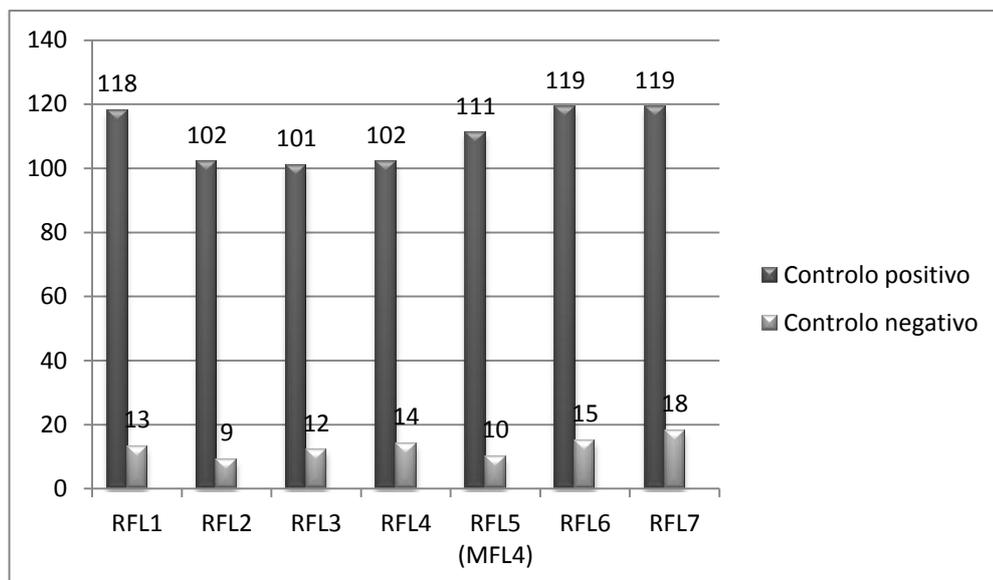
Ao aumentar o raio da característica F4 do MFL1 de 1 para 1.1 (F4 Acc, raio 1.1), o bortezomib foi de imediato detetado. Este MF (MFL3) mostrou-se bastante eficiente conseguindo identificar nos controlos positivo e negativo os valores de 101 (71%) e 9 (15%), respetivamente.

No MFL3, efetuou-se ainda uma alteração ao nível das características essenciais. Assim, considerando as características F1, F2 e F3 como essenciais, os resultados foram os seguintes: 132 *hits* ativos e 52 falsos *hits* (93%/88%). A percentagem do controlo positivo é muito boa; no entanto, a percentagem obtida no controlo negativo é demasiado elevada, o que invalida este MF.

As últimas alterações efetuadas ao MFL3 tiveram como objetivo a introdução de características duplas, ou seja, características que podem ser Acc ou Don, Aro ou Hyd. Na tabela 8 encontram-se os resultados obtidos.

Refinamento final do MFL3		
Característica(s) farmacofórica(s) alterada(s) tendo como base o MFC F1 Acc (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Acc (raio 1.1)	Número de <i>hits</i>	
	Controlo positivo	Controlo negativo
RFL1 – F4 Acc Don (raio 1.1)	118 (83%)	13 (22%)
RFL2 – F2 Aro Hyd (raio 1.4)	102 (72%)	9 (15%)
RFL3 – F3 Aro Hyd (raio 1.4)	101 (71%)	12 (20%)
RFL4 – F2 e F3 Aro Hyd (raio 1.4)	102 (72%)	14 (24%)
<b>RFL5 – F1 Acc Don (raio 1.4)</b>	<b>111 (78%)</b>	<b>10 (17%)</b>
RFL6 – F1 (raio 1.4) e F4 Acc Don (raio 1.1):	119 (84%)	15 (25%)
RFL7 – F1 Acc Don (raio 1.4) e F4 Acc Don (raio 1.1), F2 e F3 Aro Hyd (raio 1.4)	119 (84%)	18 (31%)

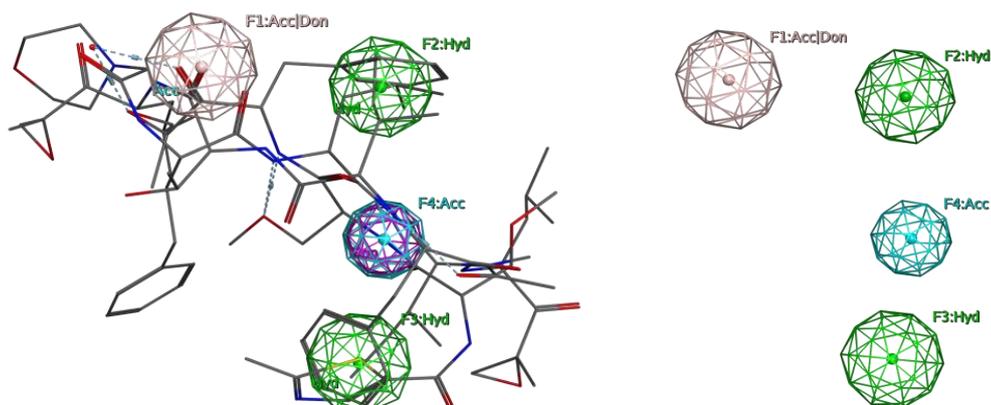
**Tabela 8** – Refinamento final do MFL3. A negrito encontra-se assinalado o melhor resultado obtido no refinamento, tendo sido assim obtido o MFL4.



**Gráfico 2** – Número de moléculas detetadas nos controlos positivo e negativo durante o refinamento do MFL3.

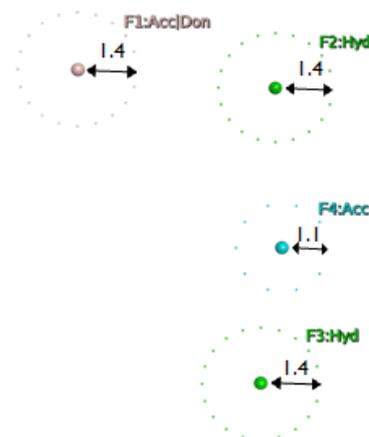
Tendo em conta os resultados obtidos no refinamento final do MFL3, optou-se por seleccionar como MF final com base no ligando o modelo MFL4 que, tendo como base o MF obtido a partir de epoxicetonas com ênfase em características Aro/Acc/Don (esquema *Unified*), apresenta as seguintes características farmacofóricas: F1 Acc|Don (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Acc (raio 1.1). Este MF apresenta uma boa razão de compostos *hit* detetados no controlo positivo e negativo (78%/17%), o que efetiva a sua validação.

Na figura 17, encontra-se o MF final obtido (MFL4) com base na estrutura de inibidores conhecidos da subunidade com atividade  $\beta$ 5c do proteassoma.



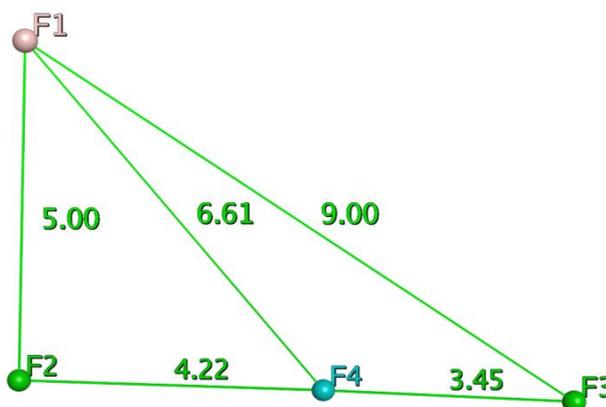
**Figura 17** – O modelo farmacofórico tridimensional efetuado com base na estrutura de inibidores da subunidade  $\beta$ 5c do proteassoma (epoxicetonas) obtido após optimização e validação, MFL4. Na imagem da esquerda o MF apresenta o alinhamento molecular das estruturas que lhe deram origem; na imagem da direita, encontra-se o MF tridimensional isoladamente. F1 Acc|Don (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Acc (raio 1.1)

Em termos dos raios de cada uma das características presentes no MFL4 (figura 18, podemos constatar que tanto a característica Acc|Don (F1) como as características Hyd (F2 e F3) apresentam o mesmo raio (1.4 Å). Por sua vez, a característica Acc (F4) apresenta um raio de 1.1 Å, o que a torna mais seletiva que as restantes características.

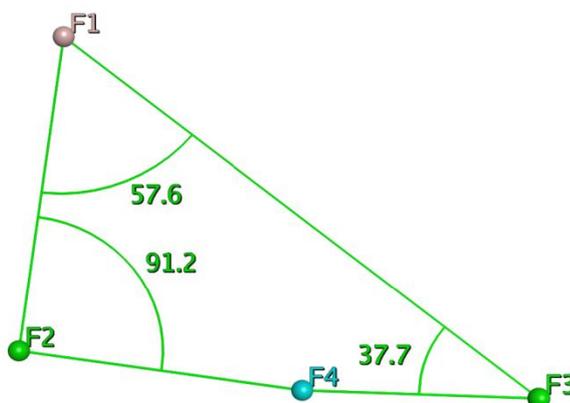


**Figura 18** – Os raios (em Å) das características farmacofóricas presentes no MFL4.

Relativamente às distâncias entre as características farmacofóricas (figura 19), verificou-se que as características mais distantes entre si são as características F1 (Acc|Don) e F3 (Hyd) que se encontram a 9.00 Å uma da outra, assim como as características Hyd F2 e F3 (distam 7.60 Å entre si).



**Figura 19** – As distâncias (em Å) entre as características farmacofóricas presentes no MFL4.



**Figura 20** – Os ângulos (em °) entre as características farmacofóricas presentes no MFL4.

Tal como foi referido na introdução teórica, os resíduos de aminoácidos relevantes para cada local ativo do proteassoma dos seres eucariotas incluem a Thr1, o Asp17, a Lys33, a Ser129, o Asp166 e a Ser169. Ou seja, são todos aminoácidos polares, pelo que interagem com grupos hidrofílicos.<sup>14</sup>

Ao nível do *pocket* de especificidade S1 da subunidade  $\beta$ 5c dos seres eucariotas, há a salientar os aminoácidos Ala20, Val31, Met45, Ala49 e Cys52, os quais, à exceção da Cys52 que é um aminoácido polar neutro (hidrofílico), são aminoácidos hidrofóbicos.<sup>17,14,39</sup> Comparando o MFL4 com os resíduos de aminoácidos relevantes para a atividade catalítica e com os aminoácidos pertencentes ao *pocket* de especificidade S1 (desempenha um papel importante no posicionamento dos ligandos), podemos constatar que o facto de termos duas características hidrofóbicas (F2 e F3) no MFL4 é importante para selecionar ligandos que possam ser bem posicionados no *pocket*; para além disso, a existência das características Acc|Don (F1) e Acc (F4) possibilitam a interação dos ligandos selecionados com a Cys52 do *pocket* de especificidade S1, assim como a interação com os aminoácidos relevantes para a atividade catalítica (a Thr1, o Asp17, a Lys33, a Ser129, o Asp166 e a Ser169) dado que são resíduos hidrofílicos.

Tendo em linha de conta a correlação existente entre as características do MFL4 e os resíduos de aminoácidos catalíticos e do *pocket* S1, assim como a elevada percentagem de compostos *hit* (78%) detetada na base de dados de compostos ativos (controlo positivo) relativamente à percentagem de compostos *hit* (17%) detetada na base de dados de compostos inativos (o controlo negativo corresponde a falsos positivos) que permitiu validar o MF em questão, podemos concluir que este farmacóforo é representativo das características gerais que uma molécula deve apresentar para ser um potencial inibidor da subunidade catalítica  $\beta$ 5c do proteassoma, podendo ser utilizado para efetuar o VS de bases de dados de modo a filtrar e diminuir o número de possíveis moléculas que possam apresentar a atividade farmacológica pretendida.

Em termos dos compostos do controlo positivo que são detetados, há a destacar os seguintes exemplos: belactosina A, bortezomib, CEP-18770, CVT-659, carfilzomib, epoxomicina, felutamida B, glidobactina A, homobelactosina, ONX0912, ONX0914, omuralide, PI-083, PR-957, ritonavir, TMC95A.

Ao nível do controlo negativo, são exemplo dos compostos detetados a luteolina, a tiropeptina A, o composto TPI01 de Lei *et al.*<sup>76</sup> e o composto 12 de Pham *et al.*<sup>37</sup>.

#### 4.2.2 – DESENVOLVIMENTO DE MODELOS FARMACOFÓRICOS TRIDIMENSIONAIS COM BASE NA ESTRUTURA TRIDIMENSIONAL DA SUBUNIDADE CATALÍTICA $\beta$ 5C DO PROTEASSOMA

Para que seja possível desenvolver modelos farmacofóricos tendo em conta a estrutura tridimensional do local ativo do recetor é necessário, antes de mais, conhecer muito bem a estrutura da proteína, nomeadamente ao nível dos aminoácidos importantes para a atividade catalítica propriamente dita, assim como os aminoácidos que formam os diversos *pockets* (nomeadamente os *pockets* de especificidade).

Em termos computacionais, existem diversos métodos que permitem identificar potenciais locais ativos de um recetor através das coordenadas tridimensionais deste. Alguns métodos recorrem à utilização das energias de interação entre o recetor e diferentes ligandos de modo a tentar localizar locais energeticamente favoráveis. No entanto, esta abordagem pressupõe a atribuição da localização dos prótons e das cargas parciais para os átomos do recetor, o que nem sempre é fácil. Para além do que foi mencionado, temos de ter em conta as energias de van der Waals (que permitem determinar regiões disponíveis), o que torna a interpretação dos níveis energéticos difícil devido à natureza de longo alcance dos potenciais eletrostáticos. Existem ainda os métodos geométricos que tentam localizar os *pockets* sem ser necessário recorrer a modelos energéticos. Esta abordagem torna-se vantajosa na medida em que não é necessário localizar os prótons.<sup>118</sup>

Para efetuar o MF com base na estrutura tridimensional do local ativo  $\beta$ 5c do proteassoma foi utilizada a funcionalidade *Site Finder* do *software* MOE 2012.10, inserindo-se esta na categoria dos métodos geométricos mencionados dado que não são utilizados modelos energéticos. Em alternativa, são tidas em conta as posições relativas e a acessibilidade dos átomos do recetor, assim como uma classificação grosseira do tipo químico.<sup>118</sup>

A metodologia em questão baseia-se nos seguintes pontos<sup>118</sup>:

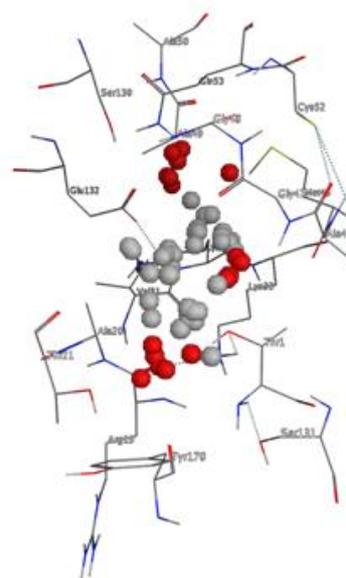
1. Identificação de regiões que apresentam densidade eletrónica elevada (sendo que tal não é o mesmo que localizar *pockets* dado que locais superficiais podem ser considerados como regiões de elevada densidade);
2. Eliminação de locais que estejam demasiado expostos ao solvente dado que têm menor probabilidade de serem bons locais ativos;
3. Classificação química dos locais em hidrofóbicos ou hidrofílicos (tal implica, por exemplo, que o conceito de hidrofílico não varie consoante o estado de protonação

ou o estado tautomérico, não existindo distinção entre átomos doadores e aceptadores).

A funcionalidade *Site Finder* utiliza um algoritmo denominado “*alpha shape construction*”, o qual permite identificar pontos na superfície de uma proteína. A metodologia em questão baseia-se na utilização de esferas alfa, sendo estas um tipo especial de esferas que estão em contacto com quatro átomos no seu limite e não contêm átomos internos. Um aspeto a salientar é o facto de cada esfera alfa ser única, ou seja, não existe uma outra esfera alfa que contacte com os mesmos quatro átomos. Para além disso, cada esfera alfa possui um centro alfa, podendo estes centros ser uma forma mais simplificada de representar as esferas em questão.<sup>118</sup>

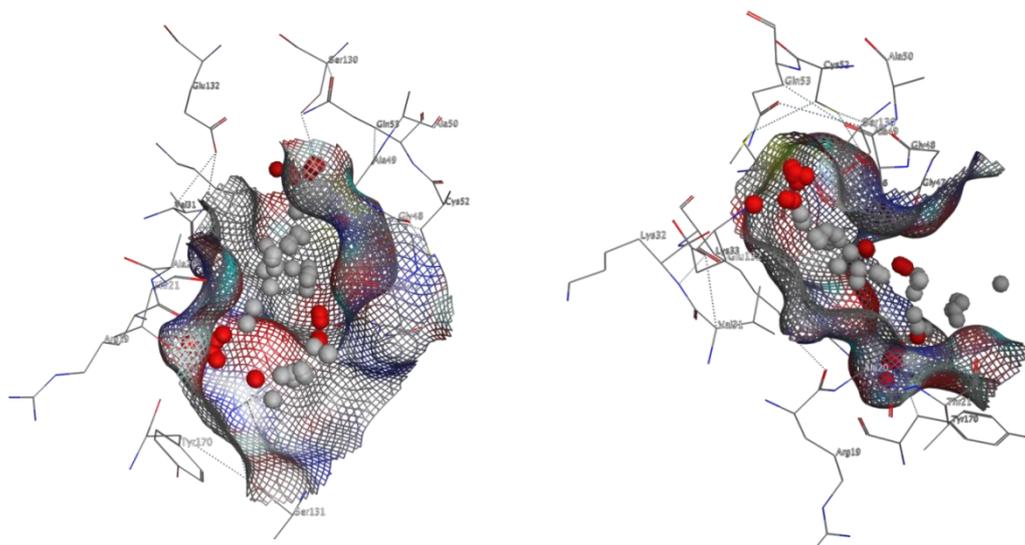
A estrutura cristalográfica utilizada para obter o farmacóforo com base no local ativo  $\beta 5c$  do proteassoma foi o código PDB 3UN8 (proteassoma 20S de levedura em complexo com PR-957), com uma resolução de 2.70 Å. Tendo como ponto de partida a estrutura tridimensional mencionada (mais especificamente, as cadeias K e L), utilizou-se a funcionalidade *Site Finder* do MOE 2012.10 para isolar a cavidade correspondente ao local ativo CT-L. Esta funcionalidade sugere-nos diversos locais que são potenciais *pockets*, pelo que temos de selecionar o local mais adequado de acordo com a informação que temos acerca dos resíduos de aminoácidos importantes para a atividade catalítica e de acordo com a importância dos resíduos de aminoácidos dos *pockets* de especificidade. Assim, foi selecionada a segunda cavidade sugerida (contém os aminoácidos Thr1, Arg19, Ala20, Thr21, Ala22, Gly23, Val31, Lys32, Lys33, Val34, Ile35, Glu36, Ile37, Asn38, Met45, Ala46, Gly47, Gly48, Ala49, Ala50, Cys52, Gln53, Ph354, Glu56, Thr57, Gly60, Ser61, Arg64, Ser131, Tyr170 e ainda Arg91, Gln94, Tyr123, Ser130, Tyr131, Glu132, Arg133) dado que esta inclui os aminoácidos Thr1 e Lys33 que são relevantes para a atividade catalítica, incluindo ainda os aminoácidos Ala20, Ala22, Val31, Met45, Gly48, Ala49 e Cys52 que são imprescindíveis para a formação dos *pockets* de especificidade. (figura 21).

Para construir um MF com base no recetor não basta somente utilizar a funcionalidade *Site Finder*. Para que todas as opções efetuadas em termos da seleção



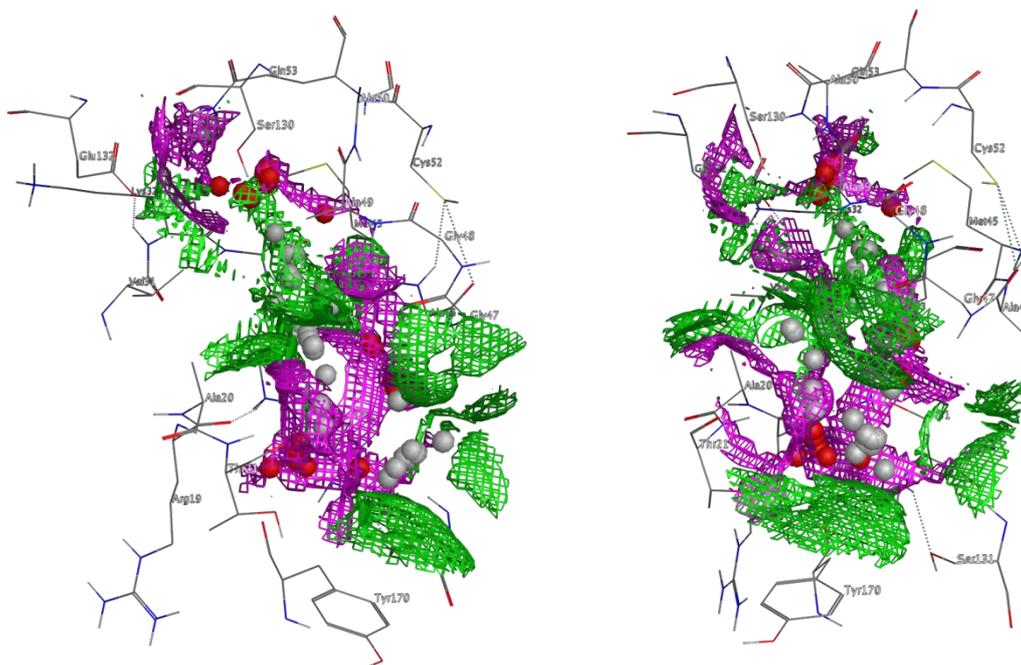
**Figura 21** – Local ativo CT-L do proteassoma identificado pela funcionalidade *Site Finder*. Os centros alfa vermelhos correspondem a pontos de contacto hidrofílicos; os centros alfa brancos são relativos a pontos de contacto hidrofóbicos.

dos centros alfa sejam devidamente fundamentadas, há que conjugar com o *Site Finder* outra funcionalidade do MOE denominada *Surfaces and Maps*. Esta funcionalidade permite visualizar diversas representações de superfícies moleculares e é útil para prever locais preferenciais para a interação com átomos do ligando. Começou-se por traçar a superfície molecular do recetor de modo a obtermos informação acerca dos resíduos que constituem o local ativo em questão (figura 22).



**Figura 22** – Superfície molecular do recetor.

Outra possibilidade de utilização da funcionalidade *Surfaces and Maps* é a determinação do mapa de preferências de contacto, o qual permite identificar as áreas preferenciais para interações hidrofóbicas ou hidrofílicas com um ligando (figura 23).



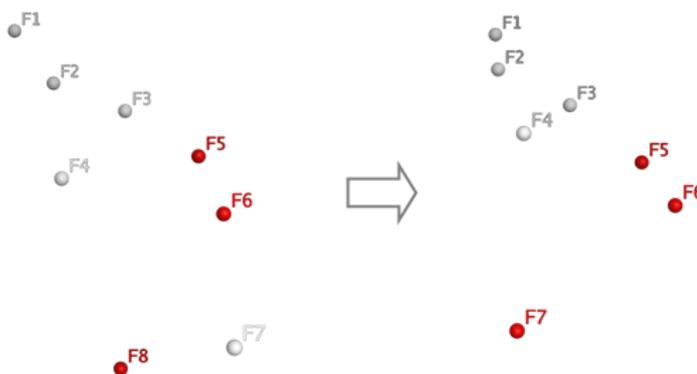
**Figura 23** – Mapa de preferências de contacto.

Legenda de cores: zona rosa – hidrofílica; zona verde – hidrofóbica.

Como podemos constatar no nosso sistema, existem centros alfa hidrofílicos (vermelhos) que se encontram nitidamente inseridas na zona hidrofílica (cor-de-rosa) do mapa de preferências de contacto. Podemos verificar também que alguns dos centros alfa hidrofóbicos (brancos) estão exatamente localizados em locais de contactos preferencialmente hidrofóbicos (verdes).

Toda a informação reunida através do *Site Finder*, do mapa da superfície molecular e do mapa de preferências de contacto, assim como todos os dados que possuímos das résiduos de aminoácidos relevantes para a atividade CT-L permite-nos refinar os centros alfa existentes de modo a obter um farmacóforo representativo para a subunidade  $\beta 5c$  do proteassoma. Para além disso, permite-nos ainda observar o tipo de interações estabelecidas entre os aminoácidos do local ativo em questão e alguns ligandos, nomeadamente o ligando da estrutura cristalográfica, o bortezomib e o carfilzomib (tendo os dados destes dois últimos ligandos sido obtidos a partir de simulações de *docking* molecular).

O primeiro passo de refinamento dos 53 centros alfa iniciais consistiu em passar estes para “*dummy atoms*” (ficam automaticamente assinalados com a letra F e o respetivo número) de modo a podermos apagar os átomos que se encontram mais afastados do local ativo ou de modo a que seja possível agrupar os átomos com características semelhantes que se encontravam muito próximos. Assim, obtivemos um conjunto final de 8 “*dummy atoms*” e, de seguida, 7 “*dummy atoms*” (figura 24).



**Figura 24** – Refinamento dos “*dummy atoms*” resultantes dos centros alfa.

Uma vez que nesta etapa já estamos a trabalhar com os “*dummy atoms*” em vez dos centros alfa propriamente ditos, torna-se possível utilizar o *Pharmacophore Elucidator*, o que permite passar os “*dummy atoms*” a características farmacofóricas. Verificámos que a disposição espacial resultante apresenta uma forma algo semelhante com o farmacóforo com

base na estrutura de ligandos obtido anteriormente. Em termos das características farmacofóricas, os átomos F1 e F2 passaram a formar a característica farmacofórica F1 dado que se encontram próximos, o átomo F3 forma a característica farmacofórica F2, o átomo F4 forma a característica farmacofórica F3 e os átomos F5 e F6 formam a característica farmacofórica F5 devido à proximidade entre estes dois átomos (figura 25). Uma vez que os centros alfa que originaram as características farmacofóricas F1, F2 e F3 estavam relacionados com características hidrofóbicas (brancas), optámos por assinalar estas características como hidrofóbicas; no caso das

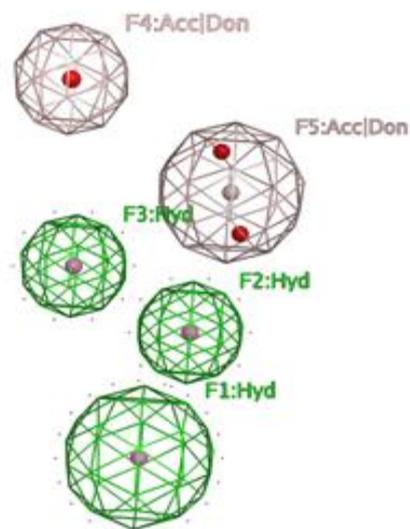
características farmacofóricas F4 e F5, estas foram assinaladas como Acc|Don devido ao facto de os centros alfa que lhes deram origem serem hidrofílicos (figura 25). Relativamente aos raios iniciais de cada característica, estes foram definidos inicialmente pela elucidação farmacofórica, sendo que F1 e F5 apresentam raios de 1.4 Å, enquanto que F1, F2 e F3 têm raios de 1 Å.

Correndo este primeiro MF nas bases de dados de controlo positivo e negativo (as mesmas utilizadas na validação dos modelos farmacofóricos obtidos com base na estrutura de ligandos), detetaram-se 61 compostos ativos (43%) e 4 compostos inativos (7%).

No entanto, dado a percentagem de compostos ativos detetados não ser tão elevada como desejávamos, efetuámos algumas alterações a este MF inicial em termos de utilização de volumes de exclusão, eventuais características essenciais, alteração dos raios de cada característica e eventual passagem de características Acc|Don para características apenas Acc ou Don de forma a torná-lo mais eficiente. Todas as alterações efetuadas e respetivos resultados podem ser consultados no anexo 7.

Em termos das conclusões a retirar dos resultados obtidos no refinamento dos MF, podemos salientar os seguintes pontos:

- A utilização de volumes de exclusão (com o objetivo de representar a estrutura do local ativo e de restringir as moléculas consideradas como *hits* ativos) não se revelou vantajosa na medida em que, apesar da percentagem de compostos do controlo



**Figura 25** – O primeiro MF obtido com base na estrutura tridimensional do recetor antes da refinamento do mesmo.

F1 Hyd (raio 1.4), F2 Hyd (raio 1), F3 Hyd (raio 1), F4 Acc|Don (raio 1), F5 Acc|Don (raio 1.4).

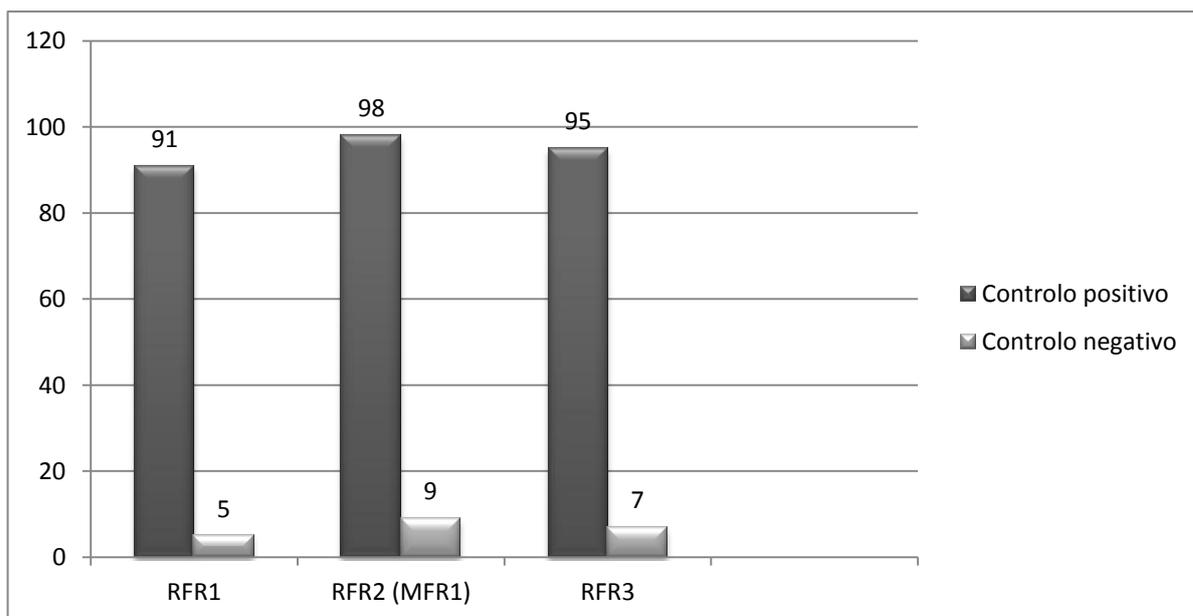
positivo aumentar (passou para 77%), a percentagem de compostos do controlo negativo aumentou bastante (passou para 88%).

- Temos de manter todas as características farmacofóricas como essenciais dado que, se apenas algumas características forem consideradas essenciais, a percentagem de compostos ativos e inativos detetada é elevada em ambas os controlos (há casos em que o MF deteta 100% dos compostos das duas bases de dados).
- Alterando progressivamente os raios de cada característica e conjugando os melhores resultados obtidos num novo MF, detetou-se 69% dos compostos do controlo positivo e 15% dos compostos do controlo negativo. Note-se que algumas características (F1, F2 e F3) apenas puderam ser aumentadas em 0.1 Å dado que o espaço disponível não permitia maior amplitude nas alterações a efetuar.
- Alterando as características Acc|Don, o melhor resultado obtido verificou-se quando F4 passou a Acc (detetou 67% de compostos ativos e 12% de compostos inativos), mantendo as restantes características iguais ao MF inicial.

Na tabela 9 encontra-se um resumo dos melhores MF obtidos, assim como as respetivas características e resultados obtidos nos controlos positivo e negativo. No gráfico 3 é possível observar a comparação do número de moléculas detetadas nos controlos positivo e negativo durante o refinamento do MF com base na estrutura do recetor.

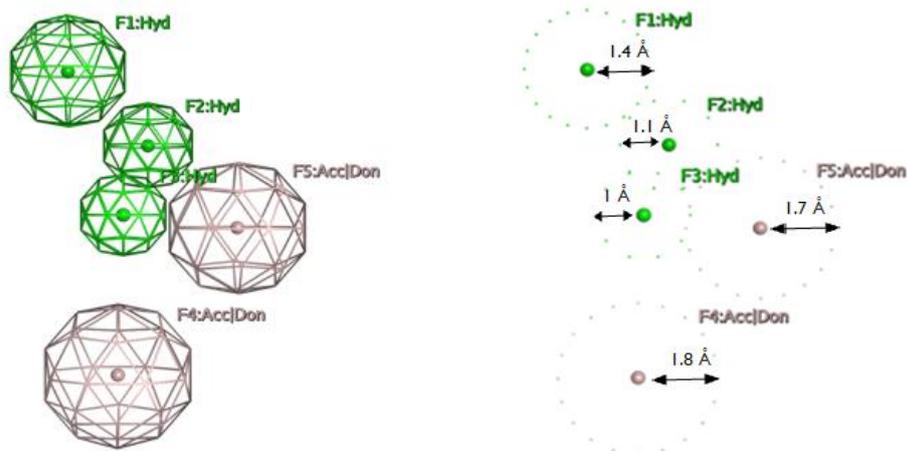
Refinamento final do modelo farmacofórico com base na estrutura do recetor		
Características farmacofóricas	Número de hits	
	Controlo positivo	Controlo negativo
RFR1 – F1 Hyd (raio 1.4), F2 Hyd (raio 1), F3 Hyd (raio 1), F4 Acc Don (raio 1.8), F5 Acc Don (raio 1.4)	91 (64%)	5 (8%)
<b>RFR2 – F1 Hyd (raio 1.4), F2 Hyd (raio 1.1), F3 Hyd (raio 1), F4 Acc Don (raio 1.8), F5 Acc Don (raio 1.7)</b> <b><u>MFR1</u></b>	<b>98 (69%)</b>	<b>9 (15%)</b>
RFR3 – F1 Hyd (raio 1.4), F2 Hyd (raio 1.1), F3 Hyd (raio 1), F4 Acc (raio 1.8), F5 Acc Don (raio 1.7)	95 (67%)	7 (12%)

**Tabela 9** – Refinamento final do modelo farmacofórico com base na estrutura do recetor. A negrito encontra-se o MF final selecionado (MFR1).

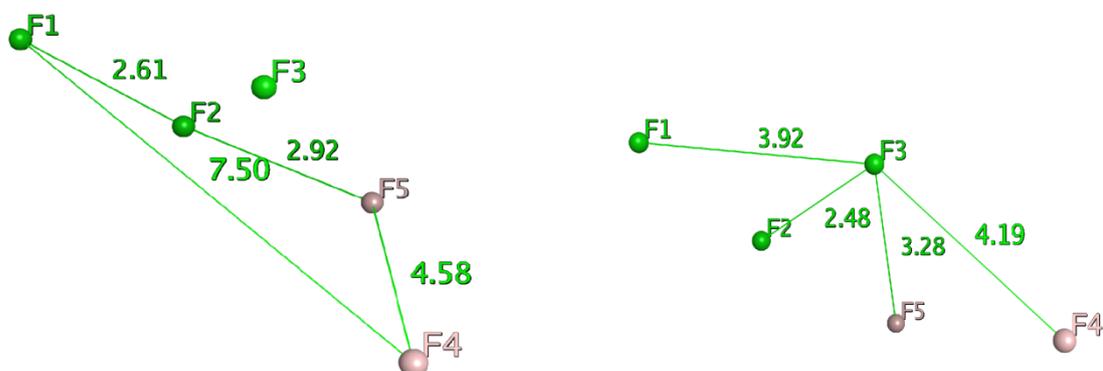


**Gráfico 3** – Número de moléculas detetadas nos controlos positivo e negativo durante o refinamento do MF com base na estrutura do recetor.

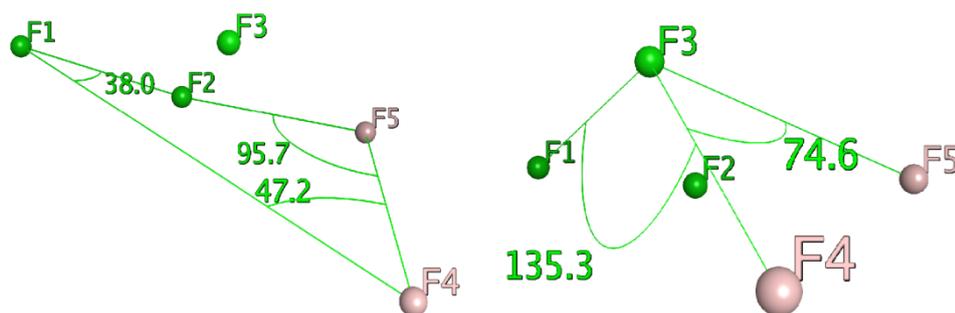
Nas figuras 26, 27 e 28 encontra-se o arranjo tridimensional do farmacóforo obtido com base na estrutura do local activo CT-L do recetor, assim como as respetivas características farmacofóricas e raios de cada uma, distâncias entre características e ângulos. Comparando este MFR1 com o MFL4 que foi obtido com base na estrutura de ligandos, verificamos que este MFR1 tem mais uma característica farmacofórica que o MFL4 (no MFL4 existem duas características Hyd – F2 e F3 – uma característica Acc|Don – F1 – e uma característica Acc – F4). Analisando as características Hyd de ambos os modelos, neste MF apenas uma das características farmacofóricas apresenta um raio de 1.4 Å (F1), sendo os raios das restantes características Hyd inferiores, enquanto que no MFL4 as duas características Hyd apresentam um raio de 1.4 Å. Quanto às características Acc|Don e Acc, no MF com base na estrutura do recetor estas apresentam raios superiores (de 1.7 Å e 1.8 Å), enquanto que no MFL4 a característica Acc|Don apresentam um raio de 1.4 Å e a característica Acc tem um raio de 1.1 Å. Em termos das distâncias entre as características, a maior distância neste MFR1 verifica-se entre as características F1 e F4 (7.5 Å), enquanto que no MFL4 a maior distância existente é entre as características F1 e F3 (9.0 Å).



**Figura 26** – O MF final obtido com base na estrutura do local ativo CT-L do proteassoma. Características farmacofóricas e raio de cada característica. F1 Hyd (raio 1.4), F2 Hyd (raio 1.1), F3 Hyd (raio 1), F4 Acc|Don (raio 1.8), F5 Acc|Don (raio 1.7).



**Figura 27** – O MF final obtido com base na estrutura do local ativo CT-L do proteassoma. Distâncias (em Å) entre as características farmacofóricas.



**Figura 28** – O MF final obtido com base na estrutura do local ativo CT-L do proteassoma. Principais ângulos entre as características farmacofóricas.

Em termos da percentagem de compostos *hit* ativos detetada, neste MF esta foi inferior à taxa detetada pelo MFL4 (69% comparativamente a 78%), podendo dever-se a

diferença em questão ao facto de que os compostos ativos não interajam com os diferentes *pockets* do local ativo da mesma forma, podendo tal estar relacionado não só com o tamanho de cada molécula, mas também com os grupos funcionais presentes. Relativamente aos falsos positivos, a taxa detetada nos dois MF é muito semelhante, apresentando este MFR1 uma percentagem de 15% e o MFL4 uma percentagem de 17%.

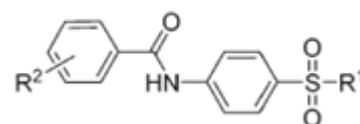
Alguns dos exemplos mais relevantes dos compostos da base de dados do controlo positivo detetados por este farmacóforo incluem: 2-aminobenzilastatina, bortezomib, carfilzomib, CEP-18770, CVT-659, glidobactina A, homobelactosina, marizomib, MG-132, MLN2238, NLVS, ONX0912, ONX0914, omuralide, PR-957, TMC95A, ritonavir.

Relativamente à base de dados de controlo negativo, são exemplos dos compostos detetados a tiropeptina A, os compostos TPI01 e TPI02 de Lei *et al.*<sup>76</sup> e o composto 28 de Pham *et al.*<sup>37</sup>

#### 4.3 – VIRTUAL SCREENING

O VS é uma técnica que permite identificar compostos ativos em bases de dados *in silico* e pode ser efetuado com base na estrutura de ligandos (*ligand based*) ou com base na estrutura do recetor (*structure/target based*).<sup>85</sup>

Basse *et al.*<sup>38</sup> recentemente utilizaram VS para identificar novos inibidores do proteassoma. Como estrutura do recetor, utilizaram as subunidades L e M do código PDB IIRU (*Bos taurus*) que, segundo os autores, correspondem às subunidades  $\beta 5c$  e  $\beta 6c$  (no entanto, tal como referimos anteriormente, a comparação dos dados obtidos na bibliografia e a estrutura cristalográfica propriamente não é coerente em termos das subunidades que formam o local ativo CT-L). Para validar o *pocket*, *dockaram* o bortezomib com o *software* Surflex, tendo o inibidor ficado posicionado de forma muito semelhante ao bortezomib cristalográfico. Para o VS utilizaram três *softwares* (FRED, Surflex e o LigandFit), tendo sido selecionadas moléculas que, posteriormente, foram testadas *in vitro* no proteassoma 20S de coelho, tendo sido avaliadas não só a atividade CT-L, mas também a atividade C-L e T-L. O composto denominado 2b (figura 29) revelou ser um inibidor promissor.



**Figura 29** – O inibidor 2b obtido por Basse *et al.* através de VS. R<sup>1</sup>=N(CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>)<sub>2</sub> e R<sup>2</sup>=4-C(CH<sub>3</sub>)<sub>3</sub>.<sup>38</sup>

No trabalho por nós desenvolvido, foram realizados diversos cálculos de VS, tanto baseados na estrutura de ligandos, como baseados na estrutura do alvo (neste caso, o proteassoma humano), com o objetivo de identificar novas moléculas com atividade inibitória.

A base de dados *in silico* escolhida para os diversos *screenings* efetuados foi a base de dados de compostos do *National Cancer Institute* (NCI), incluindo esta 265,242 moléculas devidamente protonadas e com as respetivas fórmulas e pesos moleculares, assim como as respetivas coordenadas tridimensionais de cada átomo (*download* das moléculas em formato SDF efetuado a partir de <http://cactus.nci.nih.gov/download/nci/><sup>119</sup> em Maio de 2013). A base de dados foi dividida em bases de dados mais pequenas (cada uma com cerca de 5000 moléculas) que foram pesquisadas paralelamente de modo a diminuir o tempo de cálculo.

No VS baseado na estrutura de ligandos utilizou-se como filtro o farmacóforo obtido com base em inibidores conhecidos do proteassoma humano desenvolvido com base nas epoxicetonas, tendo sido de seguida efetuado o *docking* dos compostos resultantes no modelo de homologia desenvolvido com base no local ativo CT-L ( $\beta$ 5c- $\beta$ 6c).

O VS baseado na estrutura do recetor teve como base o *docking* molecular da base de dados do NCI de modo a pesquisar compostos que interagissem favoravelmente com o local ativo do modelo de homologia com atividade CT-L e que apresentassem uma boa afinidade com o alvo.

Todos os cálculos foram realizados no *cluster* de cálculo científico da Faculdade de Farmácia da Universidade de Lisboa.

#### **4.3.1 – VIRTUAL SCREENING COM BASE NA ESTRUTURA DE LIGANDOS**

O farmacóforo obtido com base na estrutura de diversas epoxicetonas (ver 4.2.1) serviu de base a este *screening*, sendo o objetivo desta metodologia funcionar como um filtro das moléculas da base de dados do NCI tendo como base as características farmacofóricas existentes no modelo. Deste modo, o *docking* efetuado posteriormente incidiu apenas nas moléculas filtradas pelo farmacóforo cujas estruturas à partida terão alguma semelhança estrutural com moléculas ativas na subunidade CT-L do proteassoma.

Utilizando inicialmente o farmacóforo como filtro, das 265,242 moléculas da base de dados do NCI foram detetados 58,610 potenciais compostos *hit*, o que equivale a uma percentagem de 22% do total da base de dados. A aplicação deste filtro foi efetuada com o

software MOE 2012.10 e teve como objetivo selecionar apenas moléculas com características comuns a inibidores conhecidos representadas no MF, esperando-se que o conjunto de moléculas detetado esteja em parte livre de moléculas inativas.

De seguida, efectuou-se um novo *screening* utilizando-se o software GOLD 5.1.0 para proceder ao *docking* dos potenciais compostos *hit* no local ativo das subunidades  $\beta 5c$ - $\beta 6c$  e agrupá-los por ordem decrescente de afinidade (*score*). Neste segundo *screening* foi realizado um *docking* não covalente das 58,610 moléculas no modelo de homologia da subunidade catalítica  $\beta 5c$  do proteassoma humano. Em termos das condições em que o *docking* decorreu, definiu-se a Thr1 como estando no centro do local ativo no modelo de homologia das subunidades  $\beta 5c$ - $\beta 6c$  do proteassoma humano, definindo-se um raio de interação de 15 Å para o espaço de procura. Foram efetuadas 50 corridas para cada molécula da base de dados e selecionadas as 3 melhores configurações para cada molécula. A velocidade do *docking* foi definida como 30% de modo a obter uma velocidade de procura relativamente rápida, não foram colocadas moléculas de água no local ativo e a função de *scoring* utilizada foi o GoldScore.

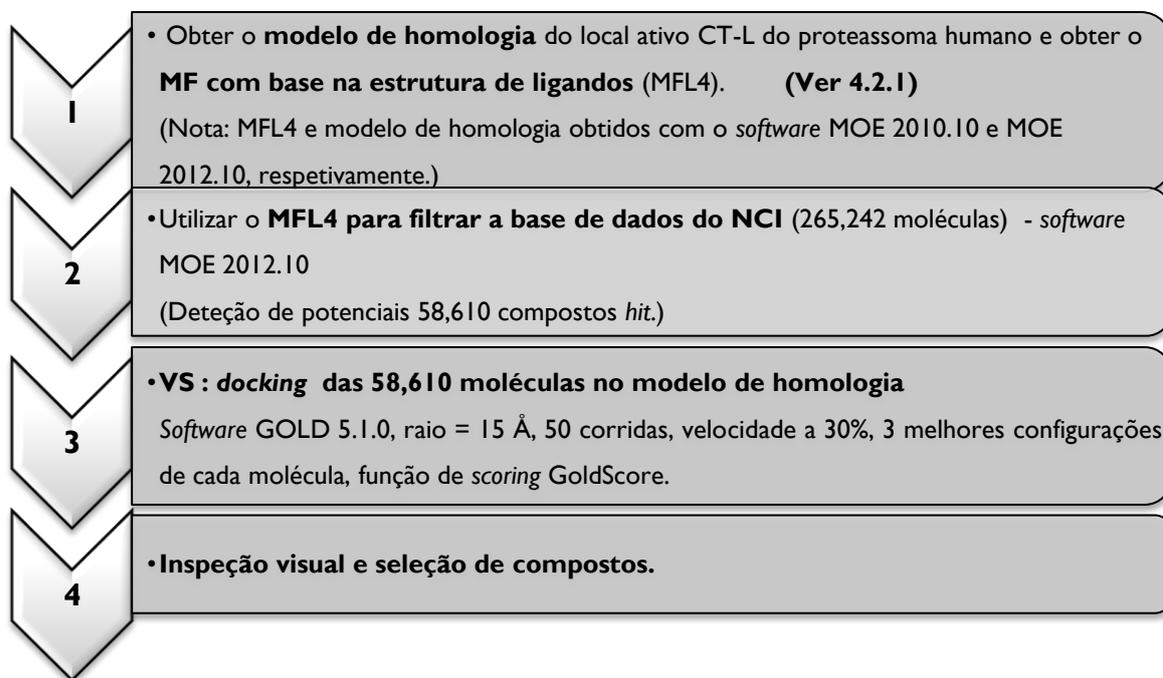


Figura 30 – VS com base na estrutura de ligandos.

### 4.3.2 – VIRTUAL SCREENING COM BASE NA ESTRUTURA DO RECETOR

Com o objetivo de identificar novos inibidores do proteassoma humano foi utilizado um protocolo de VS que consistiu no *docking* direto da base de dados do NCI no modelo de homologia desenvolvido em 4.1.

Para tal, foi efetuado o *docking* não covalente das moléculas da base de dados do NCI (265,242 moléculas) utilizando o *software* GOLD 5.1.0. A Thr1 foi definida como sendo o centro do local ativo do modelo de homologia da subunidade  $\beta$ 5c humana do *pocket*  $\beta$ 5c- $\beta$ 6c, tendo sido definido um raio de 15 Å para a interação dos ligandos. Foram efetuadas 50 corridas para cada molécula e selecionadas as três melhores configurações para cada molécula. Foi realizado um *docking* preliminar no qual a velocidade do *screening* foi definida como 30% (dado que se pretende uma maior velocidade de procura, apesar de o espaço de procura ser menos explorado), não foram colocadas moléculas de água explícitas no local ativo e a função de *scoring* utilizada foi o GoldScore.

De seguida, efetuou-se um segundo *screening* para refinar os resultados obtidos previamente. Começou-se por analisar os resultados obtidos e decidiu-se que, para este segundo *screening* apenas seriam *redockadas* as moléculas que obtiveram um *score* a partir de 60, inclusive. Este valor de *score* foi definido tendo como base os *scores* obtidos quando foram previamente *dockados* inibidores conhecidos da actividade CT-L do proteassoma na escala do nM (nomeadamente o bortezomib, o carfilzomib, a felutamida B, entre outros exemplos) no modelo de homologia utilizado. Utilizando a função GoldScore verificou-se, por exemplo, que o bortezomib obteve um *score* de 61.41 e que, de uma forma geral, os restantes inibidores *dockados* apresentaram valores superiores ou iguais a 60. Este valor de *score* mínimo estipulado levou a que fossem selecionadas 20,376 moléculas, tendo estas moléculas sido novamente *dockadas* com o *software* GOLD versão 5.1.0, sendo que, relativamente aos *screenings* anteriores, as diferenças manifestaram-se no facto de terem sido efetuadas 100 corridas para cada molécula e de a velocidade do *screening* ter sido definida como 100% dado que um *docking* mais lento do que no primeiro *screening* permite ao GOLD tentar explorar melhor o espaço de procura.

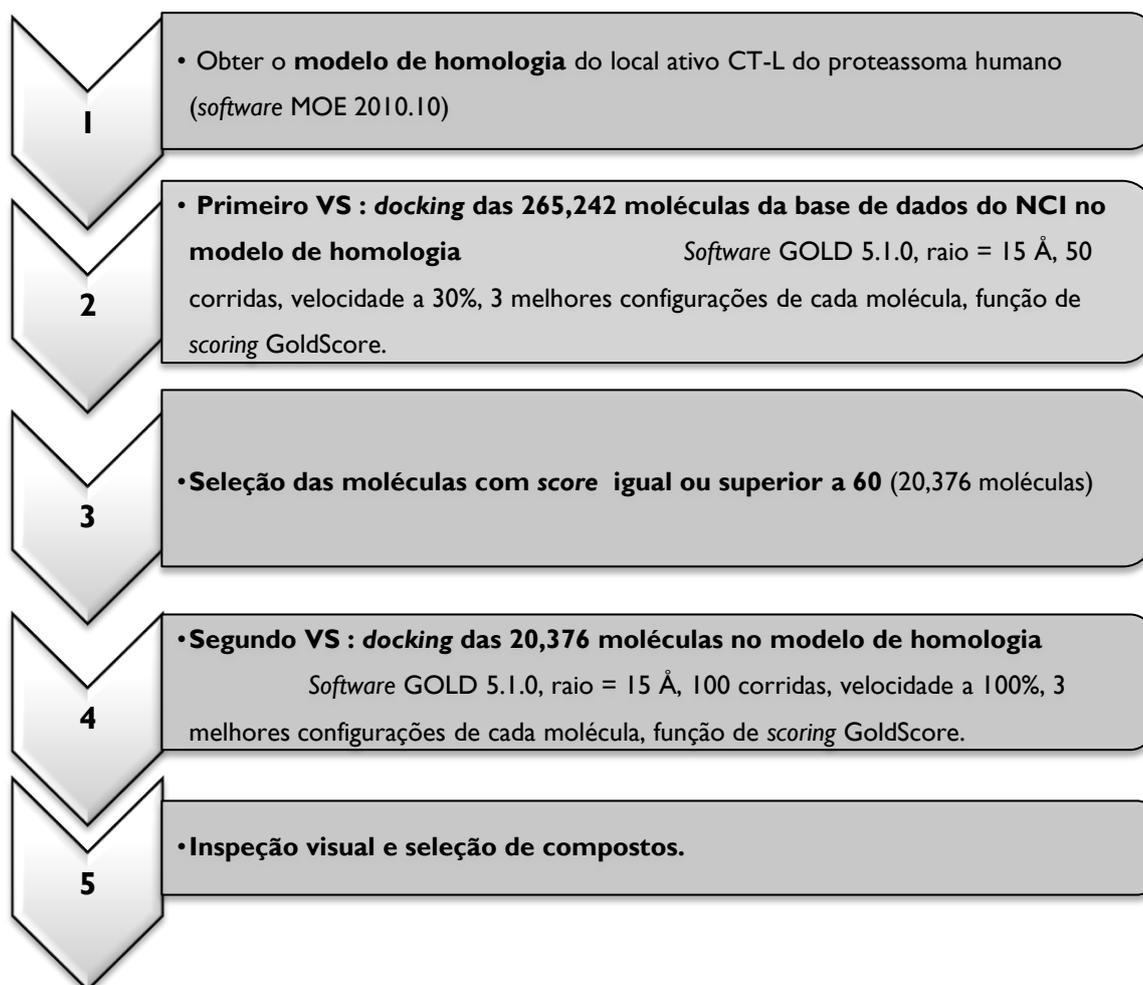


Figura 31 – VS com base na estrutura do recetor.

#### 4.3.3 – DISCUSSÃO DE RESULTADOS DO VIRTUAL SCREENING

Após o VS, os compostos *dockados* foram visualmente inspecionados, tendo sido analisadas as primeiras 150 moléculas com melhor *score* em cada um dos *screenings*. Destas 150 moléculas de cada *screening* foram excluídas, por exemplo, as moléculas que possuíam metais (nomeadamente metais pesados, como o chumbo) e compostos com estruturas tipo porfirina dado que também possuem metais na sua composição.

No caso do VS com base na estrutura de ligandos, foram selecionadas 65 moléculas. Os códigos correspondentes a cada uma das moléculas encontram-se no anexo 8, sendo possível aceder às estruturas e à informação de cada molécula no *site* PubChem (<http://pubchem.ncbi.nlm.nih.gov/>)<sup>120</sup>. No entanto, na figura 32 encontram-se algumas das moléculas selecionadas.

No caso do VS com base na estrutura do recetor, após inspeção visual foram seleccionadas 44 moléculas. No anexo 9 é possível consultar os códigos dos compostos seleccionados, estando todas os compostos e respetiva informação disponíveis no site Pubchem (<http://pubchem.ncbi.nlm.nih.gov/>)<sup>120</sup>. Na figura 33 encontram-se alguns exemplos das moléculas seleccionadas neste *screening*.

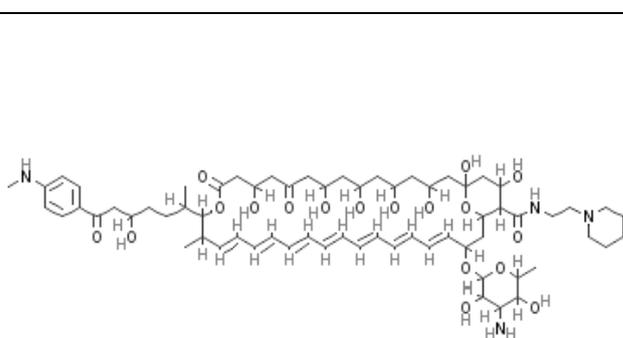
Nos anexos 8 e 9 encontram-se ainda dados relativos ao score obtido por cada composto no VS (sendo que o score representa uma estimativa da energia livre de Gibbs), informação sobre o número de átomos pesados e o valor da eficiência do ligando (*ligand efficiency* – LE) calculado para cada composto.

A LE é uma medida da energia de ligação por átomo do ligando relativamente ao recetor com o qual interage. Trata-se de um conceito relevante na descoberta de novos fármacos dado que, em conjugação com um estudo das propriedades físico-químicas e farmacológicas, permite otimizar a procura de potenciais compostos *lead*.<sup>121,122</sup>

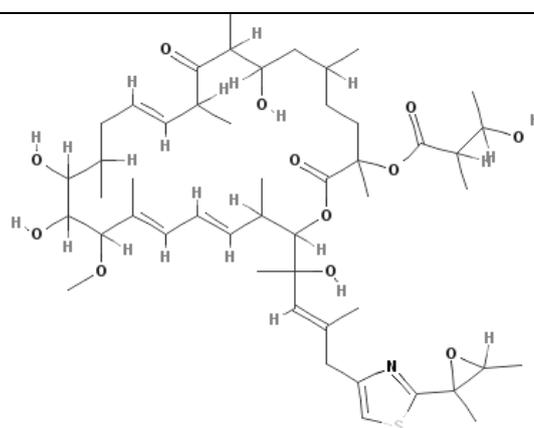
A LE permite avaliar se um ligando deve a sua potência de ação ao facto de se ajustar perfeitamente ao recetor ou por estabelecer muitas interações com este. Regra geral, a LE diminui com o aumento do número de átomos pesados do composto, podendo a biodisponibilidade ser comprometida.<sup>123</sup>

Numericamente, podemos definir a LE como sendo o quociente entre a energia livre de Gibbs ( $\Delta G$ ) e o número de átomos não hidrogénio/átomos pesados (N) do composto<sup>122</sup>:

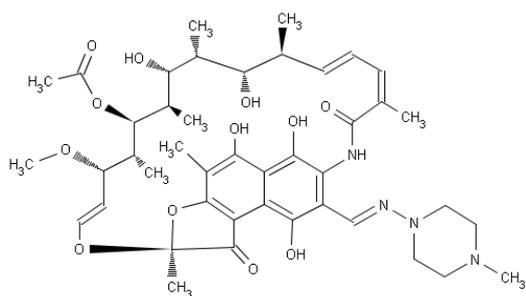
$$LE = \frac{\Delta G}{N}$$



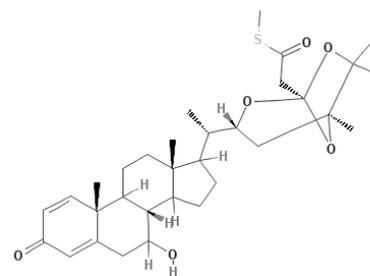
**NSC722655**



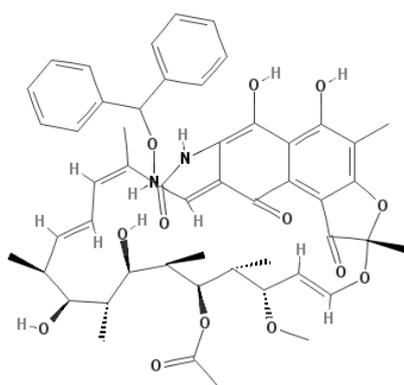
**NSC613013 (Patelazole B)**



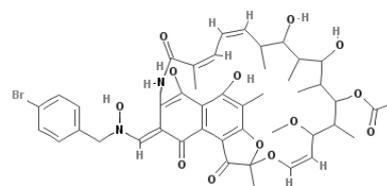
**NSC113926 (Rifampicina)**



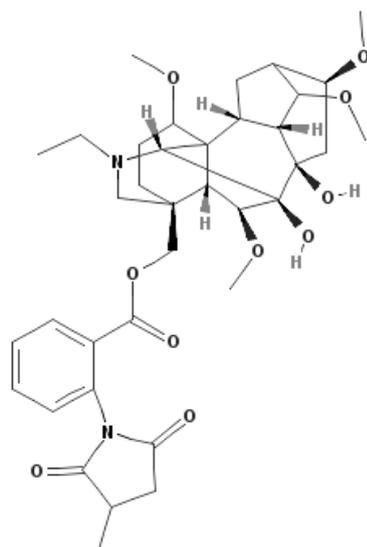
**NSC633664**



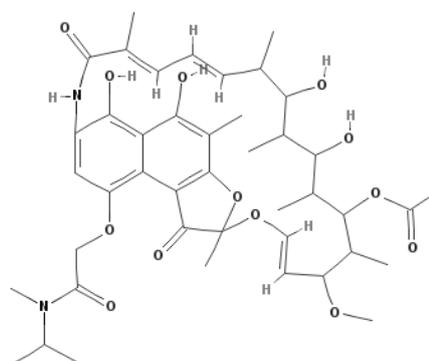
**NSC143471 (Rifamicina AF)**



**NSC146730**

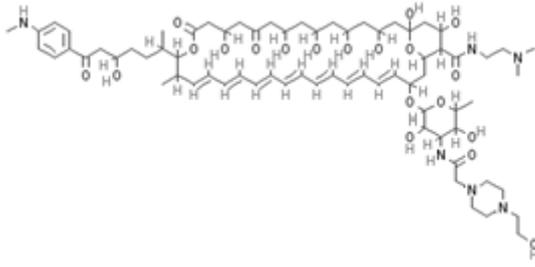


**NSC624746**

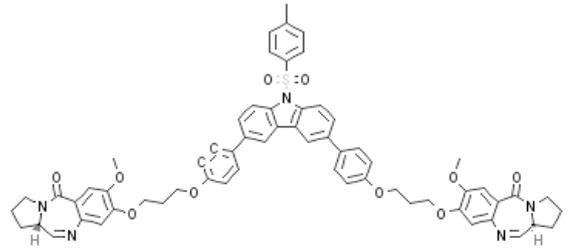


**NSC143436 (Rifamicina B  
isopropilmetilamida)**

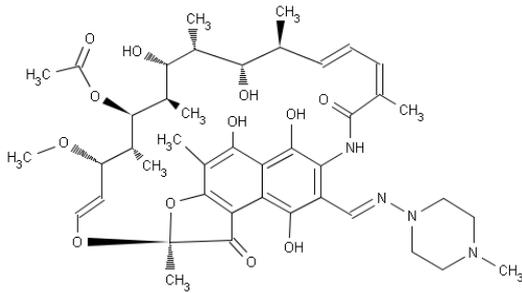
**Figura 32 – Algumas das moléculas do VS com base na estrutura de ligandos.**



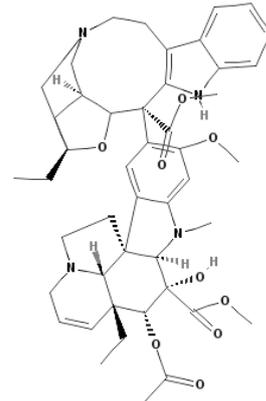
**NSC722658**



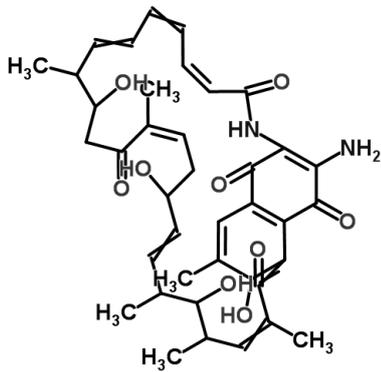
**NSC744991**



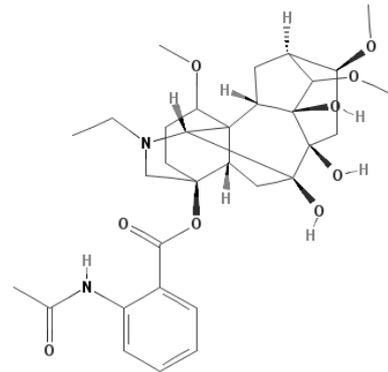
**NSC113926 (Rifampicina)**



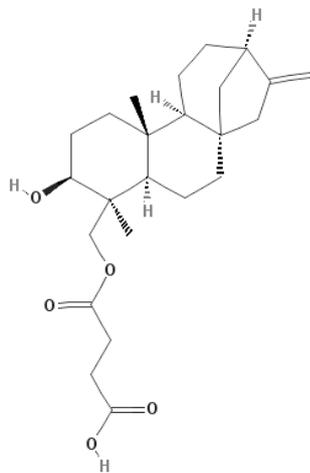
**NSC304422**



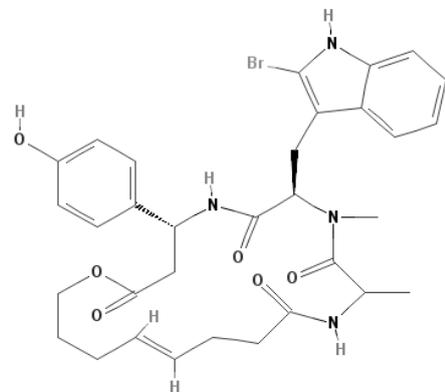
**NSC655331**



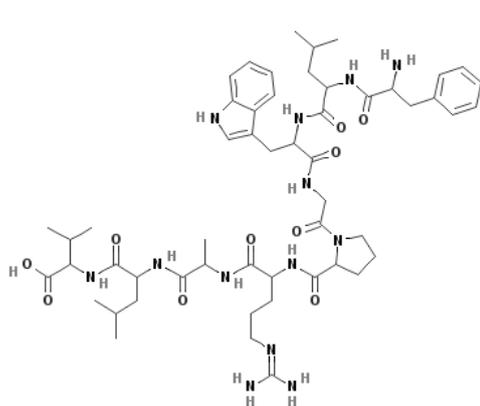
**NSC624751 (Ranaconitina)**



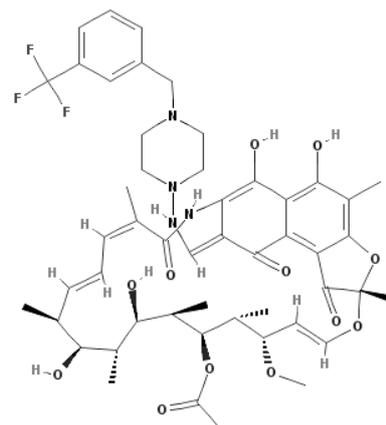
**NSC83442**



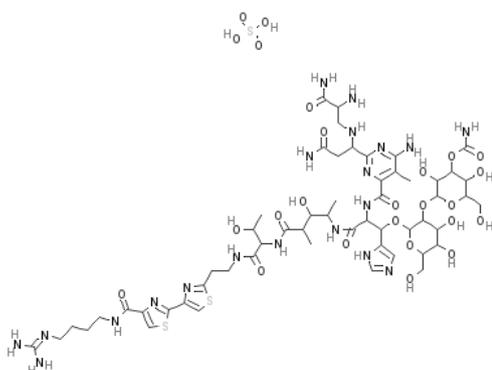
**NSC617200**



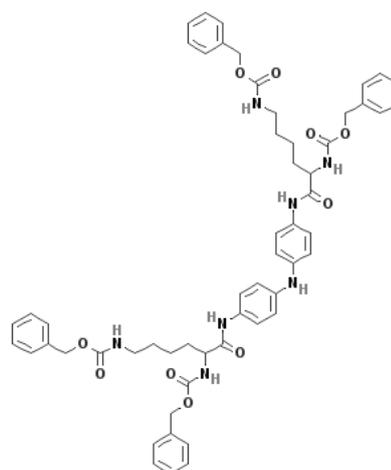
**NSC724849**



**NSC173710**



**NSC624749 (Sulfato de bleomicina)**



**NSC110279**

**Figura 33** – Algumas das moléculas do VS com base na estrutura do recetor.

Comparando as 150 moléculas inspeccionadas em cada *screening*, existem 39 moléculas em comum entre os dois *screenings*. No entanto, ao nível dos compostos selecionados, existem apenas 15 moléculas em comum cujos códigos correspondentes são: NSC722658, NSC722660, NSC722659, NSC722657, NSC722654, NSC722655, NSC722661, NSC744991, NSC726291, NSC113926, NSC208397, NSC381101, NSC624751, NSC165124 e NSC617200.

Comparando a LE dos compostos escolhidos, no caso do *screening* com base na estrutura de ligandos esta varia entre 1.23 (código NSC208397) e 3.21 (código NSC633664), existindo 18 compostos com LE maior ou igual que 2 (ou seja, 28% dos compostos selecionados neste *screening*); no *screening* com base na estrutura do recetor a LE varia entre

0.93 (código NSC 672462) e 3.94 (código NSC83442), sendo 25 os compostos com LE igual ou superior a 2 (o que corresponde a 57% dos compostos selecionados neste *screening*).

No VS com base na estrutura de ligandos, as primeiras moléculas selecionadas (por exemplo, NSC722655, NSC722659 e NSC722657) apresentam ligações peptídicas e, para além disso, têm uma estrutura algo semelhante à estrutura base dos carotenóides.

No composto NSC633664 podemos verificar que o seu núcleo base é muito semelhante ao núcleo característico dos esteróides – o núcleo ciclopentanoperhidrofenantreno – o que poderá ser relevante na medida em que os corticosteróides são utilizados, por exemplo, como imunossuppressores no tratamento de patologias autoimunes.<sup>124</sup>

Neste *screening* que utilizou o farmacóforo como filtro foram detetados diversos compostos que foram previamente alvo de bioensaios, dos quais são exemplo:

- NSC633249 – foi aplicado a linhas celulares de cancro do ovário e de leucemia, no entanto o composto mostrou-se inativo;
- NSC143416 – utilizado no NCI em ensaios antivirais HIV e tendo sido detetado no *screening* de fármacos anticancerígenos do NCI, no entanto revelou-se inativo em ambos os casos;
- NSC633249 – utilizaram-se linhas celulares de cancro do ovário e de leucemia, mas o composto mostrou-se inativo;
- NSC624751 (ranaconitina) – alcalóide diterpénico que mostrou ser inativo no ensaio antiviral de HIV do NCI;
- NSC304412 – composto antineoplásico mas que revelou ser inativo na leucemia no *screening in vivo* de fármacos anticancerígenos do NCI.<sup>120</sup>

Foram ainda detetados no *screening* com base na estrutura de ligandos alguns compostos que apresentam diversas propriedades farmacológicas interessantes, nomeadamente propriedades antibacterianas e/ou antineoplásicas:

- NSC613013 (patelazole B) e NSC613015 (patelazole C) – macrólidos citotóxicos;
- NSC113926 (rifampicina) – um derivado semi-sintético da rifamicina B que, entre outros organismos, é ativo em diversas formas de *Mycobacterium*;
- NSC56409 (*werramycin A*/ nonactina) – antibiótico que demonstrou ser ativo no ensaio inibitório do crescimento de linhas celulares neoplásicas de cancro renal e cancro do ovário;
- NSC143424 (rifamicina dibenzilamida);

- NSCI43429 (rifamicina B ciclohexilmetilamida);
- NSCI43426 (rifamicina B dibutilamida);
- NSC 143428 (rifamicina B ciclopentilmetilamida);
- NSCI43471 (rifamicina AF);
- NSCI43416 (N-dimetilrifampicina);
- NSCI65124 (3-formilrifamicina SV o-deciloxima);
- NSCI43418 (rifamida/rifamicina B dietilamida) – antibiótico macrólido.<sup>120</sup>

No VS com base na estrutura do recetor, as primeiras moléculas selecionadas (por exemplo, NSC722658, NSC722660 e NSC722659) são iguais às moléculas do *screening* anteriormente analisado, apresentando a mesma estrutura tipo carotenóide e ligações peptídicas.

No composto NSC744991 podemos constatar que se trata de uma sulfonamida, o que poderá ser relevante na medida em que as sulfonamidas são antibióticos utilizados no tratamento de doenças infecciosas causadas por diversos microorganismos (por exemplo, a sulfadiazina pode ser utilizada no tratamento de infeções urinárias).<sup>124</sup>

O composto NSC304422 apresenta um núcleo muito semelhante ao indol (o núcleo base dos alcalóides). No caso do composto NSC617200, este apresenta um núcleo indol completo e ainda ligações peptídicas.

As moléculas NSC724849 e NSCI10279 foram selecionadas devido ao facto de terem bastantes ligações peptídicas e estas estarem presentes em alguns inibidores do proteassoma.

O composto NSCI73710 poderá ter interesse devido ao facto de ter átomos de flúor (que geralmente tornam um fármaco mais resistente ao metabolismo). Para além disso, podemos verificar que tem uma ligação peptídica.

Neste *screening* podemos mencionar como exemplo de compostos que foram previamente alvo de bioensaios as seguintes moléculas:

- NSC624751 (ranaconitina) – alcalóide diterpénico que mostrou ser inativo no ensaio antiviral de HIV do NCI;
- NSC83442 – revelou ser inativo na leucemia no *screening in vivo* de fármacos anticancerígenos do NCI;
- NSC648060 – mostrou ser ativo no ensaio inibitório do crescimento de linhas celulares neoplásicas de leucemia e cancro do ovário;

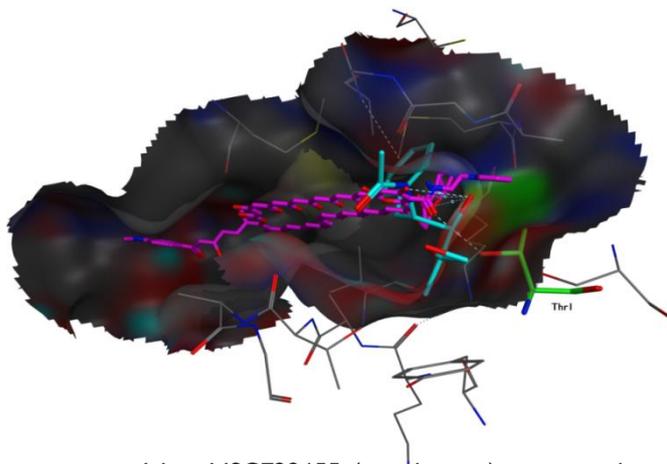
- NSC110279 – obtido no *screening* de fármacos anticancerígenos do NCI, mostrou ser ativo em ensaios biológicos;
- NSC697674 – deu resultado positivo no ensaio biológico antiviral do NCI para o HIV.<sup>120</sup>

Em termos da deteção de compostos com propriedades farmacológicas importantes (tais como propriedades antibacterianas e/ou antineoplásicas), no VS com base na estrutura do recetor temos os seguintes exemplos:

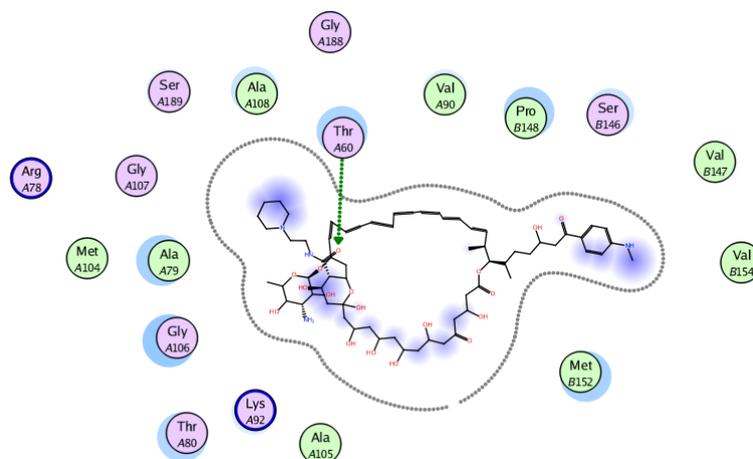
- NSC56409 (*werramycin A*/ nonactina) – antibiótico que demonstrou ser ativo no ensaio inibitório do crescimento de linhas celulares neoplásicas de cancro renal e cancro do ovário;
- NSC304422 (catarantamina) – um alcalóide da vinca que mostrou ter atividade num modelo tumoral de leucemia;
- NSC624751 (ranaconitina) – alcalóide diterpénico que mostrou ser inativo no ensaio antiviral de HIV do NCI;
- NSC624749 (bleomicina B2, sulfato) – antibiótico citostático detetado por *screening in vivo* de fármacos anticancerígenos do NCI e que mostrou ser ativo no ensaio inibitório do crescimento de modelos tumorais após linhas celulares neoplásicas de leucemia e cancro do ovário;
- NSC139209 (DMB – rifampicina) – antibiótico detetado no *screening in vivo* de fármacos anticancerígenos do NCI.<sup>120</sup>

Para além de tudo o que já foi mencionado, é extremamente importante analisar as poses de *docking* dos compostos de modo a verificarmos se estes interagem ou não com o recetor no local pretendido e quais as interações estabelecidas com este.

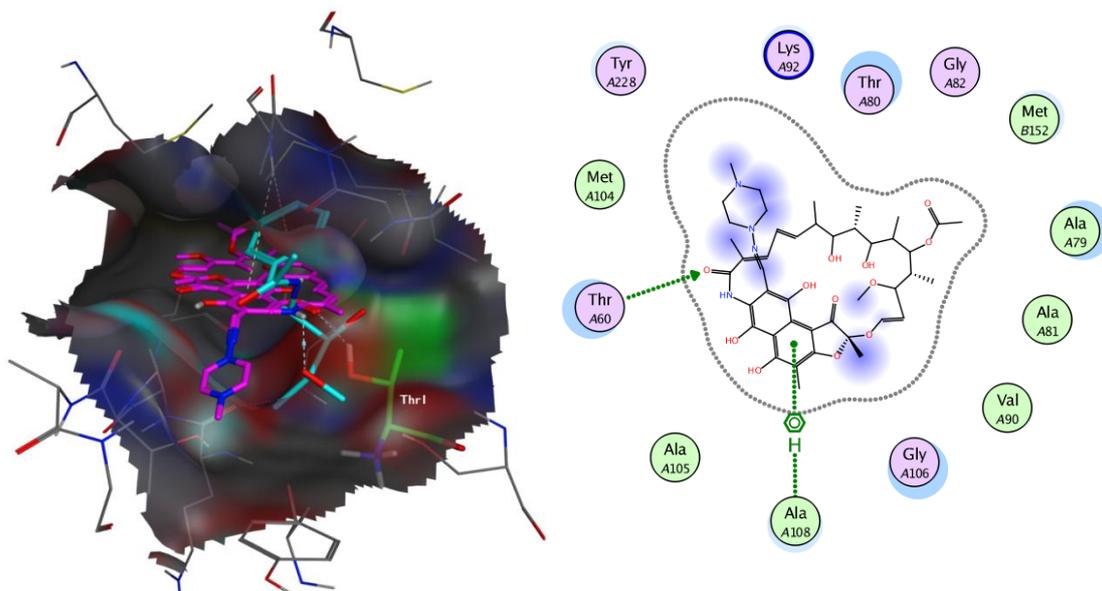
Como exemplos de poses de compostos *dockados* e respetivas interações com o recetor no VS com base na estrutura de ligandos, as figuras 34 e 35 são relativas ao ligando *dockado* com o código NSC 722655 (o ligando com melhor *score* no VS com base na estrutura de ligandos) e a figura 36 é referente ao ligando com o código NSC113926 (rifampicina). A Thr1 encontra-se a cor verde, estando o ligando *dockado* a cor-de-rosa e o PR-957 (o ligando cristalográfico do código PDB 3UN8) a azul.



**Figura 34** – O composto com o código NSC722655 (cor-de-rosa) posicionado no modelo de homologia e sobreposto com o PR-957 (azul). A Thr1 apresenta a cor verde.



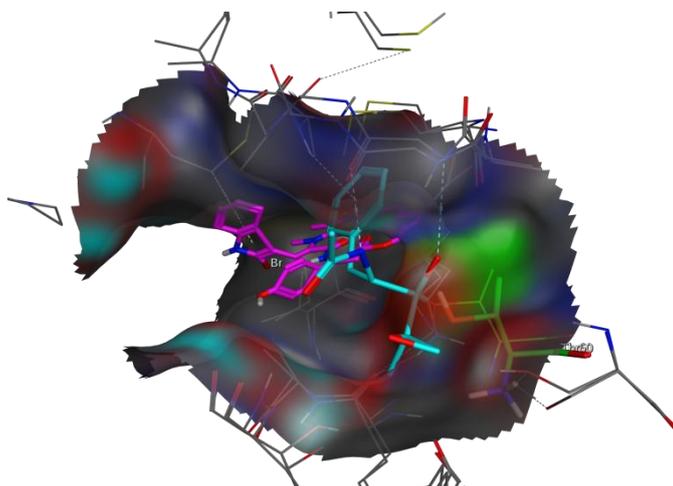
**Figura 35** – Interações do composto com o código NSC722655 *dockado* no modelo de homologia. (A Thr1 aparece representada por Thr60 dado que foi a numeração atribuída no modelo de homologia.)



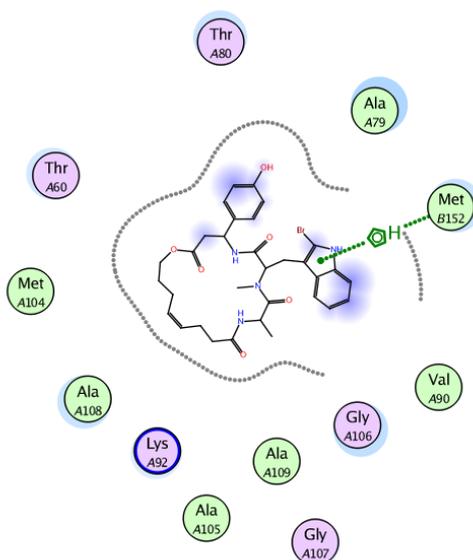
**Figura 36** – *Esquerda*: O composto com o código NSC113926 (cor-de-rosa) posicionado no modelo de homologia e sobreposto com o PR-957 (azul). A Thr1 apresenta a cor verde. *Direita*: Interações do composto com o código NSC722655 *dockado* no modelo de homologia. (A Thr1 aparece representada por Thr60 dado que foi a numeração atribuída no modelo de homologia.)

Nos dois exemplos mencionados, os ligandos ficaram bem posicionados no local ativo, chegando mesmo a estabelecer interações com a Thr I.

Relativamente ao VS com base na estrutura do recetor, o composto com o código NSC617200 encontra-se posicionado no local ativo (figura 37), no entanto interage com uma Met, encontrando-se a 2.78Å da Thr I (figura 38).



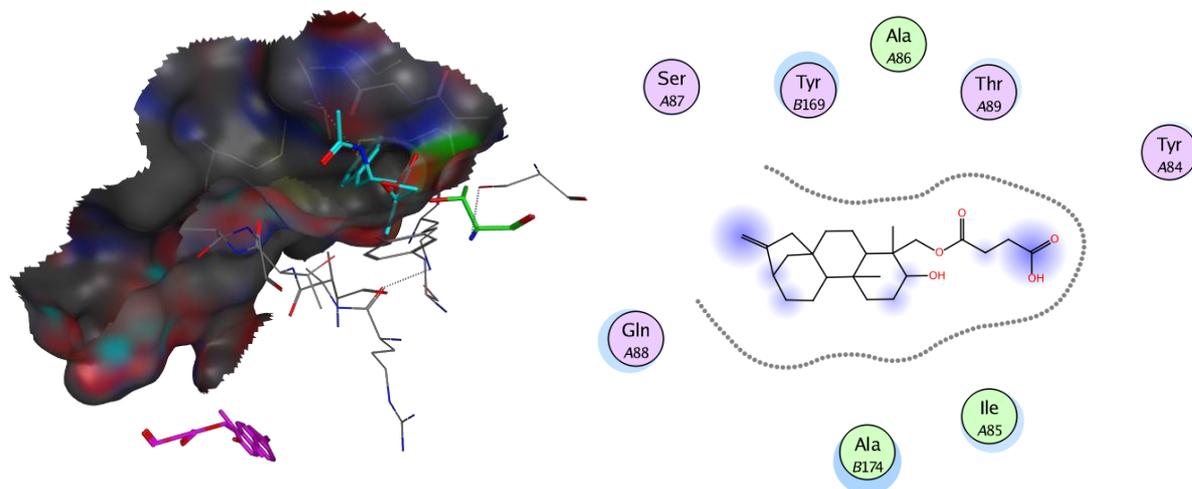
**Figura 37** – O composto com o código NSC617200 (cor-de-rosa) posicionado no modelo de homologia, encontrando-se o ligando cristalográfico PR-957 a azul e a Thr I a verde.



**Figura 38** – Interações do composto com o código NSC617200 *dockado* no modelo de homologia. (A Thr I aparece representada por Thr60 dado que foi a numeração atribuída no modelo de homologia.)

No entanto, nem sempre as moléculas ficam bem posicionadas no *pocket*, daí a importância de termos de analisar as poses dos compostos *dockados*. Por exemplo, no VS com base na estrutura do recetor, o ligando com o código NSC83442 apresentou uma

excelente LE (3.94) e, dado que se trata de uma molécula de menores dimensões que muitas das restantes moléculas da base de dados, seria de esperar que se conseguisse posicionar bem no *pocket*. No entanto, verificou-se que este ligando ficou bastante afastado da Thr1 (a 12.72 Å), não estabelecendo qualquer tipo de interações com esta (figura 39).



**Figura 39** – *Esquerda*: O composto com o código NSC83442 (cor-de-rosa) posicionado no modelo de homologia. A azul encontra-se o ligando cristalográfico PR-957 e a Thr1 apresenta a cor verde. *Direita*: Não se verificaram interações entre o composto com o código NSC83442 e a Thr1.

Tendo em conta que existem diversos compostos resultantes tanto do VS com base na estrutura de ligandos como do VS com base na estrutura do recetor que possuem ligações peptídicas na sua estrutura (por exemplo, os compostos NSC722655 e NSC724849), podemos concluir que estes compostos apresentam algumas semelhanças com inibidores peptídicos do proteassoma (dos quais são exemplo o TMC95A e o ritonavir).

O facto de terem sido detetadas algumas moléculas com núcleos iguais ou muito semelhantes aos núcleos bases de alguns grupos de compostos (por exemplo, os alcalóides) ou moléculas com atividade farmacológica conhecida, nomeadamente a rifampicina e o sulfato de bleomicina, tal poderá ser importante para estudos futuros em termos de eventuais novas aplicações terapêuticas.

## V – CONSIDERAÇÕES FINAIS E PERSPETIVAS FUTURAS

O sistema ubiquitina-proteassoma é essencial à homeostase, sendo o proteassoma um alvo importante para intervenção terapêutica em diversas patologias, nomeadamente neoplasias, doenças neurodegenerativas e patologias inflamatórias.

Sendo o proteassoma uma protease multicatalítica, os seus inibidores podem atuar em diversos locais, nomeadamente ao nível do local ativo C-L, T-L e CT-L.

Neste trabalho foi desenvolvido e validado um modelo de homologia para o local ativo CT-L que teve como base o código PDB 3UN8 (proteassoma 20S de levedura em complexo com o inibidor PR-957). Ao sobrepor o modelo de homologia com a estrutura cristalográfica utilizada como *template*, obteve-se um RMSD de 0.74 Å. Para além disso, o gráfico de Ramachandran mostrou que a maioria dos resíduos se encontra em zonas nucleares (nomeadamente os resíduos importantes na atividade catalítica). O *docking* de inibidores conhecidos (tais como o bortezomib e o carfilzomib) que atuam na subunidade catalítica  $\beta 5c$  mostrou que os compostos ficaram bem posicionados no local ativo em questão.

No decorrer deste trabalho foram desenvolvidos modelos farmacofóricos com o objetivo de diferenciar possíveis inibidores de moléculas não ativas, tendo dois desses modelos sido validados.

O farmacóforo desenvolvido com base na estrutura de inibidores conhecidos da subunidade catalítica  $\beta 5c$  do proteassoma, designado por MFL4, teve como ponto de partida a estrutura de epoxicetonas e detetou 78% dos compostos da base de dados de controlo positivo e 17% dos compostos da base de dados de controlo negativo, o que demonstra a sua seletividade para a deteção de potenciais compostos *hit*. O MFL4 apresenta as seguintes características farmacofóricas: F1 Acc|Don (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Acc (raio 1.1).

O segundo farmacóforo desenvolvido, MFR1, teve como base a estrutura tridimensional do local ativo CT-L do código PDB 3UN8 (cadeias K e L) e apresenta as seguintes características: F1 Hyd (raio 1.4), F2 Hyd (raio 1.1), F3 Hyd (raio 1), F4 Acc|Don (raio 1.8), F5 Acc|Don (raio 1.7). Uma vez que detetou 69% dos compostos da base de dados de controlo positivo e 15% dos compostos da base de dados de controlo negativo, podemos constatar que apresenta uma boa seletividade na deteção de potenciais compostos *hit*.

O VS realizado teve como base duas abordagens distintas: VS com base na estrutura de ligandos (no qual se utilizou o MFL4 como filtro) e VS com base na estrutura do recetor (modelo de homologia). Em ambos, a base de dados utilizada foi a base do NCI e foi utilizado o modelo de homologia obtido previamente para *dockar* os compostos.

Foram analisadas as 150 moléculas moléculas de cada *screening* que apresentaram melhor score, tendo sido selecionadas 65 moléculas no VS com base na estrutura de ligandos e 44 moléculas no VS com base na estrutura do recetor. Algumas das moléculas selecionadas (por exemplo, os compostos NSC722655 e NSC724849) apresentam algumas semelhanças com inibidores peptídicos do proteassoma dado que possuem ligações peptídicas. A presença de compostos com estruturas iguais ou muito parecidas com núcleos base de alguns compostos (por exemplo, o núcleo indol dos alcalóides), assim como a deteção de moléculas com atividade farmacológica conhecida e que são utilizadas na clínica, nomeadamente a rifampicina, poderá ser importante para estudos futuros em termos de eventuais novas aplicações terapêuticas.

O facto de existirem 15 moléculas em comum (NSC722658, NSC722660, NSC722659, NSC722657, NSC722654, NSC722655, NSC722661, NSC744991, NSC726291, NSC113926, NSC208397, NSC381101, NSC624751, NSC165124 e NSC617200) entre as moléculas selecionadas de entre as 150 primeiras moléculas de cada *screening* demonstra que não só o MFL4 selecionou estas moléculas como tendo características comuns a alguns inibidores conhecidos da subunidade catalítica  $\beta 5c$  proteassoma como, à partida, também foram algumas das moléculas melhor posicionadas no *docking*, tendo portanto com uma maior afinidade prevista.

Apesar de a eficiência do ligando nos dar alguma informação acerca da interação do ligando com a proteína, é extremamente importante visualizar o posicionamento de cada molécula no local ativo para verificar se de facto ocorrem as interações pretendidas e se o composto se encontra bem posicionado.

Resumindo, neste trabalho foram construídos com sucesso o primeiro modelo de homologia para as subunidades  $\beta 5c$  e  $\beta 6c$  do proteassoma humano e ainda modelos farmacofóricos de inibidores do proteassoma. Estes resultados permitiram realizar o VS de uma base de dados de compostos significativa (a base de compostos do NCI) e selecionar compostos que se prevê terem potencial atividade inibitória CT-L no proteassoma humano.

Futuramente, será importante proceder à realização de todo o procedimento computacional para os locais ativos C-L e T-L de modo a selecionar potenciais inibidores que atuem nestas subunidades e identificar compostos que tenham atividade em mais de um local ativo. Para além disso, terá de ser efetuada toda a avaliação biológica das moléculas selecionadas não só no trabalho presente mas também em trabalhos futuros de modo a obtermos informação mais detalhada.

## VI – ANEXOS

## ANEXO I

<b>Código PDB</b>	<b>Descrição do código PDB</b>	<b>Cadeia(s) e genes que codificam para as subunidades catalíticas</b>	<b>Resolução e data de revisão</b>
<b>1PMA</b>	Proteassoma de <i>Thermoplasma acidophilum</i>	1, 2, B, P, Q, R, S, T, U, V, W, X, Y, Z: subunidades $\beta$	3.40 Å 24/02/2009
<b>1Q5Q</b>	Proteassoma 20S de <i>Rhodococcus</i>	H I J K L M N: subunidade I tipo $\beta$	2.60 Å 13/07/2011
<b>3KRD</b>	Estrutura cristalográfica do proteassoma de <i>Mycobacterium Tuberculosis</i> em complexo com Felutamida B	2 C E G H J L N P R T V X Z: subunidade $\beta$ do proteassoma	2.50 Å 13/07/2011
<b>2FHH</b>	Estrutura cristalográfica do proteassoma de <i>Mycobacterium Tuberculosis</i> em complexo com o inibidor MLN-273	2 C E G H J L N P R T V X Z: subunidade $\beta$ do proteassoma	2.99 Å 13/07/2011
<b>3MG0</b>	Estrutura do proteassoma 20S de levedura com o bortezomib	K Y: PRE2* 2 N: PRE3* H V: PUP1*	2.68 Å 13/07/2011
<b>2F16</b>	Estrutura cristalográfica do proteassoma 20S de levedura com o bortezomib	K Y: PRE2* 2 N: PRE3* H V: PUP1*	2.80 Å 13/07/2011
<b>2ZCY</b>	Complexo do proteassoma 20S de levedura com syringolin A	K Y: PRE2* I N: PRE3* H V: PUP1*	2.90 Å 13/07/2011
<b>3UN4</b>	Proteassoma 20S de levedura em complexo com PR-957 (morfolina)	K Y: PRE2* N b: PRE3* H V: PUP1*	3.40 Å 28/03/2012
<b>3UN8</b>	Proteassoma 20S de levedura em complexo com PR-957 (epóxido)	K Y: PRE2* N b: PRE3* H V: PUP1*	2.70 Å 28/03/2012
<b>3SHJ</b>	Proteassoma em complexo com o derivado hidroxiureia HUI0	K Y: PRE2* 2 N: PRE3* H V: PUP1*	2.80 Å 18/01/2012
<b>3DY4</b>	Estrutura cristalográfica do proteassoma 20S de levedura em complexo com <i>spiolactacystin</i>	K Y: PRE2* 2 N: PRE3* H V: PUP1*	2.80 Å 19/01/2010
<b>3E47</b>	Estrutura cristalográfica do proteassoma 20S de levedura em complexo com Homobelactosina C	K Y: PRE2* 2 N: PRE3* H V: PUP1*	3.00 Å 13/07/2011
<b>1IRU</b>	Estrutura cristalográfica do proteassoma 20S de mamífero com resolução de 2.75 Å ( <i>Bos taurus</i> )	I M: $\beta$ 1 K Y: $\beta$ 2 L Z: $\beta$ 5	2.75 Å 13/07/2011
<b>3UNB</b>	Proteassoma 20S constitutivo de rato em complexo com PR-957	L Z n 2: $\beta$ 1 J X l z: $\beta$ 2 H V j x: $\beta$ 10 ( $\beta$ 2i) I K Y m: $\beta$ 8 ( $\beta$ 5i) 4 N b p: $\beta$ 9 ( $\beta$ 1i)	2.90 Å 28/03/2012
<b>3UNE</b>	Proteassoma 20S constitutivo de rato	2 L Z n: $\beta$ 1 J X l z: $\beta$ 2 I K Y m: $\beta$ 5	3.20 Å 28/03/2012
<b>3UNF</b>	Imunoproteassoma 20S de rato em complexo com PR-957	L Z: $\beta$ 1 J X: $\beta$ 2 H V: $\beta$ 10 ( $\beta$ 2i) K Y: $\beta$ 8 ( $\beta$ 5i) N b: $\beta$ 9 ( $\beta$ 1i)	2.90 Å 28/03/2012
<b>3UNH</b>	Imunoproteassoma 20S de rato	L Z: $\beta$ 1 J X: $\beta$ 2	3.20 Å 28/03/2012

H V:  $\beta 10$  ( $\beta 2i$ )  
 K Y:  $\beta 8$  ( $\beta 5i$ )  
 N b:  $\beta 9$  ( $\beta 1i$ )

**Estruturas cristalográficas publicadas após elaboração dos modelos de homologia**

<b>4BGR</b>	A estrutura do proteassoma 26S com ligação a ATP- $\gamma$ S auxilia na compreensão do mecanismo de translocação do substrato dependente de nucleótido	1: $\beta 1$ 2: $\beta 2$ 5: $\beta 5$	9.8 Å (microscopia eletrónica) 15/05/2013
<b>4EU2</b>	Estrutura cristalográfica do proteassoma 20S com o novo inibidor K-7174	L Z: PRE2* H V: PRE3* I W: PUP1*	2.51 Å 01/05/2013
<b>4INR</b>	Proteassoma 20S de levedura em complexo com a vinilsulfona LUI02	K Y: PRE2* N b: PRE3* H V: PUP1*	2.70 Å 27/02/2013
<b>4INT</b>	Proteassoma 20S de levedura em complexo com a vinilsulfona LUI22	K Y: PRE2* N b: PRE3* H V: PUP1*	2.90 Å 27/02/2013
<b>4INU</b>	Proteassoma 20S de levedura em complexo com a vinilsulfona LUI12	K Y: PRE2* N b: PRE3* H V: PUP1*	3.10 Å 27/02/2013
<b>4JSQ</b>	Proteassoma 20S de levedura em complexo com o mimético linear dimerizado de TMC-95A - $\gamma$ CP:4e	N b: $\beta 1$ H V: $\beta 2$ K Y: $\beta 5$	2.80 Å 01/05/2013
<b>4JSU</b>	Proteassoma 20S de levedura em complexo com o mimético linear dimerizado de TMC-95A - $\gamma$ CP:3a	N b: $\beta 1$ H V: $\beta 2$ K Y: $\beta 5$	2.90 Å 01/05/2013
<b>4JT0</b>	Proteassoma 20S de levedura em complexo com o mimético linear dimerizado de TMC-95A - $\gamma$ CP:4a	N b: $\beta 1$ H V: $\beta 2$ K Y: $\beta 5$	3.10 Å 05/06/2013

\*São pró-formas das subunidades  $\beta 1$  (PRE3),  $\beta 2$  (PUP1) e  $\beta 5$  (PRE2).  
 Com excepção do PDB 4BGR – obtido por microscopia eletrónica – as restantes estruturas foram obtidas por cristalografia de raio X).

**Tabela 10** – Exemplos de estruturas cristalográficas do proteassoma.

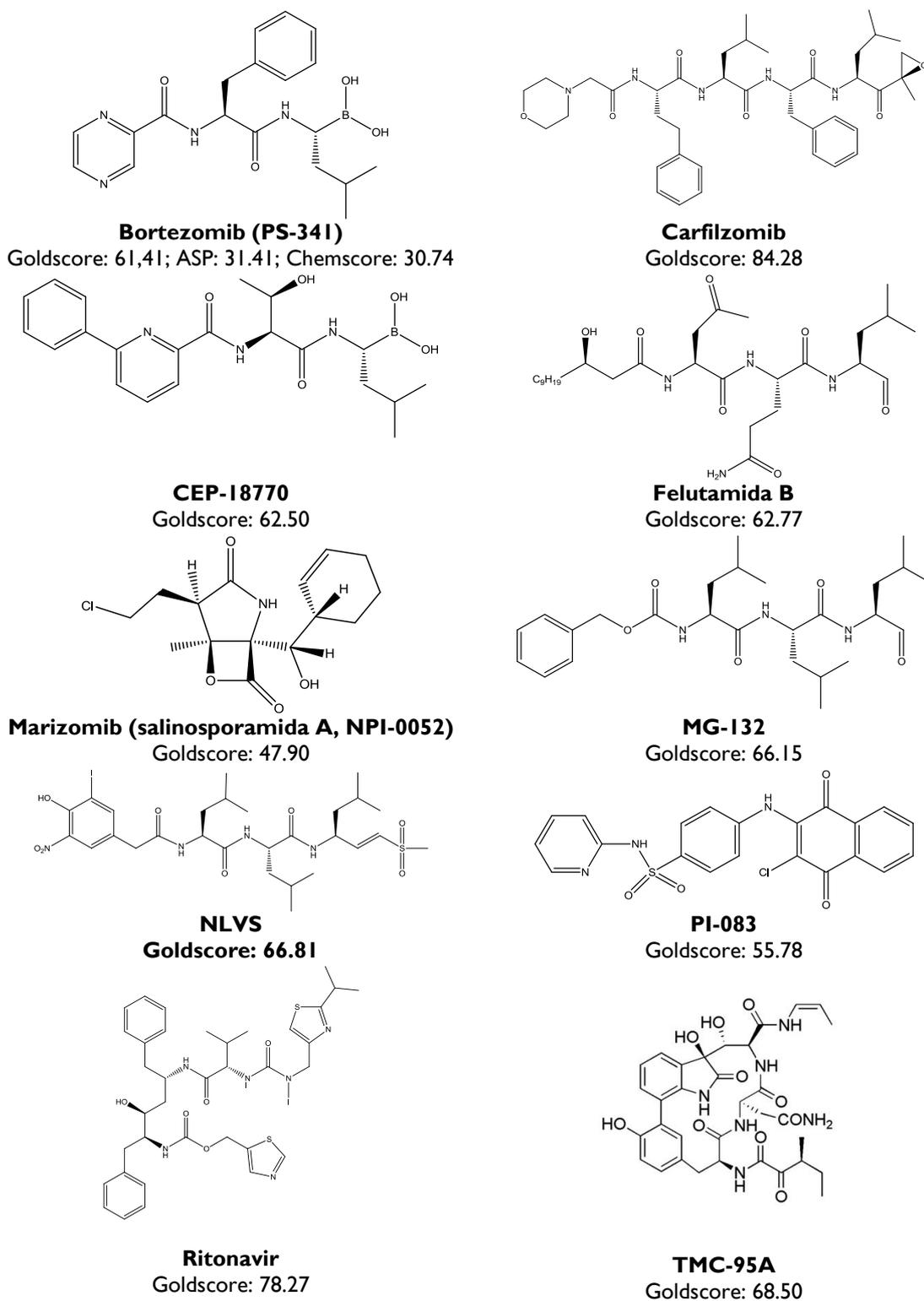
## ANEXO 2

<b>Subunidade</b>	<b>Códigos Uniprot a comparar</b>	<b>Porcentagem de identidade</b>
<b>β1c</b>	<i>Homo sapiens</i> (P20618) e <i>Mus musculus</i> (O09061)	94%
	<i>Homo sapiens</i> (P20618) e <i>S. cerevisiae</i> (P38624)	55%
	<i>Mus musculus</i> (O09061) e <i>S. cerevisiae</i> (P38624)	45%
	<i>Bos taurus</i> (Q2TBX6) e <i>Homo sapiens</i> (P20618)	94%
	<i>Bos taurus</i> (Q2TBX6) e <i>Mus musculus</i> (O09061)	93%
	<i>Bos taurus</i> (Q2TBX6) e <i>S. cerevisiae</i> (P38624)	18%
<b>β2c</b>	<i>Homo sapiens</i> (P49721) e <i>Mus musculus</i> (Q9R1P3)	97%
	<i>Homo sapiens</i> (P49721) e <i>S. cerevisiae</i> (P25043)	19%
	<i>Mus musculus</i> (Q9R1P3) e <i>S. cerevisiae</i> (P25043)	19%
	<i>Bos taurus</i> (Q5E9K0) e <i>Homo sapiens</i> (P49721)	99%
	<i>Bos taurus</i> (Q5E9K0) e <i>Mus musculus</i> (Q9R1P3)	95%
	<i>Bos taurus</i> (Q5E9K0) e <i>S. cerevisiae</i> (P25043)	14%
<b>β3c</b>	<i>Homo sapiens</i> (P49720) e <i>Mus musculus</i> (Q9R1P1)	99%
	<i>Homo sapiens</i> (P49720) e <i>S. cerevisiae</i> (P25451)	55%
	<i>Mus musculus</i> (Q9R1P1) e <i>S. cerevisiae</i> (P25451)	55%
	<i>Bos taurus</i> (P33672) e <i>Homo sapiens</i> (P49720)	99%
	<i>Bos taurus</i> (P33672) e <i>Mus musculus</i> (Q9R1P1)	99%
	<i>Bos taurus</i> (P33672) e <i>S. cerevisiae</i> (P25451)	54%
<b>β4c</b>	<i>Homo sapiens</i> (P28070) e <i>Mus musculus</i> (P99026)	94%
	<i>Homo sapiens</i> (P28070) e <i>S. cerevisiae</i> (P22141)	15%
	<i>Mus musculus</i> (P99026) e <i>S. cerevisiae</i> (P22141)	15%
	<i>Bos taurus</i> (Q3T108) e <i>Homo sapiens</i> (P28070)	95%
	<i>Bos taurus</i> (Q3T108) e <i>Mus musculus</i> (P99026)	93%
	<i>Bos taurus</i> (Q3T108) e <i>S. cerevisiae</i> (P22141)	15%
<b>β5c</b>	<i>Homo sapiens</i> (P28074) e <i>Mus musculus</i> (O55234)	95%
	<i>Homo sapiens</i> (P28074) e <i>S. cerevisiae</i> (P30656)	67%
	<i>Mus musculus</i> (O55234) e <i>S. cerevisiae</i> (P30656)	67%
	<i>Bos taurus</i> (Q32KL2) e <i>Homo sapiens</i> (P28074)	96%
	<i>Bos taurus</i> (Q32KL2) e <i>Mus musculus</i> (O55234)	93%
	<i>Bos taurus</i> (Q32KL2) e <i>S. cerevisiae</i> (P30656)	51%
<b>β6c</b>	<i>Homo sapiens</i> (P28072) e <i>Mus musculus</i> (Q60692)	94%
	<i>Homo sapiens</i> (P28072) e <i>S. cerevisiae</i> (P23724)	50%
	<i>Mus musculus</i> (Q60692) e <i>S. cerevisiae</i> (P23724)	24%
	<i>Bos taurus</i> (Q3MHN0) e <i>Homo sapiens</i> (P28072)	95%
	<i>Bos taurus</i> (Q3MHN0) e <i>Mus musculus</i> (Q60692)	92%
	<i>Bos taurus</i> (Q3MHN0) e <i>S. cerevisiae</i> (P23724)	21%
<b>β7c</b>	<i>Homo sapiens</i> (Q99436) e <i>Mus musculus</i> (P70195)	96%
	<i>Homo sapiens</i> (Q99436) e <i>S. cerevisiae</i> (P30657)	16%
	<i>Mus musculus</i> (P70195) e <i>S. cerevisiae</i> (P30657)	15%
	<i>Bos taurus</i> (Q2TBP0) e <i>Homo sapiens</i> (Q99436)	96%
	<i>Bos taurus</i> (Q2TBP0) e <i>Mus musculus</i> (P70195)	96%
	<i>Bos taurus</i> (Q2TBP0) e <i>S. cerevisiae</i> (P30657)	15%
<b>β1i</b> (nota: no caso do <i>S. cerevisiae</i> é β1c a subunidade utilizada dado que não tem subunidades indutivas)	<i>Homo sapiens</i> (P28065) e <i>Mus musculus</i> (P28076)	89%
	<i>Homo sapiens</i> (P28065) e <i>S. cerevisiae</i> (P38624)	45%
	<i>Mus musculus</i> (P28076) e <i>S. cerevisiae</i> (P38624)	45%
	<i>Bos taurus</i> (Q3SZC2) e <i>Homo sapiens</i> (P28065)	92%
	<i>Bos taurus</i> (Q3SZC2) e <i>Mus musculus</i> (P28076)	89%
	<i>Bos taurus</i> (Q3SZC2) e <i>S. cerevisiae</i> (P38624)	41%

<b>β2i</b> (nota: no caso do <i>S. cerevisiae</i> é β2c a subunidade utilizada dado que não tem subunidades indutivas)	<i>Homo sapiens</i> (P40306) e <i>Mus musculus</i> (O35955)	89%
	<i>Homo sapiens</i> (P40306) e <i>S. cerevisiae</i> (P38624)	47%
	<i>Mus musculus</i> (O35955) e <i>S. cerevisiae</i> (P38624)	56%
	<i>Bos taurus</i> (Q3T0T1) e <i>Homo sapiens</i> (P40306)	87%
	<i>Bos taurus</i> (Q3T0T1) e <i>Mus musculus</i> (P28076)	85%
	<i>Bos taurus</i> (Q3T0T1) e <i>S. cerevisiae</i> (P38624)	45%
<b>β5i</b> (nota: no caso do <i>S. cerevisiae</i> é β5c a subunidade utilizada dado que não tem subunidades indutivas)	<i>Homo sapiens</i> (P28062) e <i>Mus musculus</i> (P28063)	89%
	<i>Homo sapiens</i> (P28062) e <i>S. cerevisiae</i> (P30656)	54%
	<i>Mus musculus</i> (P28063) e <i>S. cerevisiae</i> (P30656)	57%
	<i>Bos taurus</i> (Q3T112) e <i>Homo sapiens</i> (P28062)	89%
	<i>Bos taurus</i> (Q3T112) e <i>Mus musculus</i> (P28063)	84%
	<i>Bos taurus</i> (Q3T112) e <i>S. cerevisiae</i> (P30656)	48%

**Tabela II** – Percentagem de identidade das subunidades do proteassoma/imunoproteassoma entre diversos organismos.

## ANEXO 3

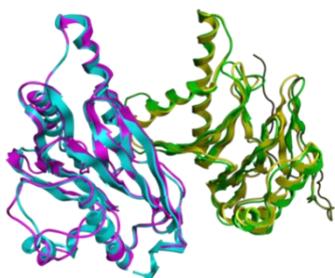


**Figura 40** – Alguns inibidores conhecidos *dockados* da subunidade catalítica  $\beta 5c$  na escala do nM e respectivos scores obtidos.

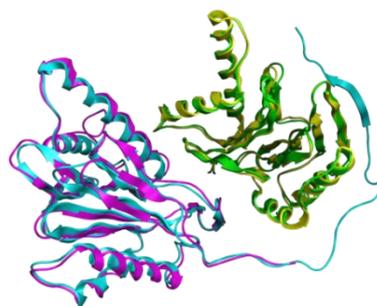
## ANEXO 4

Subunidade catalítica a modelar por homologia	Cadeias de cada subunidade do pocket	Código PDB de partida	Código Uniprot ( <i>Homo sapiens</i> )
$\beta 1c$	$\beta 1c$ – cadeia N	3UN8	P20618
	$\beta 2c$ – cadeia H	3UN8	P49721
$\beta 2c$	$\beta 2c$ – cadeia H	3UN8	P49721
	$\beta 3c$ – cadeia I	3UN8	P49720
$\beta 5c$	$\beta 5c$ – cadeia K	3UN8	P28074
	$\beta 6c$ – cadeia L	3UN8	P28072
$\beta 1i$	$\beta 1c$ – cadeia N	3UN8	P28065
	$\beta 2c$ – cadeia H	3UN8	P40306
$\beta 2i$	$\beta 2c$ – cadeia H	3UN8	P40306
	$\beta 3c$ – cadeia I	3UN8	P49720
$\beta 5i$	$\beta 5c$ – cadeia K	3UN8	P28062
	$\beta 6c$ – cadeia L	3UN8	P28072

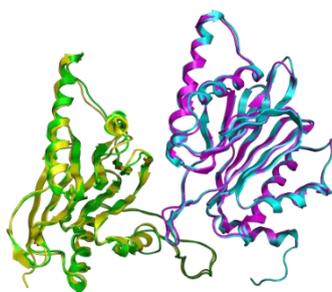
**Tabela 12** – Resumo dos códigos utilizados para efetuar cada modelo de homologia.



$\beta 1c/\beta 2c$



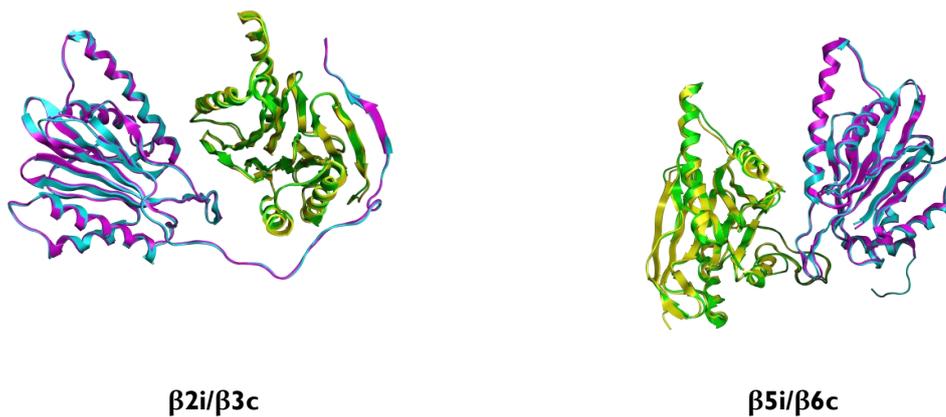
$\beta 2c/\beta 3c$



$\beta 5c/\beta 6c$



$\beta 1i/\beta 2i$



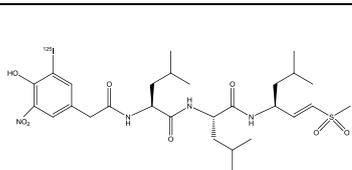
**Figura 41** – Sobreposição dos modelos de homologia dos *pockets* das subunidades catalíticas do proteassoma e do imunoproteassoma com as subunidades de levedura iniciais utilizadas.

(Foram utilizadas as mesmas subunidades de levedura na construção dos modelos constitutivos e indutivos devido ao facto de a levedura não expressar subunidades indutivas. A descrição de cada modelo diz respeito ao modelo de homologia obtido.)

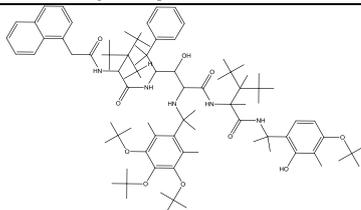
**Legenda de cores:** azul – subunidade catalítica da levedura; rosa – subunidade catalítica homóloga humana; amarelo – subunidade complementar do *pocket* da levedura, verde – subunidade complementar homóloga humana.

# Anexo 5

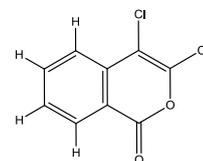
## Base de dados de controlo positivo (compostos com atividade na subunidade $\beta 5c$ )



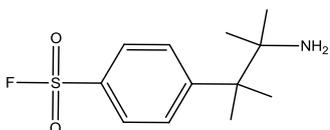
125I-NIP



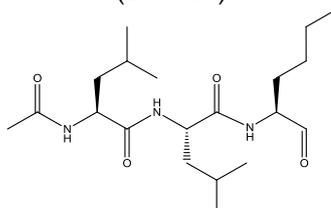
2-aminobenzilastatina  
(derivado)



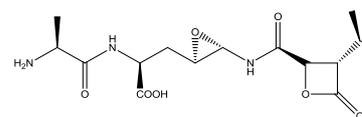
3, 4-dicloroisocumarina



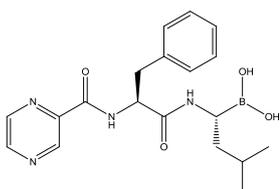
4-(2-aminoethyl)-  
benzenesulfonylfluoride



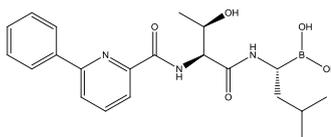
ALLN



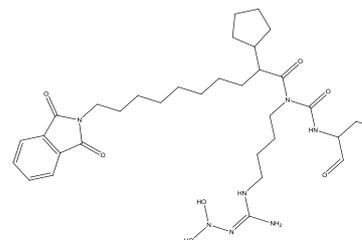
Belactosina A



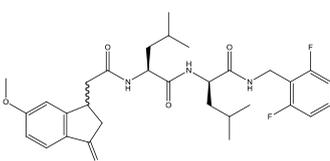
Bortezomib



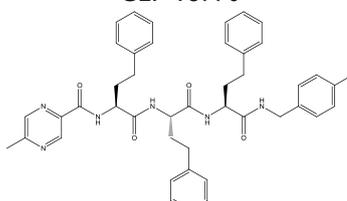
CEP-18770



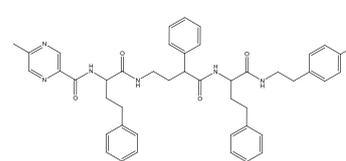
CEP-1612



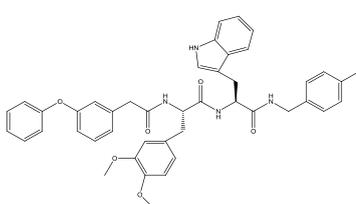
CVT-659



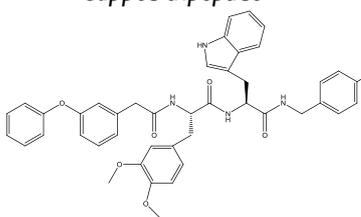
Capped tripeptide\*



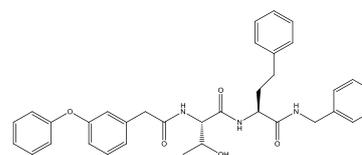
Composto 1\*



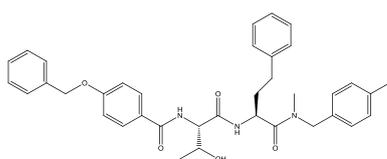
Composto 2\*



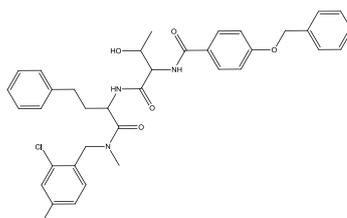
Composto 3\*



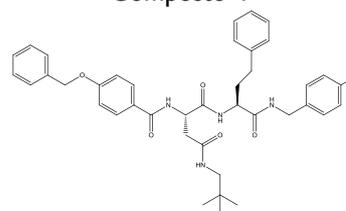
Composto 4\*



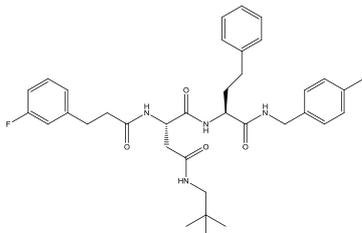
Composto 7\*



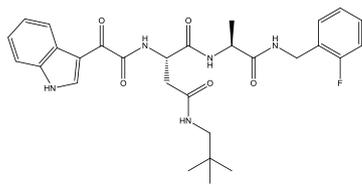
Composto 8\*



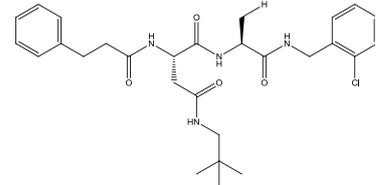
Composto 9\*



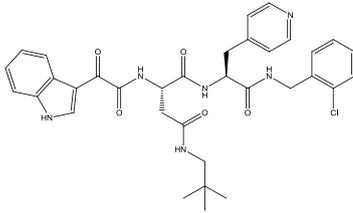
**Composto 10\***



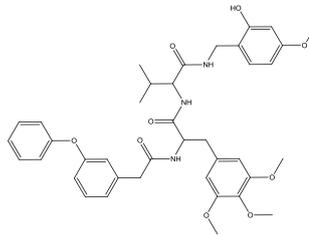
**Composto 11\***



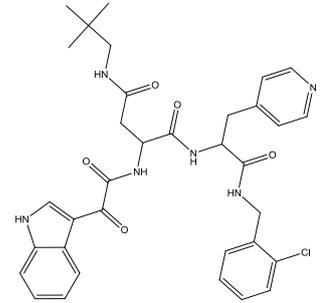
**Composto 12\***



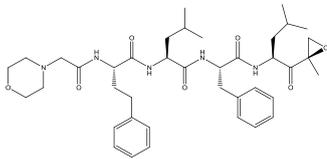
**Composto 16\***



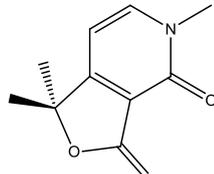
**Capped dipeptide 1\*\***



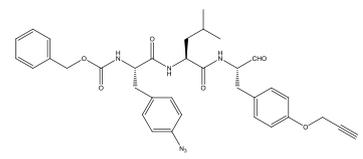
**Capped dipeptide 2\*\***



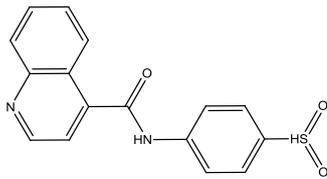
**Carfilzomib**



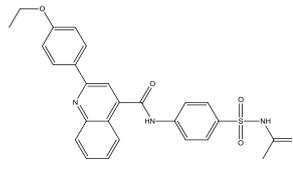
**Cerpegina**



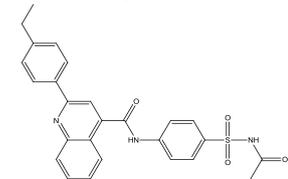
**Composto 6\*\*\***



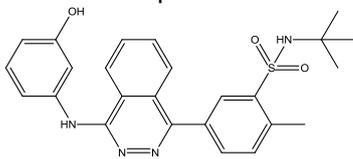
**Composto 3#**



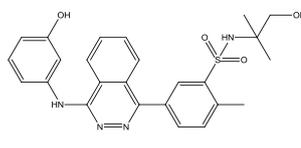
**Composto 3a#**



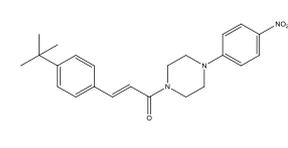
**Composto 3b#**



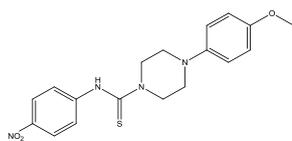
**Composto 4a#**



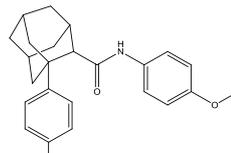
**Composto 4c#**



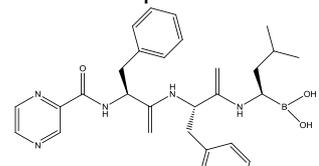
**Composto 6#**



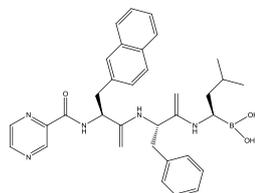
**Composto 7#**



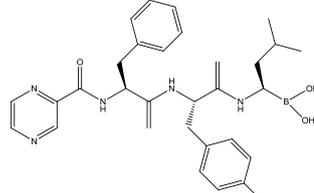
**Composto 8a#**



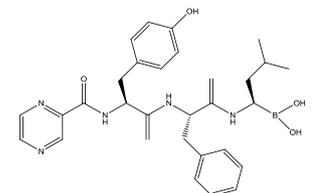
**Composto 1###**



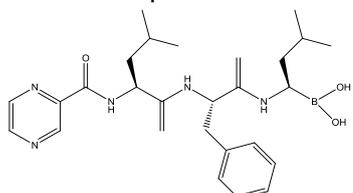
**Composto 2###**



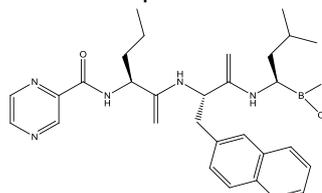
**Composto 3###**



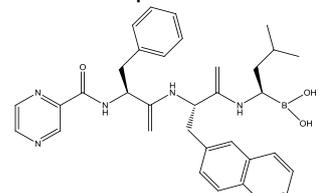
**Composto 4###**



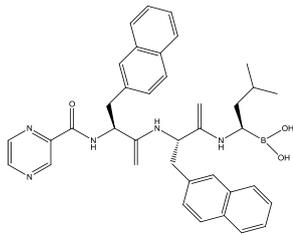
**Composto 5###**



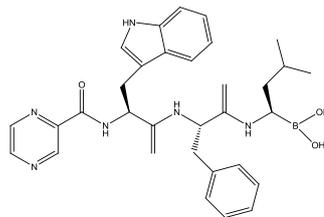
**Composto 6###**



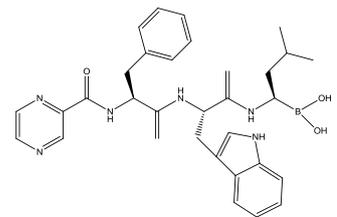
**Composto 7###**



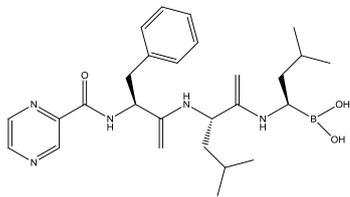
Composto 8##



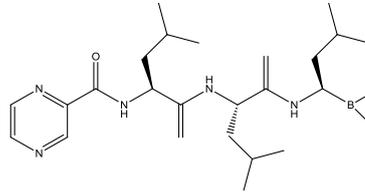
Composto 9##



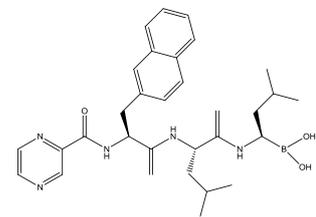
Composto 10##



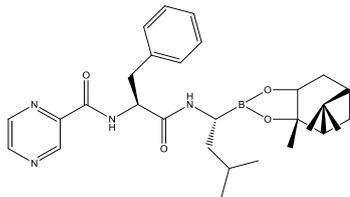
Composto 11##



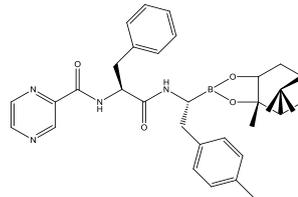
Composto 12##



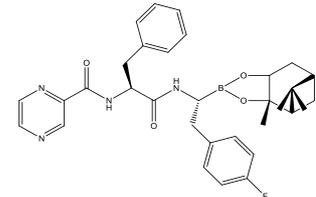
Composto 13##



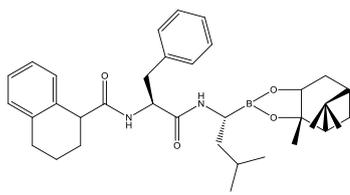
Composto 12◇



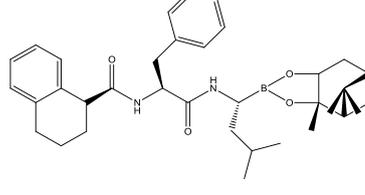
Composto 13◇



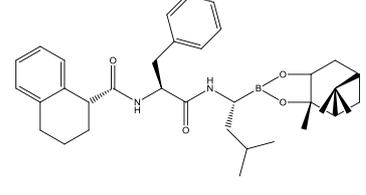
Composto 14◇



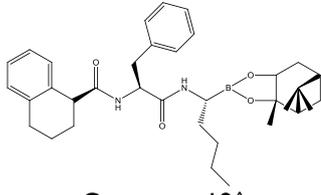
Composto 15◇



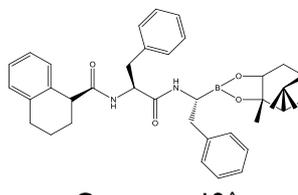
Composto 16◇



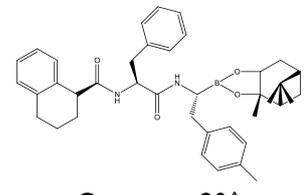
Composto 17◇



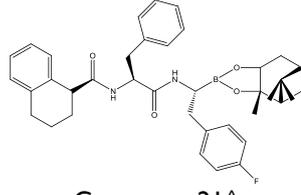
Composto 18◇



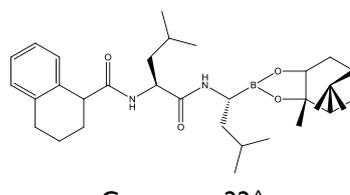
Composto 19◇



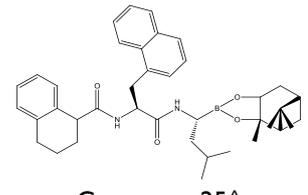
Composto 20◇



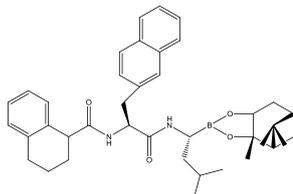
Composto 21◇



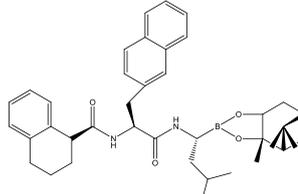
Composto 22◇



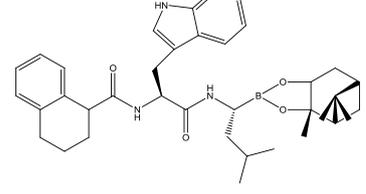
Composto 25◇



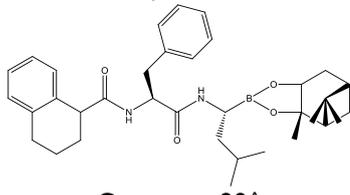
Composto 26◇



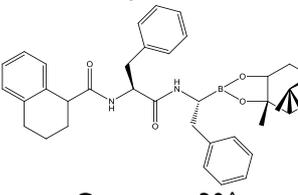
Composto 27◇



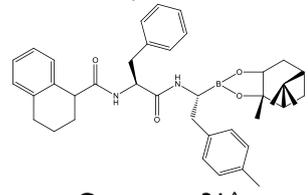
Composto 28◇



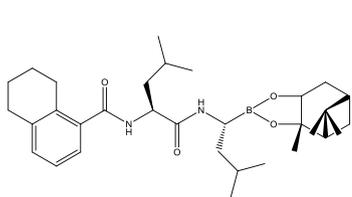
Composto 29◇



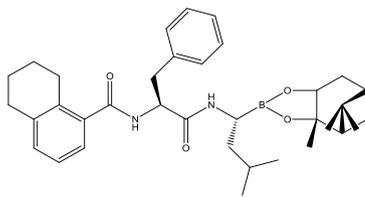
Composto 30◇



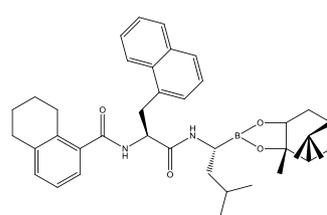
Composto 31◇



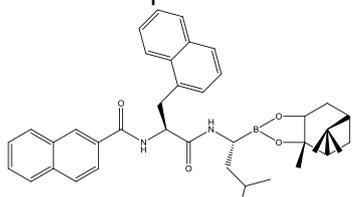
Composto 32



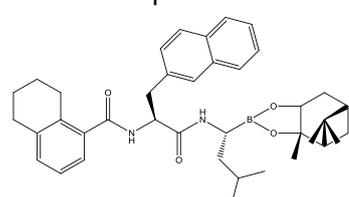
Composto 34



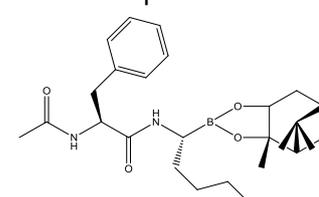
Composto 35



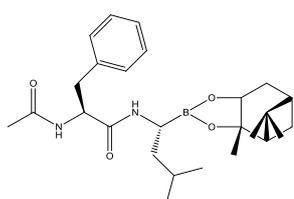
Composto 36



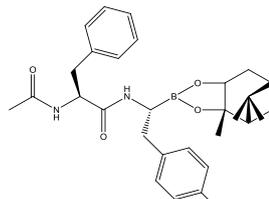
Composto 37



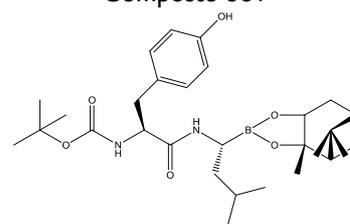
Composto 38



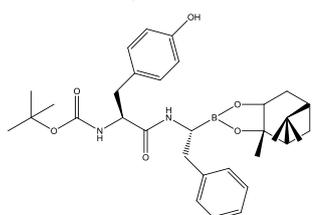
Composto 39



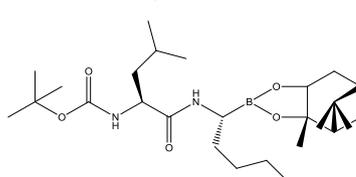
Composto 40



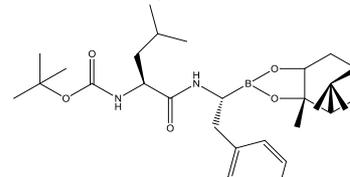
Composto 41



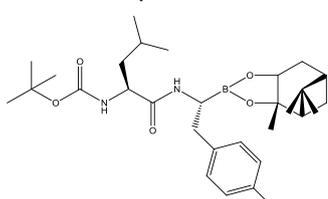
Composto 42



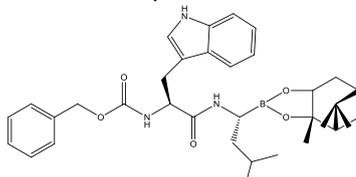
Composto 43



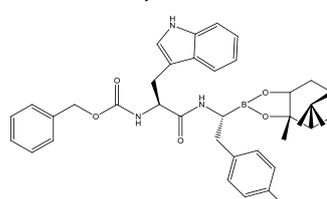
Composto 44



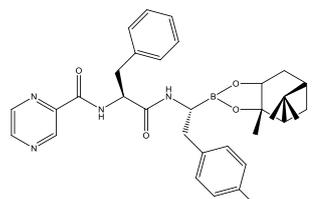
Composto 45



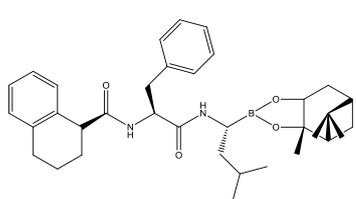
Composto 46



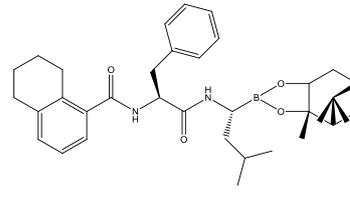
Composto 47



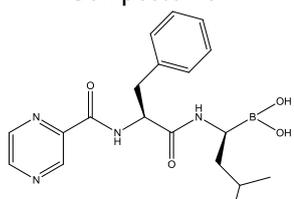
Composto 48



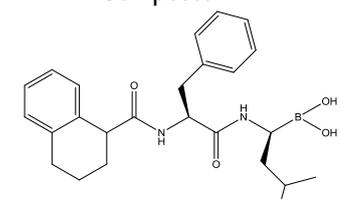
Composto 49



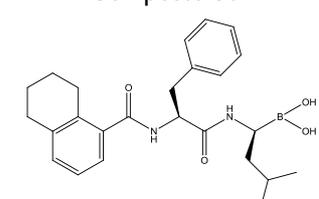
Composto 50



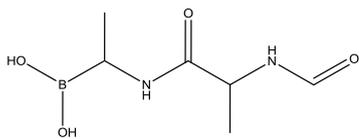
Composto 51



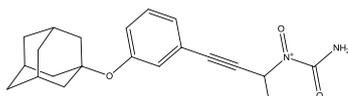
Composto 52



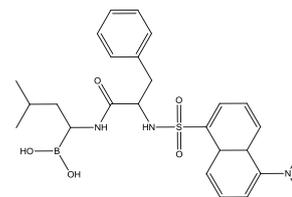
Composto 53



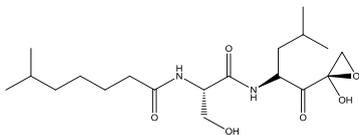
**Composto modelo**



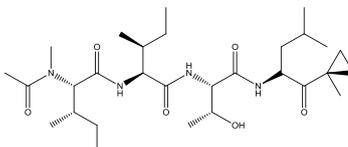
**Derivado HU10**



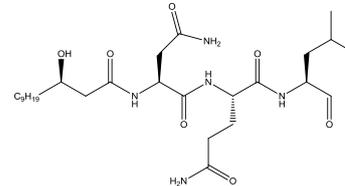
**DFLB**



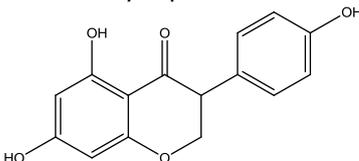
**Dihydroponemycin**



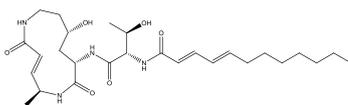
**Epoxomicina**



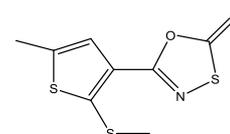
**Felutamida B**



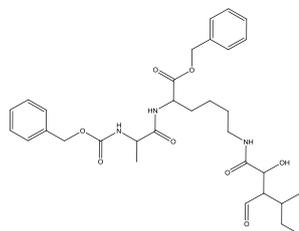
**Genisteína**



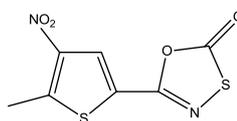
**Glidobactina A**



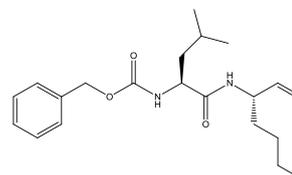
**GL5**



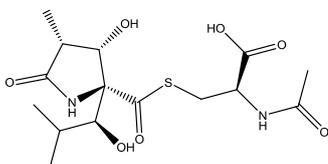
**Homobelactosina**



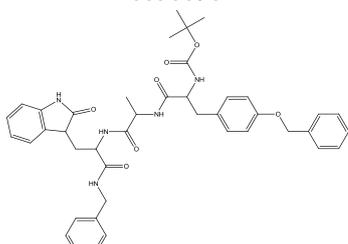
**HT1171**



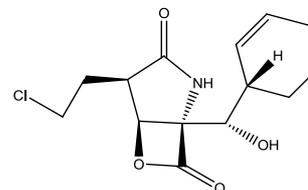
**IPSI**



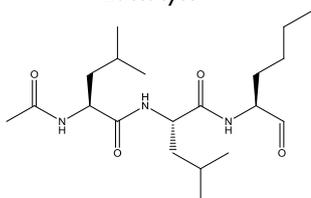
**Lactacystin**



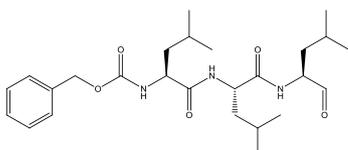
**Ligando a (PDB 3NZJ)**



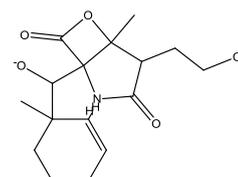
**Marizomib**



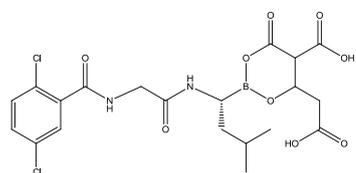
**MG-101**



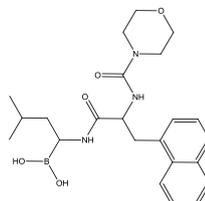
**MG-131**



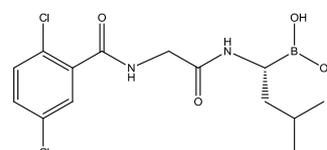
**ML858**



**ML9708**



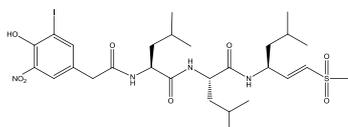
**MLN273**



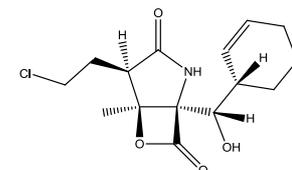
**MLN2238**



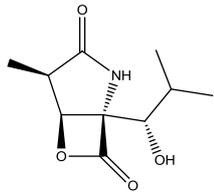
**MVI51**



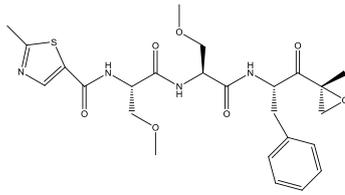
**NLVS**



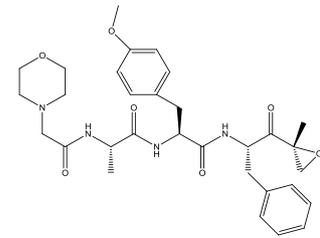
**NPI005**



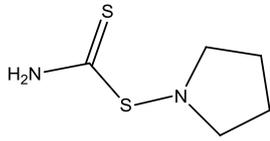
**Omuralide**



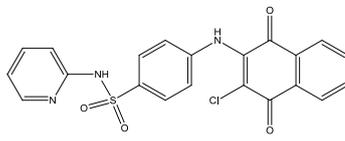
**ONX0912**



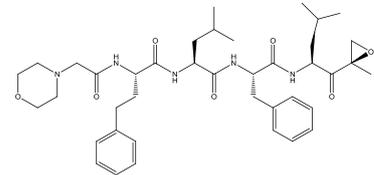
**ONX0914**



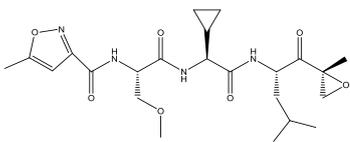
**PDTc**



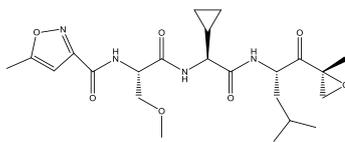
**PI-083**



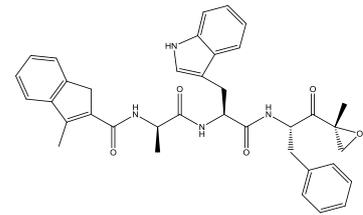
**PR-171**



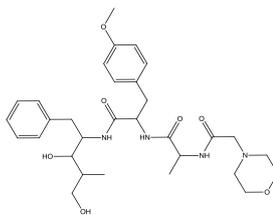
**PR-825**



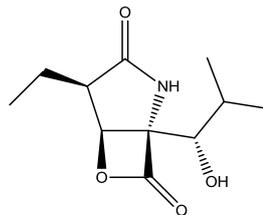
**PR-893**



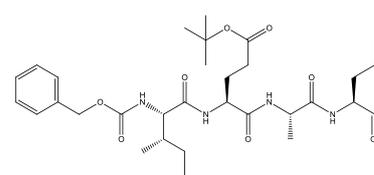
**PR-924**



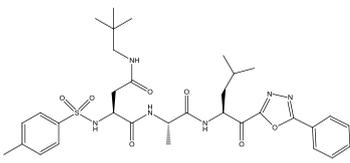
**PR-957**



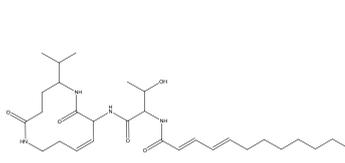
**PS-519**



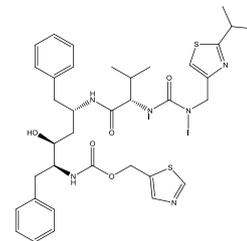
**PSI**



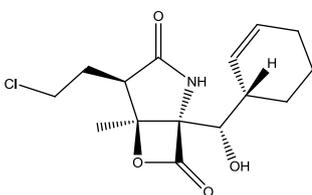
**Peptide-2-keto-1,3,4-oxadizole**



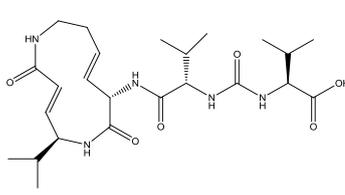
**Quimera SyLA-gliodobactina**



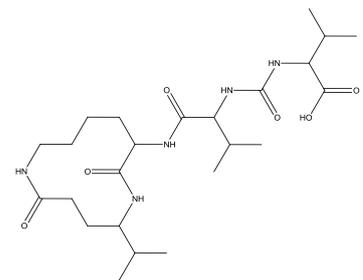
**Ritonavir**



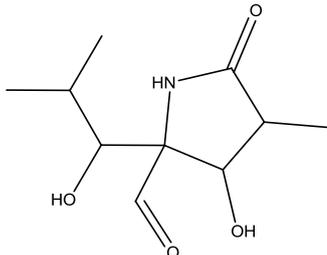
**Salinosporamida**



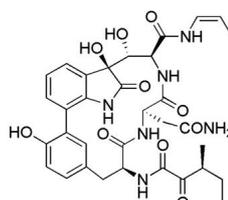
**Syringolin A (SylA)**



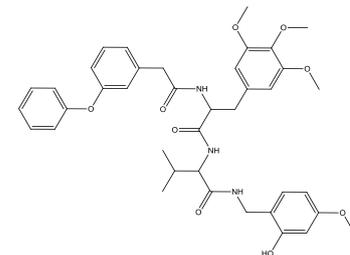
**Syringolin B (SylB)**



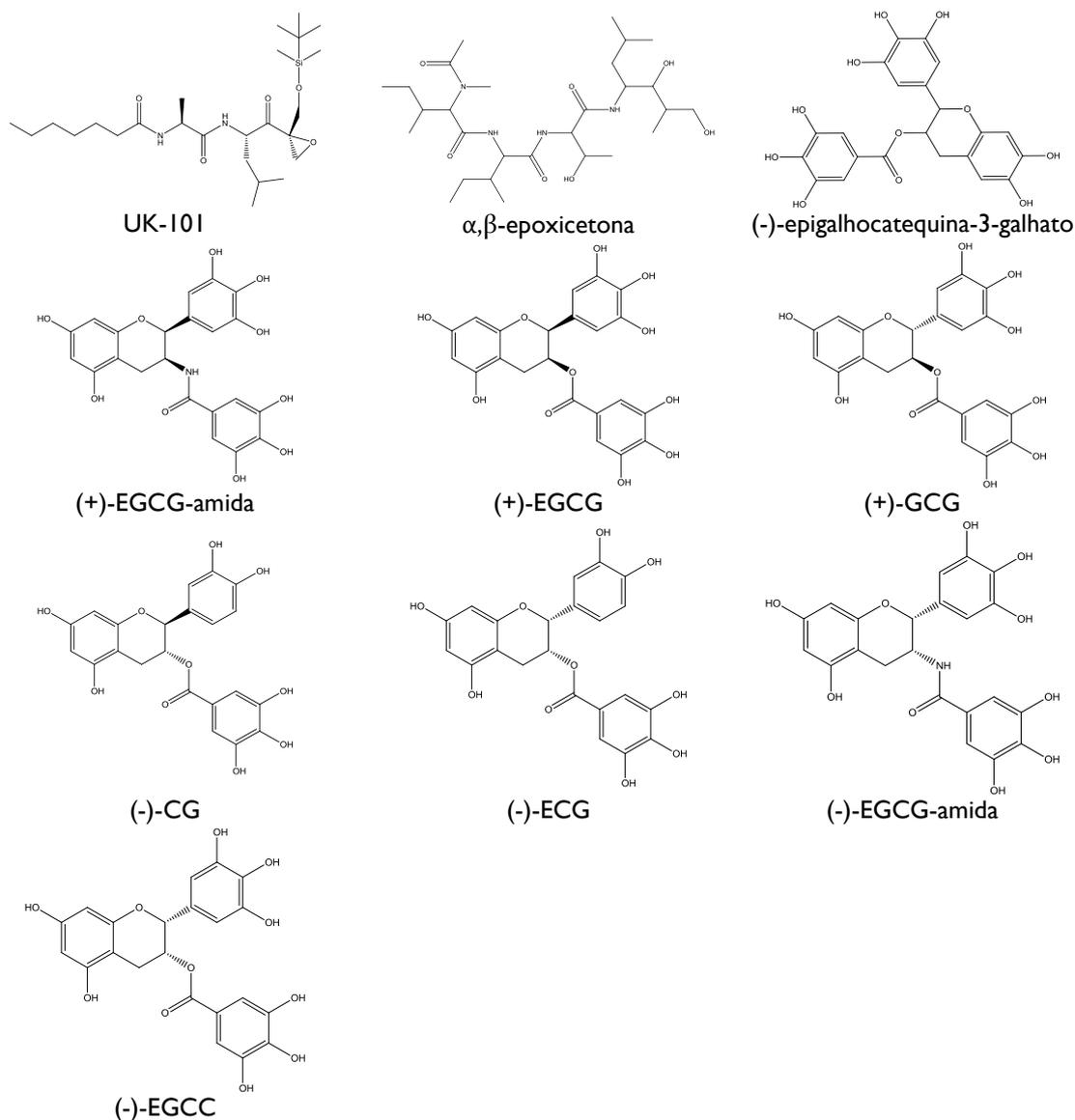
**Spirolactactystin**



**TMC95A**



**Trimetoxi L-fenilalanina**



**Figura 42** – Base de dados de controlo positivo utilizada na etapa de modelação farmacofórica.

\* BLACKBURN, C. [et al.] – **Characterization of a new series of non-covalent proteasome inhibitors with exquisite potency and selectivity for the 20S beta5-subunit.** The Biochemical journal. 430 (2010) 461–76.

\*\* KISSELEV, A. F. ; LINDEN, W. A. V. D. ; OVERKLEEF, H. S. – **Proteasome inhibitors: an expanding army attacking a unique target.** Chemistry & biology. 19 (2012) 99–115.

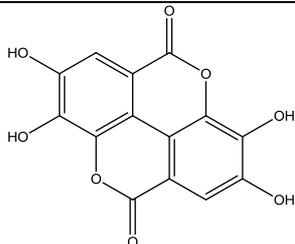
\*\*\* NEILSEN, P. M. [et al.] – **New 26S proteasome inhibitors with high selectivity for chymotrypsin-like activity and p53-dependent cytotoxicity.** ACS chemical biology. 8 (2013) 353–359.

# BASSE, N. [et al.] – **Novel organic proteasome inhibitors identified by virtual and in vitro screening.** Journal of medicinal chemistry. 53 (2010) 509–13.

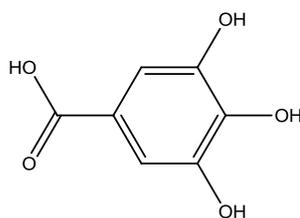
## Lei, M. [et al.] – **Progress of computer-aided drug design (CADD) of proteasome inhibitors.** Current topics in medicinal chemistry. 11 (2011) 2931–2944.

◇ ZHU, Y. [et al.] – **Design, synthesis, biological evaluation, and structure-activity relationship (SAR) discussion of dipeptidyl boronate proteasome inhibitors, part I: comprehensive**

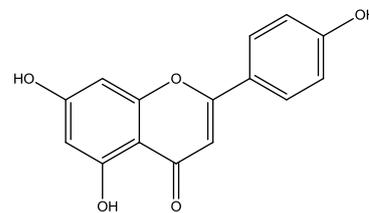
**Base de dados de controlo negativo (compostos sem atividade na subunidade  $\beta 5c$ )**



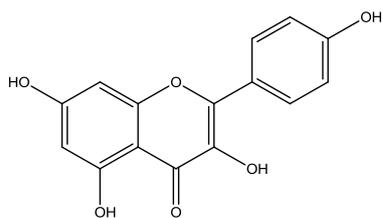
Ácido elágico



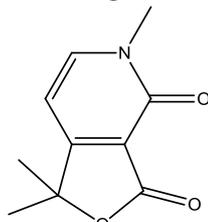
Ácido gálico



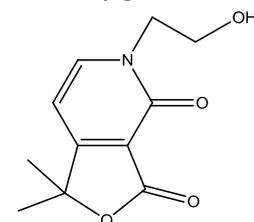
Apigenina



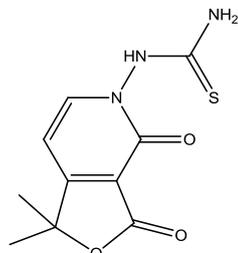
Canferol



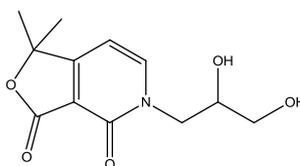
Composto 1•



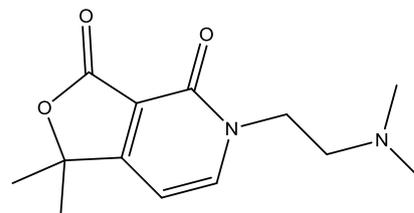
Composto 2•



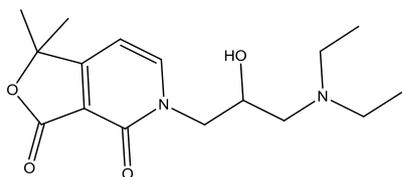
Composto 3•



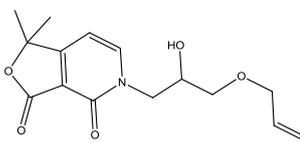
Composto 4•



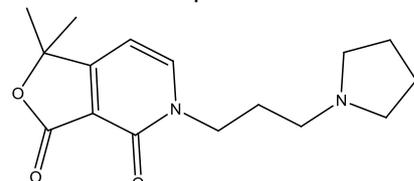
Composto 5•



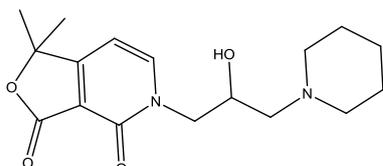
Composto 6•



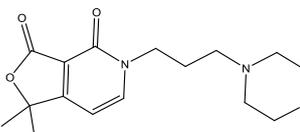
Composto 7•



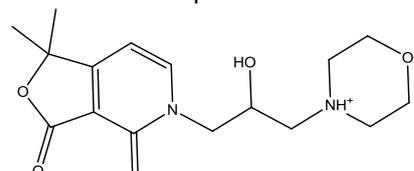
Composto 8•



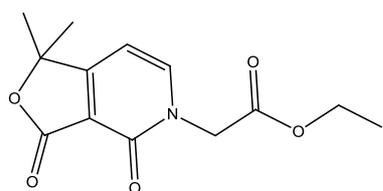
Composto 9•



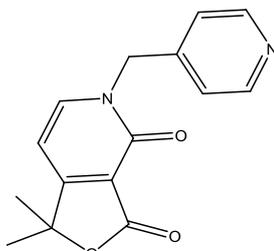
Composto 10•



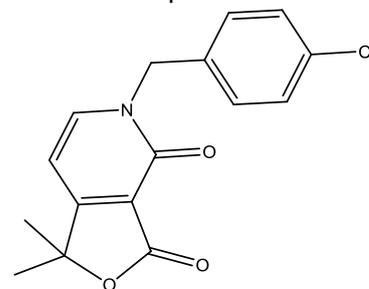
Composto 11•



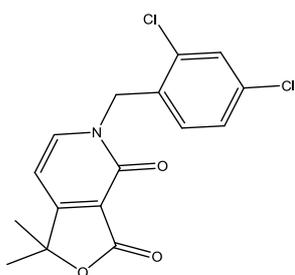
Composto 12•



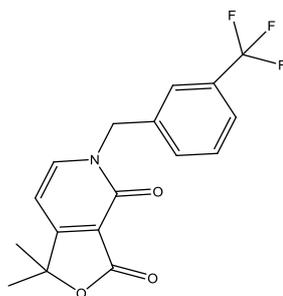
Composto 13•



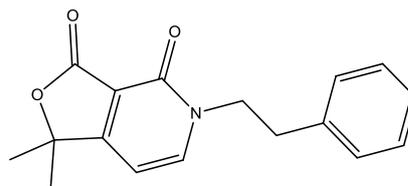
Composto 14•



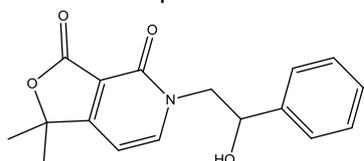
Composto 15•



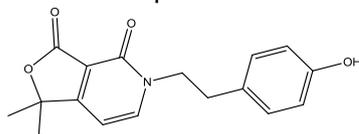
Composto 16•



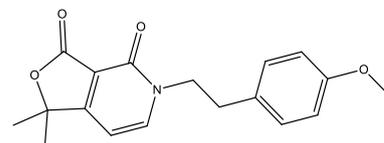
Composto 17•



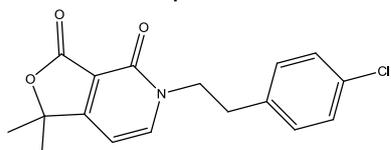
Composto 18•



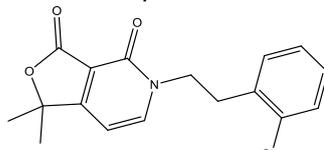
Composto 19•



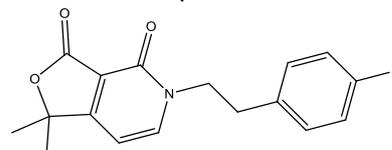
Composto 20•



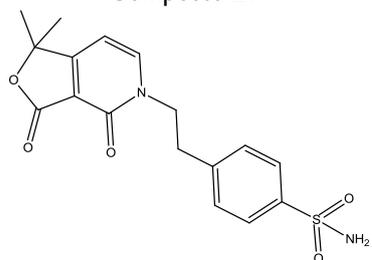
Composto 21•



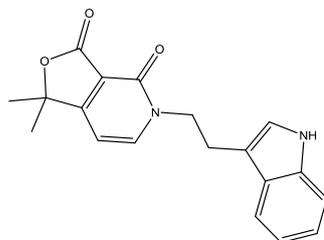
Composto 22•



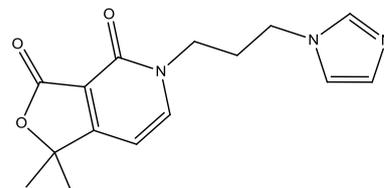
Composto 23•



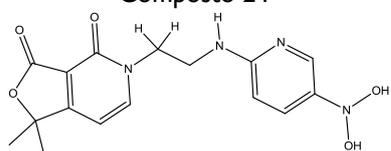
Composto 24•



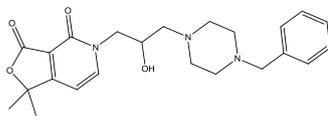
Composto 25•



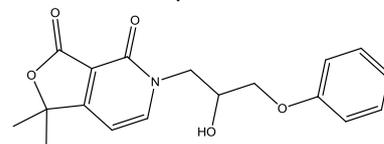
Composto 26•



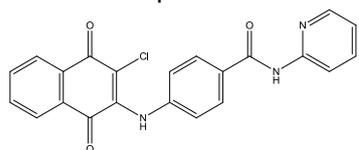
Composto 27•



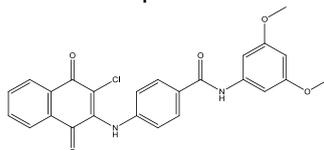
Composto 28•



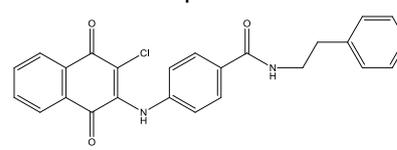
Composto 29•



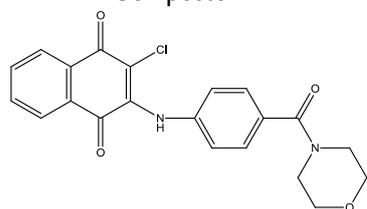
Composto 1Δ



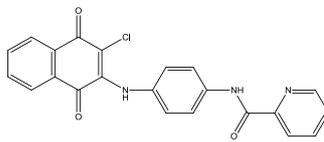
Composto 2Δ



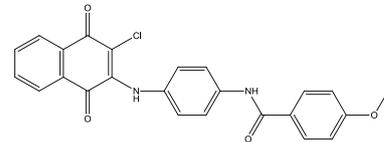
Composto 3Δ



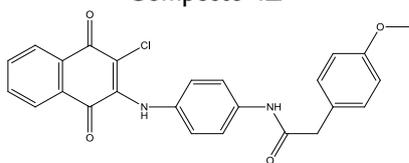
Composto 4Δ



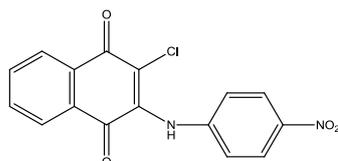
Composto 5Δ



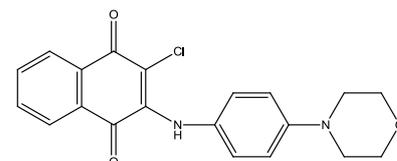
Composto 6Δ



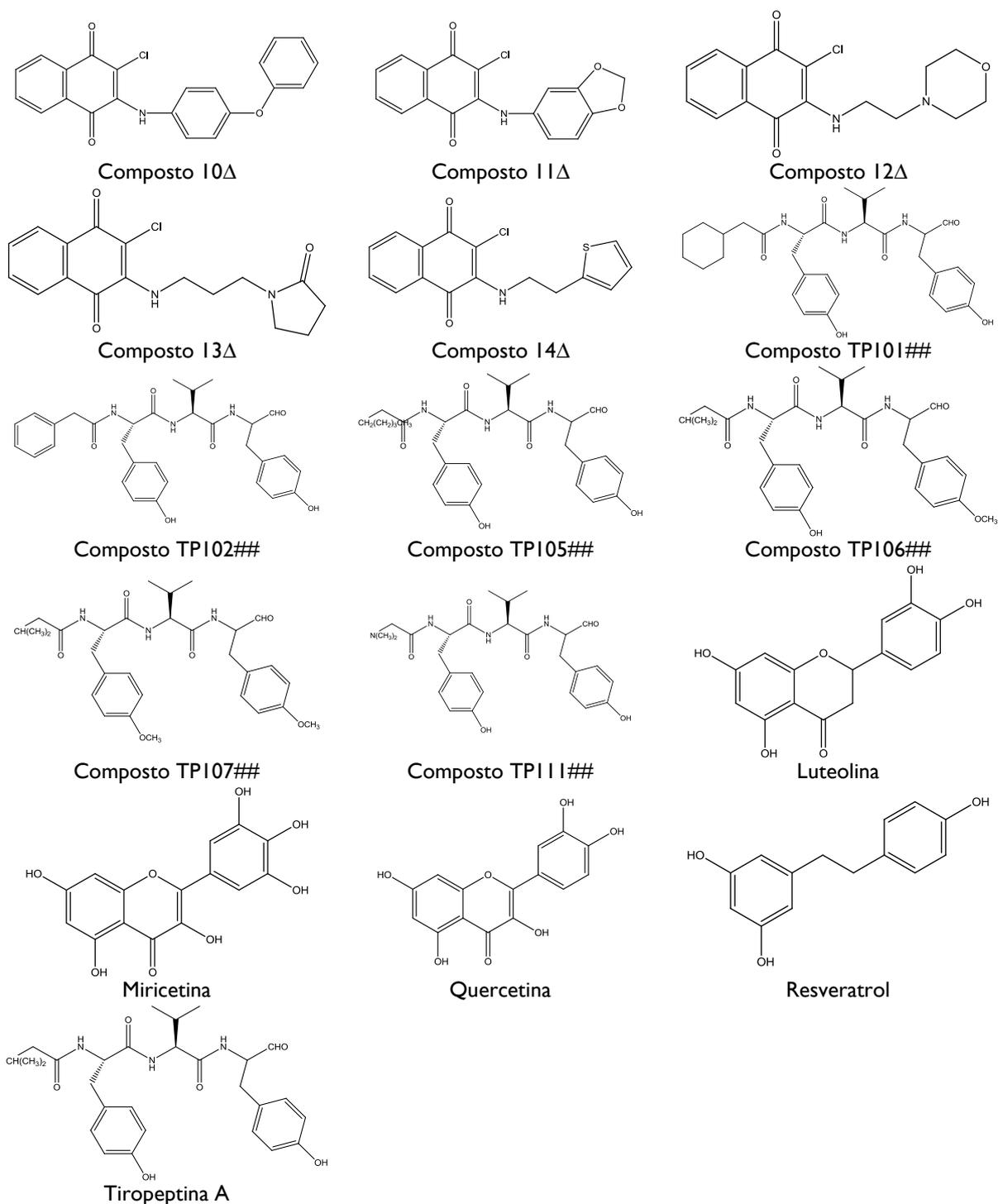
Composto 7Δ



Composto 8Δ



Composto 9Δ



**Figura 43** – Base de dados de controlo negativo utilizada na etapa de modelação farmacofórica.

- PHAM, T. H. [et al.] – **A new series of N5 derivatives of the 1,1,5-trimethyl furo[3,4-c]pyridine-3,4-dione (cerpegin) selectively inhibits the post-acid activity of mammalian 20S proteasomes.** *Bioorganic & medicinal chemistry letters.* 22 (2012) 3822–3827.
- $\Delta$  XU, K. [et al.] – **Design and synthesis of naphthoquinone derivatives as antiproliferative agents and 20S proteasome inhibitors.** *Bioorganic & Medicinal Chemistry Letters* 22 (2012) 2772–2774.
- ## Lei, M. [et al.] – **Progress of computer-aided drug design (CADD) of proteasome inhibitors.** *Current topics in medicinal chemistry.* 11 (2011) 2931–2944.

## Anexo 6

	<b>Epoxicetonas com ênfase em características Aro/Acc/Don, esquema unified</b> F1 Acc (raio 1.4); F2 e F3 Hyd (raio 1.4); F4 Don (raio 1)	<b>Vinilsulfonas sem ênfase em características Aro/Acc/Don, esquema unified</b> F1, F2 e F3 Hyd (raio 1.4); F4 Acc2 (raio 1)
<b>Aumentar o raio de todas as características do farmacóforo</b>	F1, F2, F3 (1.4 → 2) e F4 (1 → 1.5)  124 ativos e 29 inativos (87%/49%)	F1, F3 (1.4 → 1.5), F2 (1.4 → 2.0), F4 (1 → 1.4)  111 ativos e 20 inativos (78%/34%)  Nota: em F1 e F3 os raios só podem ser aumentados até 1.5 dado que estas características se encontram muito próximas.
<b>Aumentar o raio das características individualmente e mantendo as restantes características constantes</b>	F1 (1.4 → 2): 110 ativos, 11 inativos (77%/19%) F2 (1.4 → 2): 103 ativos, 15 inativos (73%/25%) F3 (1.4 → 2): 103 ativos, 11 inativos (73%/19%) F4 (1 → 1.4): 105 ativos, 13 inativos (74%/22%) F4 (1 → 2): 113 ativos, 17 inativos (80%/29%)	F1 (1.4 → 1.5): 97 ativos, 16 inativos (68%/27%) F2 (1.4 → 2): 106 ativos, 16 inativos (75%/27%) F3 (1.4 → 1.5): 97 ativos, 15 inativos (68%/25%) F4 (1 → 1.4): 101 ativos, 16 inativos (71%/27%) F4 (1 → 2): 111 ativos, 29 inativos (78%/49%)
<b>Aumentar o raio das características hidrofóbicas em simultâneo</b>	F2 (2), F3 (2) → 108 ativos, 21 inativos (76%, 36%)	F1 (1.5), F2 (2), F3 (1.5) → 108 ativos, 18 inativos (76%, 31%)
<b>Aumentar progressivamente o raio de algumas características, mantendo as restantes constantes</b>	F1 (1.4 → 1.5): 98 ativos, 8 inativos (69%/14%) <b>F1 (1.4 → 1.6): 102 ativos, 8 inativos (72%/14%)</b> F1 (1.4 → 1.7): 103 ativos, 9 inativos (73%/15%) F1 (1.4 → 1.8): 104 ativos, 9 inativos (73%/15%) F1 (1.4 → 1.9): 107 ativos, 10 inativos (75%/17%) F1 (1.4 → 2.0): 110 ativos, 11 inativos (77%/19%)  F2 (1.4 → 1.5): 75 ativos, 6 inativos (52%/10%) F2 (1.4 → 1.6): 75 ativos, 6 inativos (52%/10%) <b>F2 (1.4 → 1.7): 79 ativos, 6 inativos (56%/10%)</b> F2 (1.4 → 1.8): 79 ativos, 6 inativos (56%/10%) F2 (1.4 → 1.9): 79 ativos, 7 inativos (56%/12%) F2 (1.4 → 2.0): 81 ativos, 7 inativos (57%/12%)  F3 (1.4 → 1.5): 98 ativos, 8 inativos (69%/14%) <b>F3 (1.4 → 1.6): 99 ativos, 8 inativos (70%/14%)</b> F3 (1.4 → 1.7): 100 ativos, 9 inativos (70%/15%) F3 (1.4 → 1.8): 101 ativos, 11 inativos (71%/19%) F3 (1.4 → 1.9): 102 ativos, 11 inativos (72%/19%) F3 (1.4 → 2.0): 103 ativos, 11 inativos (73%/11%)  F4 (1 → 1.1): 102 ativos, 9 inativos (72%/15%) <b>F4 (1 → 1.2): 103 ativos, 10 inativos (73%/17%)</b> F4 (1 → 1.3): 104 ativos, 10 inativos (73%/17%) <b>F4 (1 → 1.4): 105 ativos, 13 inativos (74%/22%)</b> F4 (1 → 1.5): 105 ativos, 13 inativos (74%/22%) <b>F4 (1 → 1.6): 106 ativos, 13 inativos (75%/22%)</b> F4 (1 → 1.7): 110 ativos, 14 inativos (77%/24%) F4 (1 → 1.8): 110 ativos, 16 inativos (77%/27%) F4 (1 → 1.9): 113 ativos, 17 inativos (80%/29%) F4 (1 → 2.0): 113 ativos, 17 inativos (80%/29%)	<b>F1 (1.4 → 1.5): 97 ativos, 16 inativos (68%/27%)</b>  F2 (1.4 → 1.5): 96 ativos, 15 inativos (68%/25%) F2 (1.4 → 1.6): 98 ativos, 15 inativos (69%/25%) F2 (1.4 → 1.7): 99 ativos, 15 inativos (70%/25%) F2 (1.4 → 1.8): 100 ativos, 15 inativos (70%/25%) <b>F2 (1.4 → 1.9): 104 ativos, 15 inativos (73%/25%)</b> F2 (1.4 → 2.0): 106 ativos, 16 inativos (75%/27%)  <b>F3 (1.4 → 1.5): 97 ativos, 15 inativos (68%/25%)</b>  F4 (1 → 1.1): 98 ativos, 15 inativos (69%/25%) F4 (1 → 1.2): 98 ativos, 16 inativos (69%/27%) F4 (1 → 1.3): 101 ativos, 16 inativos (71%/27%) F4 (1 → 1.4): 101 ativos, 16 inativos (71%/27%) F4 (1 → 1.5): 102 ativos, 16 inativos (72%/27%) <b>F4 (1 → 1.6): 104 ativos, 16 inativos (73%/27%)</b> F4 (1 → 1.7): 107 ativos, 22 inativos (75%/37%) F4 (1 → 1.8): 108 ativos, 22 inativos (76%/37%) F4 (1 → 1.9): 110 ativos, 25 inativos (77%/42%) F4 (1 → 2.0): 111 ativos, 29 inativos (78%/49%)  F1 (1.5) + F2 (1.9) + F3 (1.5) + F4 (1.6) → 112 ativos, 21 inativos (79%/36%)

	<p>FI (1.6) + F4 (1.2) → 107 ativos, 10 inativos (75%/17%)</p> <p>FI (1.6) + F4 (1.4) → 109 ativos, 14 inativos (77%/24%)</p> <p>FI (1.6) + F4 (1.6) → 111 ativos, 14 inativos (78%/24%)</p>	
<b>Variar as características essenciais</b>	<p>FI, F2, F4 → 129 ativos, 43 inativos (91%/73%)</p> <p>FI, F3, F4 → 132 ativos, 52 inativos (93%/88%)</p> <p>FI, F2 → 142 ativos, 58 inativos (100%/100%)</p> <p>FI, F3 → 139 ativos, 57 inativos (98%/97%)</p> <p>F2, F4 → 137 ativos, 47 inativos (96%/80%)</p> <p>F3, F4 → 137 ativos, 47 inativos (96%/80%)</p> <p>FI → 142 ativos, 58 inativos (100%/100%)</p> <p>F2 → 142 ativos, 58 inativos (100%/100%)</p> <p>F3 → 142 ativos, 58 inativos (100%/100%)</p> <p>F4 → 142 ativos, 58 inativos (100%/100%)</p>	<p>FI, F2, F3 → 136 ativos, 52 inativos (96%/88%)</p> <p>FI, F2 → 139 ativos, 57 inativos (98%/97%)</p> <p>FI, F3 → 141 ativos, 55 inativos (99%/93%)</p> <p>FI, F4 → 136 ativos, 58 inativos (96%/88%)</p> <p>F2, F4 → 132 ativos, 57 inativos (93%/97%)</p> <p>F3, F4 → 136 ativos, 58 inativos (96%/100%)</p> <p>FI → 142 ativos, 58 inativos (100%/100%)</p> <p>F2 → 142 ativos, 58 inativos (100%/100%)</p> <p>F3 → 142 ativos, 58 inativos (100%/100%)</p> <p>F4 → 136 ativos, 58 inativos (96%/100%)</p>
<b>Transformar as características hidrofóbicas em aceitadoras</b>	<p>F2 → 72 ativos, 8 inativos (51%/14%)</p> <p>F3 → 76 ativos, 8 inativos (54%/14%)</p>	<p>FI → 87 ativos, 13 inativos (61%/22%)</p> <p>F2 → 79 ativos, 10 inativos (56%/17%)</p> <p><b>F3 → 88 ativos, 12 inativos (62%/20%)</b></p>
<b>Transformar as características hidrofóbicas em doadoras</b>	<p>F2 → 64 ativos, 12 inativos (45%/12%)</p> <p>F3 → 56 ativos, 6 inativos (39%/10%)</p>	<p>FI → 73 ativos, 16 inativos (51%/27%)</p> <p>F2 → 60 ativos, 9 inativos (42%/15%)</p> <p><b>F3 → 94 ativos, 12 inativos (66%/12%)</b></p>
<b>Transformar as características hidrofóbicas em aromáticas</b>	<p>F2 → 37 ativos, 5 inativos (26%/8%)</p> <p>F3 → 51 ativos, 4 inativos (36%/7%)</p> <p>F2 e F3 → 19 ativos, 0 inativos (13%/0%)</p>	<p>FI → 53 ativos, 5 inativos (37%/8%)</p> <p><b>F2 → 58 ativos, 4 inativos (41%/7%)</b></p> <p>F3 → 41 ativos, 18 inativos (29%/31%)</p>
<b>Transformar as características aceitadoras em doadoras e as características doadoras em aceitadoras</b>	<p>FI passa a doador → 67 ativos, 5 inativos (47%/8%)</p> <p><b>F4 (r=1) passa a aceitador → 102 ativos, 7 inativos (72%/12%)</b></p> <p style="text-align: center;">↓</p> <p>F4 (1.0 → 1.1): 101 ativos, 9 inativos (71%/15%)</p> <p>F4 (1.0 → 1.2): 101 ativos, 11 inativos (71%/19%)</p> <p>F4 (1.0 → 1.3): 105 ativos, 12 inativos (74%/20%)</p> <p>F4 (1.0 → 1.4): 105 ativos, 15 inativos (74%/25%)</p> <p>F4 (1.0 → 1.5): 105 ativos, 17 inativos (74%/29%)</p> <p>F4 (1.0 → 1.6): 111 ativos, 17 inativos (78%/29%)</p> <p>F4 (1.0 → 1.7): 115 ativos, 18 inativos (81%/31%)</p> <p>F4 (1.0 → 1.8): 118 ativos, 20 inativos (83%/34%)</p> <p>F4 (1.0 → 1.9): 117 ativos, 21 inativos (82%/36%)</p> <p>F4 (1.0 → 2.0): 118 ativos, 22 inativos (83%/37%)</p>	<p>F4 (Acc2) passa a Don → 50 ativos, 6 inativos (35%/10%)</p> <p>F4 (Acc2) passa a Don2 → 64 ativos, 10 inativos (45%/17%)</p> <p><b>F4 (Acc2) passa a Acc → 68 ativos, 6 inativos (48%/10%)</b></p>

Nota: raio em Å; F = feature (característica farmacofórica)

**Tabela 13** – Refinamento das características farmacofóricas dos melhores modelos farmacofóricos iniciais obtidos com base na estrutura de epoxicetonas e vinilsulfonas e respetivos resultados na deteção de compostos das bases de dados do controlo positivo e negativo.

(A negrito encontram-se assinalados os melhores MF obtidos em cada alteração efetuada.)

**Alterações efectuadas ao farmacóforo com base na estrutura do recetor**

<b>Adicionar volumes de exclusão ao farmacóforo 6I ativos, 4 inativos (43%/7%)</b>	109 ativos e 52 inativos (77%/88%)
<b>Ignorar uma característica de cada vez no farmacóforo 6I ativos, 4 inativos (43%/7%)</b>	F1: 120 ativos, 38 inativos (85%/64%) F2: 121 ativos, 23 inativos (85%/40%) F3: 127 ativos, 21 inativos (89%/36%) F4: 117 ativos, 18 inativos (82%/31%) F5: 107 ativos, 36 inativos (75%/61%)
<b>Variar as características essenciais no farmacóforo 6I ativos, 4 inativos (43%/7%)</b>	F1 essencial: 142 ativos, 59 inativos (100%/100%) F2 essencial: 142 ativos, 59 inativos (100%/100%) F3 essencial: 142 ativos, 59 inativos (100%/100%) F4 essencial: 142 ativos, 59 inativos (100%/100%) F5 essencial: 142 ativos, 59 inativos (100%/100%)  F1, F2 e F3 essenciais: 128 ativos, 42 inativos (90%/71%) F4 e F5 essenciais: 140 ativos, 59 inativos (99%/100%)
<b>Diminuir progressivamente o raio de algumas características, mantendo as restantes constantes no farmacóforo 6I ativos, 4 inativos (43%/7%)</b>	F1 (1.4 → 1.3): 53 ativos, 3 inativos (37%/5%) F1 (1.4 → 1.2): 47 ativos, 3 inativos (33%/5%) F1 (1.4 → 1.1): 43 ativos, 2 inativos (30%/3%) F1 (1.4 → 1.0): 38 ativos, 1 inativos (27%/2%)  F2 (1 → 0.9): 55 ativos, 4 inativos (39%/7%) F2 (1 → 0.8): 45 ativos, 3 inativos (30%/5%) F2 (1 → 0.7): 34 ativos, 2 inativos (24%/3%)  F3 (1 → 0.9): 55 ativos, 3 inativos (39%/5%) F3 (1 → 0.8): 42 ativos, 1 inativos (30%/2%) F3 (1 → 0.7): 31 ativos, 1 inativos (22%/2%)  F4 (1 → 0.9): 62 ativos, 3 inativos (44%/5%) F4 (1 → 0.8): 54 ativos, 3 inativos (38%/5%) F4 (1 → 0.7): 42 ativos, 3 inativos (30%/5%)  F5 (1.4 → 1.3): 59 ativos, 3 inativos (42%/5%) F5 (1.4 → 1.2): 58 ativos, 3 inativos (41%/5%) F5 (1.4 → 1.1): 48 ativos, 3 inativos (34%/5%) F5 (1.4 → 1.0): 45 ativos, 2 inativos (32%/5%)
<b>Aumentar progressivamente o raio de algumas características, mantendo as restantes constantes no farmacóforo 6I ativos, 4 inativos (43%/7%)</b>	F1 (1.4 → 1.5): 64 ativos, 4 inativos (45%/7%) F2 (1 → 1.1): 71 ativos, 4 inativos (50%/7%) F3 (1 → 1.1): 71 ativos, 5 inativos (50%/8%)  F4 (1 → 1.1): 69 ativos, 4 inativos (49%/7%) F4 (1 → 1.2): 71 ativos, 4 inativos (50%/7%) F4 (1 → 1.3): 75 ativos, 4 inativos (53%/7%) F4 (1 → 1.4): 79 ativos, 4 inativos (56%/7%) F4 (1 → 1.5): 82 ativos, 4 inativos (56%/7%) F4 (1 → 1.6): 84 ativos, 5 inativos (59%/8%) F4 (1 → 1.7): 87 ativos, 5 inativos (61%/8%) <b>F4 (1 → 1.8): 91 ativos, 5 inativos (64%/8%)</b> F4 (1 → 1.9): 93 ativos, 6 inativos (65%/10%) F4 (1 → 2.0): 94 ativos, 7 inativos (66%/12%)
<b>(nota: por questões de espaço, algumas características só podem ser aumentadas 0.1)</b>	

F5 (1.4 → 1.5): 70 ativos, 5 inativos (49%/8%)  
 F5 (1.4 → 1.6): 71 ativos, 5 inativos (50%/8%)  
 F5 (1.4 → 1.7): 76 ativos, 5 inativos (54%/8%)  
 F5 (1.4 → 1.8): 78 ativos, 6 inativos (5%/10%)  
 F5 (1.4 → 1.9): 81 ativos, 6 inativos (57%/10%)  
 (Não se aumentou o raio de F5 para 2 porque ia sobrepor F3.)

Tentar um farmacóforo com as seguintes características:

**F1 (raio = 1.4) Hyd**  
**F2 (raio = 1.1) Hyd**  
**F3 (raio = 1) Hyd**  
**F4 (raio = 1.8) Acc|Don**  
**F5 (raio = 1.7) Acc|Don** } **98 ativos, 9 inativos (69%/15%)**

**Variar as características essenciais no Ph4 com 98 ativos, 9 inativos (69%/15%)**

F4 e F5 essenciais → 140 ativos, 59 inativos (99%/100%)

F1, F2, F3, F4 essenciais → 119 ativos, 37 inativos (83%/63%)

**Ignorar e esconder F5 no Ph4 com 98 ativos, 9 inativos (69%/15%) e sem qualquer característica essencial**

119 ativos, 36 inativos (83%/61%)

**Transformar as características Aceitadoras|Doadoras em características apenas aceitadoras ou doadoras no Ph4 com 98 ativos, 9 inativos (69%/15%) (Nota: F1, F2 e F3 permanecem inalterados)**

F4 e F5 Acc → 83 ativos, 4 inativos (58%/7%)  
**F4 Acc → 95 ativos, 7 inativos (67%/12%)**  
 F5 Acc → 90 ativos, 7 inativos (63%/12%)

**Transformar as características Aceitadoras|Doadoras em características apenas aceitadoras ou doadoras no Ph4 com 91 ativos, 5 inativos (64%/8%) (Nota: F1, F2 e F3 permanecem inalterados)**

F4 e F5 Don → 48 ativos, 4 inativos (34%/7%)  
 F4 Don → 73 ativos, 6 inativos (51%/10%)  
 F5 Don → 71 ativos, 6 inativos (50%/10%)

F4 e F5 Acc → 70 ativos, 3 inativos (49%/5%)  
 F4 Acc → 84 ativos, 4 inativos (59%/7%)  
 F5 Acc → 80 ativos, 4 inativos (56%/7%)

F4 e F5 Don → 27 ativos, 3 inativos (19%/5%)  
 F4 Don → 53 ativos, 4 inativos (37%/7%)  
 F5 Don → 58 ativos, 4 inativos (41%/7%)

**Os três melhores modelos farmacofóricos nesta etapa de refinamento apresentam as seguintes características:**

- F1 (raio = 1.4) Hyd, F2 (raio = 1) Hyd, F3 (raio = 1) Hyd, F4 (raio = 1.8) Acc|Don, F5 (raio = 1.4) Acc|Don – 91 ativos, 5 inativos (64%/8%)
- F1 (raio = 1.4) Hyd, F2 (raio = 1.1) Hyd, F3 (raio = 1) Hyd, F4 (raio = 1.8) Acc|Don, F5 (raio = 1.7) Acc|Don – 98 ativos, 9 inativos (69%/15%)
- F1 (raio = 1.4) Hyd, F2 (raio = 1.1) Hyd, F3 (raio = 1) Hyd, F4 (raio = 1.8) Acc, F5 (raio = 1.7) Acc|Don – 95 ativos, 7 inativos (67%/12%)

**Tabela 14** – Refinamento das características farmacofóricas do primeiro modelo farmacofórico obtido com base na estrutura tridimensional do recetor.

(A negrito encontram-se assinalados os melhores MF obtidos nesta etapa de refinamento.)

## Anexo 8

<i>Virtual screening com base na estrutura de ligandos</i>			
<b>Molécula (código NCI)</b>	<b>Score</b>	<b>Número de átomos pesados</b>	<b>Eficiência do ligando</b>
<b>NSC722655</b>	193.91	88	2.20
<b>NSC722659</b>	193.60	95	2.04
<b>NSC722657</b>	190.59	88	2.17
<b>NSC722660</b>	189.93	98	1.94
<b>NSC722654</b>	188.60	88	2.14
<b>NSC722661</b>	183.31	94	1.95
<b>NSC722658</b>	173.90	97	1.79
<b>NSC744991</b>	159.21	79	2.01
<b>NSC726291</b>	145.68	83	1.76
<b>NSC613013</b>	144.37	63	2.29
<b>NSC113926</b>	132.39	59	2.24
<b>NSC613015</b>	130.34	64	2.04
<b>NSC143420</b>	130.27	67	1.94
<b>NSC143423</b>	128.68	64	2.01
<b>NSC633664</b>	125.15	39	3.21
<b>NSC143424</b>	122.92	68	1.81
<b>NSC143429</b>	122.86	61	2.01
<b>NSC143426</b>	120.56	62	1.94
<b>NSC381101</b>	120.23	50	2.40
<b>NSC143433</b>	118.70	65	1.82
<b>NSC143428</b>	117.57	60	1.96
<b>NSC143471</b>	117.09	66	1.77
<b>NSC143416</b>	116.72	58	2.01
<b>NSC165124</b>	115.79	63	1.84
<b>NSC143441</b>	115.32	60	1.92
<b>NSC143421</b>	115.28	59	1.95
<b>NSC624751</b>	114.75	43	2.67
<b>NSC143427</b>	114.69	58	1.98
<b>NSC143418</b>	114.63	58	1.98
<b>NSC239378</b>	113.04	68	1.66
<b>NSC143417</b>	112.87	55	2.05
<b>NSC143430</b>	112.57	59	1.91
<b>NSC143442</b>	111.80	60	1.86
<b>NSC143422</b>	111.58	56	1.99
<b>NSC143439</b>	110.79	60	1.85

<b>NSC617200</b>	110.68	43	2.57
<b>NSC143425</b>	109.69	60	1.83
<b>NSC145618</b>	109.32	56	1.95
<b>NSC146725</b>	108.88	63	1.73
<b>NSC165125</b>	108.69	64	1.70
<b>NSC143432</b>	108.53	59	1.84
<b>NSC146730</b>	108.12	61	1.77
<b>NSC624746</b>	107.95	49	2.20
<b>NSC144138</b>	107.93	63	1.71
<b>NSC633249</b>	107.62	77	1.40
<b>NSC208397</b>	107.29	87	1.23
<b>NSC143450</b>	107.24	59	1.82
<b>NSC145665</b>	107.08	64	1.67
<b>NSC146719</b>	106.92	58	1.84
<b>NSC146721</b>	106.44	62	1.72
<b>NSC210761</b>	106.43	63	1.70
<b>NSC144108</b>	105.94	60	1.77
<b>NSC145657</b>	105.87	66	1.60
<b>NSC239379</b>	105.33	71	1.48
<b>NSC145620</b>	105.09	57	1.84
<b>NSC143436</b>	104.45	58	1.80
<b>NSC212551</b>	104.36	68	1.53
<b>NSC304412</b>	104.16	54	1.93
<b>NSC624754</b>	104.11	47	2.22
<b>NSC143431</b>	104.08	60	1.73
<b>NSC179498</b>	104.07	73	1.43
<b>NSC143463</b>	104.02	55	1.89
<b>NSC212549</b>	104.02	69	1.51
<b>NSC145619</b>	104.01	57	1.82
<b>NSC143487</b>	103.95	58	1.79

**Tabela 15** – Compostos selecionados no VS com base na estrutura de ligandos.

## Anexo 9

<i>Virtual screening com base na estrutura do recetor</i>			
<b>Molécula (código NCI)</b>	<b>Score</b>	<b>Número de átomos pesados</b>	<b>Eficiência do ligando</b>
<b>NSC722658</b>	199.83	97	2.06
<b>NSC722660</b>	199.66	98	2.04
<b>NSC722659</b>	197.94	95	2.08
<b>NSC722657</b>	197.81	88	2.25
<b>NSC722654</b>	193.20	88	2.20
<b>NSC722655</b>	193.14	88	2.19
<b>NSC722656</b>	184.43	85	2.17
<b>NSC722661</b>	182.01	94	1.94
<b>NSC744991</b>	165.25	79	2.09
<b>NSC726291</b>	145.49	83	1.75
<b>NSC113926</b>	134.24	59	2.28
<b>NSC208397</b>	128.70	87	1.48
<b>NSC304422</b>	124.87	59	2.12
<b>NSC381101</b>	121.47	50	2.43
<b>NSC208394</b>	120.47	69	1.75
<b>NSC90797</b>	119.91	82	1.46
<b>NSC655331</b>	118.11	50	2.36
<b>NSC624751</b>	118.10	43	2.75
<b>NSC165124</b>	115.86	63	1.83
<b>NSC83442</b>	114.22	29	3.94
<b>NSC617200</b>	114.13	43	2.65
<b>NSC624746</b>	113.43	49	2.31
<b>NSC724849</b>	111.44	76	1.47
<b>NSC173710</b>	111.30	69	1.61
<b>NSC624748</b>	110.52	44	2.51
<b>NSC648060</b>	108.83	41	2.65
<b>NSC624749</b>	108.19	103	1.05
<b>NSC672462</b>	108.15	116	0.93
<b>NSC139209</b>	107.96	67	1.61
<b>NSC717182</b>	106.63	75	1.42
<b>NSC110279</b>	105.71	73	1.48
<b>NSC624742</b>	105.67	38	2.78
<b>NSC631389</b>	105.59	62	1.70
<b>NSC165151</b>	105.28	63	1.67

---

<b>NSC624743</b>	105.03	32	3.28
<b>NSC688141</b>	104.89	63	1.66
<b>NSC624744</b>	104.69	37	2.83
<b>NSC615283</b>	104.03	38	2.73
<b>NSC697674</b>	103.82	58	1.79
<b>NSC624755</b>	103.45	45	2.30
<b>NSC624740</b>	103.15	35	2.95
<b>NSC405956</b>	102.65	86	1.19
<b>NSC624745</b>	102.58	35	2.93
<b>NSC691703</b>	102.03	70	1.46

---

**Tabela 16** – Compostos selecionados no VS com base na estrutura do recetor.

## VII – LISTA DE REFERÊNCIAS

1. BRANCOLINI, C. – **Inhibitors of the ubiquitin-proteasome system and the cell death machinery: how many pathways are activated?** *Current molecular pharmacology*. 1 (2008) 24–37.
2. WANG, J. ; MALDONADO, M. – **The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases.** *Cellular & molecular immunology*. 3 (2006) 255–61.
3. JUNG, T. ; GRUNE, T. – **Structure of the proteasome.** *Progress in molecular biology and translational science*. 109 (2012) 1–39.
4. KIM, H. M. ; YU, Y. ; CHENG, Y. – **Structure characterization of the 26S proteasome.** *Biochimica et biophysica acta*. 1809 (2011) 67–79.
5. GUO, N., PENG, Z. – **MGI32, a proteasome inhibitor, induces apoptosis in tumor cells.** *Asia-Pacific journal of clinical oncology*. 9 (2013) 6–11.
6. BLACKBURN, C. [et al.] – **Characterization of a new series of non-covalent proteasome inhibitors with exquisite potency and selectivity for the 20S beta5-subunit.** *The Biochemical journal*. 430 (2010) 461–76.
7. CHENG, Y. – **Towards an atomic model of the 26S Proteasome.** *Curr Opin Struct Biol*. 19 (2009) 203–208.
8. BEDFORD, L. [et al.] – **Assembly, structure, and function of the 26S proteasome.** *Trends in cell biology*. 20 (2010) 391–401.
9. FONSECA, P. C.; MORRIS, E. P. – **Structure of the human 26S proteasome: subunit radial displacements open the gate into the proteolytic core.** *The Journal of biological chemistry*. 283 (2008) 23305–23314.
10. FONSECA, P. C. ; HE, J. ; MORRIS, E. P. – **Molecular model of the human 26S proteasome.** *Molecular cell*. 46 (2012) 54–66.
11. KISSELEV, A. F. [et al.] – **The caspase-like sites of proteasomes, their substrate specificity, new inhibitors and substrates, and allosteric interactions with the trypsin-like sites.** *The Journal of biological chemistry*. 278 (2003) 35869–35877.
12. CHEN, D. [et al.] – **Bortezomib as the First Proteasome Inhibitor Anticancer Drug: Current Status and Future Perspectives.** *Curr Cancer Drug Targets*. 11 (2011) 239–253.
13. SHRINGARPURE, R. [et al.] – **Ubiquitin conjugation is not required for the degradation of oxidized proteins by proteasome.** *The Journal of biological chemistry*. 278 (2003) 311–8.

14. HUBER, E. [et al.] – **Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity.** *Cell*. 148 (2012) 727–738.
15. ADAMS, J. – **The proteasome: structure, function, and role in the cell.** *Cancer Treatment Reviews*. 29 (2003) 3–9.
16. GINODI, I. [et al.] – **Precise score for the prediction of peptides cleaved by the proteasome.** *Bioinformatics*. 24 (2008) 477–83.
17. BORISSENKO, L. ; GROLL, M. – **20S proteasome and its inhibitors: crystallographic knowledge for drug development.** *Chemical reviews*. 107 (2007) 687–717.
18. KAO, A. [et al.] – **Mapping the structural topology of the yeast 19S proteasomal regulatory particle using chemical cross-linking and probabilistic modeling.** *Mol Cell Proteomics*. 11 (2012) 1566–1577.
19. JUNG, T. ; CATALGOL, B. ; GRUNE, T. – **The proteasomal system.** *Molecular aspects of medicine*. 30 (2009) 191–296.
20. YAMAGUCHI, H. ; HSU, J. L. ; HUNG, M. C. – **Regulation of ubiquitination-mediated protein degradation by survival kinases in cancer.** *Frontiers in oncology* 2 (2012) 1–9.
21. MIRABELLA, A. [et al.] – **Specific cell-permeable inhibitor of proteasome trypsin-like sites selectively sensitizes myeloma cells to bortezomib and carfilzomib.** *Chem Biol*. 18 (2011) 608–618.
22. TOMKO, R.; HOCHSTRASSER, M. – **Order of the proteasomal ATPases and eukaryotic proteasome assembly.** *Cell Biochem Biophys*. 60 (2011) 13–20.
23. MARQUES, A. J. [et al.] – **Catalytic mechanism and assembly of the proteasome.** *Chemical reviews*. 109 (2009) 1509–36.
24. KISSELEV, A. F.; LINDEN, W. A. V. D. ; OVERKLEEF, H. S. – **Proteasome inhibitors: an expanding army attacking a unique target.** *Chemistry & biology*. 19 (2012) 99–115.
25. ZOLK, O.; SCHENKE, C. ; SARIKAS, A. – **The ubiquitin-proteasome system: focus on the heart.** *Cardiovascular research*. 70 (2006) 410–21.
26. HUBER, E. M.; GROLL, M. – **Inhibitors for the immuno- and constitutive proteasome: current and future trends in drug development.** *Angewandte Chemie (International edition in english)*. 51 (2012) 8708–20.

27. LANDER, G. C. [et al.] – **Complete subunit architecture of the proteasome regulatory particle.** *Nature.* 482 (2012) 186–91.
28. GOLDBERG, A. – **Development of proteasome inhibitors as research tools and cancer drugs.** *The Journal of cell biology.* 199 (2012) 583–8.
29. HADELER, K. P. ; KUTTLER, C. ; NUSSBAUM, A. K. – **Cleaving proteins for the immune system.** *Mathematical biosciences.* 188 (2004) 63–79.
30. GROLL, M. [et al.] – **Crystal structure of the 20 S proteasome:TMC-95A complex: a non-covalent proteasome inhibitor.** *Journal of molecular biology.* 311 (2001) 543–548.
31. OINONEN, C. ; ROUVINEN, J. – **Structural comparison of Ntn-hydrolases.** *Protein science.* 9 (2000) 2329–37.
32. DIEZ-RIVERO, C. M.; LAFUENTE, E. M. ; RECHE, P. – **Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome.** *BMC bioinformatics.* 11 (2010) 1–13.
33. ZHU, Y. [et al.] – **Design, synthesis, biological evaluation, and structure-activity relationship (SAR) discussion of dipeptidyl boronate proteasome inhibitors, part I: comprehensive understanding of the SAR of alpha-amino acid boronates.** *Journal of medicinal chemistry.* 52 (2009) 4192–4199.
34. LEI, B. [et al.] – **Molecular basis of the selectivity of the immunoproteasome catalytic subunit LMP2-specific inhibitor revealed by molecular modeling and dynamics simulations.** *The journal of physical chemistry.* 114 (2010) 12333–12339.
35. SCHECHTER, I.; BERGER, A. – **On the size of the active site in proteases. I. Papain.** *Biochemical and Biophysical Research Communications.* 27 (1967) 157–162.
36. GROLL, M. [et al.] – **Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome.** *Structure.* 14 (2006) 451–456.
37. PHAM, T. H. [et al.] – **A new series of N5 derivatives of the 1,1,5-trimethyl furo[3,4-c]pyridine-3,4-dione (cerpegin) selectively inhibits the post-acid activity of mammalian 20S proteasomes.** *Bioorganic & medicinal chemistry letters.* 22 (2012) 3822–3827.
38. BASSE, N. [et al.] – **Novel organic proteasome inhibitors identified by virtual and in vitro screening.** *Journal of medicinal chemistry.* 53 (2010) 509–13.

39. UNNO, M. [et al.] – **The Structure of the mammalian 20S proteasome at 2.75 Å resolution.** *Structure*. 10 (2002) 609–618.
40. RENTSCH, A. [et al.] – **Synthesis and Pharmacology of Proteasome Inhibitors.** *Angewandte Chemie (International edition in English)*. 52 (2013) 2–41.
41. WEI, D. [et al.] – **Fundamental reaction pathway and free energy profile for inhibition of proteasome by Epoxomicin.** *Journal of the American Chemical Society*. 134 (2012) 10436–10450.
42. SCREEN, M. [et al.] – **Nature of pharmacophore influences active site specificity of proteasome inhibitors.** *The Journal of biological chemistry*. 285 (2010) 40125–40134.
43. NEILSEN, P. M. [et al.] – **New 26S proteasome inhibitors with high selectivity for chymotrypsin-like activity and p53-dependent cytotoxicity.** *ACS chemical biology*. 8 (2013) 353–359.
44. KORTUEM, K. ; STEWART, K. – **Carfilzomib.** *Blood*. 121 (2013) 893–897.
45. KAZI, A. [et al.] – **Discovery of a novel proteasome inhibitor selective for cancer cells over non-transformed cells.** *Cell Cycle*. 8 (2009) 1940–1951.
46. KISSELEV, A. F.; CALLARD, A. ; GOLDBERG, A. L. – **Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate.** *The Journal of biological chemistry*. 281 (2006) 8582–8590.
47. MOLINEAUX, S. M. – **Targeting proteasomal protein degradation in cancer.** *Clinical Cancer Research*. 18 (2011) 1–18.
48. KISSELEV, A. F. – **Joining the army of proteasome inhibitors.** *Chemistry & biology*. 15 (2008) 419–421.
49. WILT, L. [et al.] – **Bortezomib and TRAIL: a perfect match for apoptotic elimination of tumour cells?** *Critical reviews in oncology/hematology*. 85 (2013) 363–372.
50. **ClinicalTrials.gov.** U.S. National Institutes of Health. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://clinicaltrials.gov/>
51. HINES, J. [et al.] – **Proteasome Inhibition by Fellutamide B Induces Nerve Growth Factor Synthesis.** *Chem Biol*. 15 (2008) 501–512.
52. GROLL, M. [et al.] – **Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of  $\alpha'$ , $\beta'$ -Epoxyketone Proteasome Inhibitors.** *Journal of the American Chemical Society*. 122 (2000) 1237–1238.

53. KIKUCHI, J. [et al.] – **Homopiperazine derivatives as a novel class of proteasome inhibitors with a unique mode of proteasome binding.** PLoS One. 8 (2013) e60649.
54. MOREIRA, V. [et al.] – **Novel oleanolic vinyl boronates: synthesis and antitumor activity.** European Journal of Medicinal chemistry. 63 (2013) 46–56.
55. GALLERANI, E. [et al.] – **A first in human phase I study of the proteasome inhibitor CEP-18770 in patients with advanced solid tumours and multiple myeloma.** European Journal of Cancer. 49 (2013) 290–6.
56. KUPPERMAN, E. [et al.] – **Evaluation of the proteasome inhibitor MLN9708 in preclinical models of human cancer.** Cancer Res. 70 (2010) 1970–1980.
57. PIVA, R. [et al.] – **CEP-18770: a novel, orally active proteasome inhibitor with a tumor-selective pharmacologic profile competitive with bortezomib.** Blood. 111 (2008) 2765–2775.
58. GALLASTEGUI, N. [et al.] – **Hydroxyureas as noncovalent proteasome inhibitors.** Angewandte Chemie (International edition in English). 51 (2012) 247–9.
59. GRAWERT, M. A. [et al.] – **Elucidation of the alpha-keto-aldehyde binding mechanism: a lead structure motif for proteasome inhibition.** Angewandte Chemie (International edition in English). 50 (2011) 542–544.
60. TAHA, MUTASEM et al. – **Virtual Screening.** 2012, ISBN 978-953-51-0308-0.
61. TRIGGLE, DAVID et al. – **Comprehensive medicinal chemistry II, volume 4: Computer-Assisted Drug Design.** 2006, ISBN 0-08-044513-6.
62. LUSHER, S. et al. – **Computer-Aided Drug Design.** In: International Computational Drug Discovery Practical Course, Nijmegen – The Netherlands. 2007.
63. LIAO, C. [et al.] – **Software and resources for computational medicinal chemistry.** 3 (2012) 1057–1085.
64. WERMUTH, CAMILLE – **The Practice of Medicinal Chemistry.** 3rd Ed. 2008, ISBN: 9780123741943.
65. GAO, Q. [et al.] – **Pharmacophore based drug design approach as a practical process in drug discovery.** Current computer-aided drug design. 6 (2010) 37–49.
66. WALTERS, W., MURCKO, M. – **Prediction of “drug-likeness”.** Advanced drug delivery reviews. 54 (2002) 255–71.
67. COSCONATI, S. [et al.] – **Virtual Screening with AutoDock: theory and practice.** 5 (2011) 597–607.

68. LIPINSKI, C. [et al.] – **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Advanced drug delivery reviews.* 46 (2001) 3–26.
69. HOLTJE, Hans-Dieter et al. – **Molecular Modeling - basic principals and applications.** 3rd Ed. 2008, Weinheim, Germany: Wiley-VCH Verlag GmbH, 1996. ISBN 3-527-29384-1.
70. KLEBE, G. – **Virtual ligand screening: strategies, perspectives and limitations.** *Drug discovery today.* 11 (2006) 580–594.
71. SCHNEIDER, G.; BÖHM, H.J. – **Virtual screening and fast automated docking methods.** *Drug discovery today.* 7, (2002) 64–70.
72. LILJEFORS, T. [et al.] – **Computational medicinal chemistry for drug discovery.** (2004).
73. **Protein Data Bank.** Research Collaboratory for Structural Bioinformatics (RCSB). [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://www.rcsb.org/pdb/home/home.do>
74. CAVASOTTO, C.; PHATAK, S. – **Homology modeling in drug discovery: current trends and applications.** *Drug discovery today.* 14 (2009) 676–683.
75. VYAS, V. [et al.] – **Homology modeling a fast tool for drug discovery: current perspectives.** *Indian journal of pharmaceutical sciences.* 74 (2012) 1–17.
76. LEI, M. [et al.] – **Progress of computer-aided drug design (CADD) of proteasome inhibitors.** *Current topics in medicinal chemistry.* 11 (2011) 2931–2944.
77. **Basic Local Alignment Search Tool (BLAST).** The National Center for Biotechnology Information, Bethesda (USA). [Acedido a 28 de Março de 2013]. Disponível na Internet: <http://blast.ncbi.nlm.nih.gov/BLAST.cgi>
78. **Molecular Operating Environment (MOE).** Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, 2012.
79. **Modeller.** University of California, San Francisco, USA. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://salilab.org/modeller/>
80. **The Universal Protein Resource (UniProt).** The Uniprot Consortium. [Acedido a 24 de Agosto de 2013]. Disponível na Internet: <http://www.uniprot.org>
81. KARADAGHI, S. – **Torsion angles and the Ramachandran Plot in protein structures.** *Protein structure and structural bioinformatics* (2012). [Acedido a 20 de

- Agosto de 2013]. Disponível na Internet: <http://www.proteinstructures.com/Structure/Structure/Ramachandran-plot.html>
82. ARNOLD, K. [et al.] – **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.** *Bioinformatics*. 22, (2006) 195–201.
  83. **Protein Model Portal** . Protein Structure Bioinformatics group, SIB - Swiss Institute of Bioinformatics, Biozentrum University of Basel. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://www.proteinmodelportal.org/>
  84. **Modbase.** University of California, San Francisco. [Acedido a 2 de Setembro de 2013]. Disponível na internet: [http://modbase.compbio.ucsf.edu/modbase/modbase\\_help.html](http://modbase.compbio.ucsf.edu/modbase/modbase_help.html)
  85. HEIN, M.; ZILIAN, D.; SOTRIFFER, C. – **Docking compared to 3D-pharmacophores: the scoring function challenge.** *Drug Discovery Today: Technologies*. 7 (2010) e229–e236.
  86. SAKKIAH, S. [et al.] – **3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors.** *European journal of medicinal chemistry*. 45 (2010) 2132–2140.
  87. DS Catalyst. Accelrys, San Diego (USA). [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://accelrys.com/>
  88. **Phase.** Schrodinger. [Acedido a 2 de Setembro de 2013]. Disponível na internet: <http://www.schrodinger.com/productpage/14/13/>
  89. DS Structure Based Pharmacophore. Accelrys, San Diego (USA). [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://accelrys.com/>
  90. FANG, C.; XIAO, Z.; GUO, Z. – **Generation and validation of the first predictive pharmacophore model for cyclin-dependent kinase 9 inhibitors.** *Journal of Molecular Graphics and Modelling*. 29 (2011) 800–808.
  91. HUANG-SHENG. L. [et al.] – **Pharmacophore modeling and virtual screening to design the potential influenza virus endonuclease inhibitors.** *Journal of the chinese chemical society*. 59 (2012) 1430–1438.
  92. GOZALBES, R. [et al.] – **Development and validation of a pharmacophore-based QSAR model for the prediction of CNS activity.** *ChemMedChem*. 4 (2009) 204–209.

93. MISHRA, R.; SINGH, J. – **Generation, validation, and utilization of a three-dimensional pharmacophore model for EP3 antagonists.** J. Chem. Inf. Model. 50 (2010) 1502–1509.
94. KIANI, Y.; KALSOOM, S.; RIAZ, N. – **In silico ligand-based pharmacophore model generation for the identification of novel *Pneumocystis carinii* DHFR inhibitors.** Medicinal chemistry research. 22 (2013) 949–963.
95. MILLETTI, F. [et al.] – **Tautomer enumeration and stability prediction for virtual screening on large chemical databases.** Journal of chemical information and modeling. 49 (2009) 68–75.
96. LEACH, A.R. [et al.] – **Three-dimensional pharmacophore methods in drug discovery.** Journal of medicinal chemistry. 53, (2009) 539–558.
97. BOHM, H., [et al.] – **Oxygen and nitrogen in competitive situations: which is the hydrogen-bond acceptor?** Chemistry. 2 (1996) 1509–1513.
98. NOVIKOV, F.; CHILOV, G. – **Molecular docking: theoretical background, practical applications and perspectives.** Mendeleev Communications. 19 (2009) 237–242.
99. **FlexX.** BioSolveIT. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://www.biosolveit.de/flexx/index.html?ct=1>
100. **Genetic optimisation for ligand docking (GOLD).** The Cambridge Crystallographic Data Centre. [Acedido a 2 de Setembro de 1013]. Disponível na Internet: <http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx>
101. **Glide.** Schrodinger. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://www.schrodinger.com/productpage/14/5/>
102. SHOICHET, B. K. [et al.] – **Lead discovery using molecular docking.** Current opinion in chemical biology. 6 (2002) 439–446.
103. **Surflex-Dock.** Tripos International. St. Louis, USA. [Acedido a 2 de Agosto de 2013]. Disponível na internet: [http://www.tripos.com/index.php?family=modules,SimplePage...&page=surflex\\_dock&se=0](http://www.tripos.com/index.php?family=modules,SimplePage...&page=surflex_dock&se=0)
104. **ICM-Docking.** Molsoft L.L.C.. [Acedido a 5 de Setembro de 2013]. Disponível na Internet: <http://www.molsoft.com/docking.html>
105. **GlamDock.** Chil<sup>2</sup>. [Acedido a 5 de Setembro de 2013]. Disponível na Internet: <http://www.chil2.de/Glamdock.html>

106. **Moldock.** Molegro Virtual Docker. Prismet, Denmark. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://www.molegro.com/mvd-product.php>
107. MCINNES, C. – **Virtual screening strategies in drug discovery.** Current opinion in chemical biology. 11 (2007) 494–502.
108. SCHUSTER, D. [et al.] – **Development and validation of an in silico P450 profiler based on pharmacophore models.** Current Drug Discovery Technologies. 3 (2006) 1–48.
109. WANG, J.; LIN, J. – **Scoring functions for prediction of protein-ligand interactions.** Current pharmaceutical design. 19 (2013) 2174–2182.
110. VERDONK, M. L. [et al.] – **Improved protein-ligand docking using GOLD.** Proteins: structure, function and genetics. 52 (2003) 609–623.
111. SASTRY, G. M. [et al.] – **Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.** Journal of computer-aided molecular design. 27 (2013) 221–34.
112. KENYON, V. [et al.] – **Novel human lipoxigenase inhibitors discovered using virtual screening with homology models.** Journal of medicinal chemistry. 49 (2006) 1356–1363.
113. LI, H. [et al.] – **Identification of novel falcipain-2 inhibitors as potential antimalarial agents through structure-based virtual screening.** Journal of medicinal chemistry. 52 (2009) 4936–4940.
114. Cambridge Crystallographic Data Centre GOLD User Guide. (2011).
115. ANNAMALA, M.K.; INAMPUDI, K.K.; GURUPRASAD, L. – **Docking of phosphate and trehalose analog inhibitors into M. tuberculosis mycolyltransferase Ag85C: Comparison on the two scoring fitness functions GoldScore and ChemScore, in the GOLD software.** Bioinformation. 1 (2007) 339–351.
116. LEI, M. [et al.] – **Pharmacophore modeling, docking studies, and synthesis of novel dipeptide proteasome inhibitors containing boron atoms.** J. Chem. Inf. Model. 49 (2009) 2092–2100.
117. XU, K. [et al.] – **Design and synthesis of naphthoquinone derivatives as antiproliferative agents and 20S proteasome inhibitors.** Bioorganic & medicinal chemistry letters. 22 (2012) 2772–2774.
118. LABUTE, P. – **An Integrated Application in MOE for the Visualization and Analysis of Protein Active Sites with Molecular Surfaces, Contact Statistics**

- and Electrostatic Maps.** Chemical Computing Group. [Acedido a 3 de Julho de 2013]. Disponível na Internet: <https://www.chemcomp.com/journal/sitefind.htm>
119. **Structure files of NCI open database compounds.** CADD Group cheminformatics tools and user services. [Acedido a 2 de Maio de 2013]. Disponível na Internet: <http://cactus.nci.nih.gov/download/nci>
120. **The PubChem project.** The National Center for Biotechnology Information. [Acedido a 2 de Setembro de 2013]. Disponível na Internet: <http://pubchem.ncbi.nlm.nih.gov/>
121. KUNTZ, I. D. [et al]. – **The maximal affinity of ligands.** Proceedings of the National Academy of Sciences of the United States of America. 96 (1999) 9997–10002.
122. ABAD-ZAPATERO, C. – **Ligand efficiency indices for effective drug discovery.** Expert opinion on drug discovery. 2 (2007) 469–488.
123. REYNOLDS, C. H.; BEMBENEK, S. D.; TOUNGE, B. A. – **The role of molecular size in ligand efficiency.** Bioorganic & medicinal chemistry letters. 17 (2007) 4258–4261.
124. GUIMARÃES, Serafim; MOURA, Daniel; SILVA, Patrício Soares. – **Terapêutica medicamentosa e suas bases farmacológicas.** 5ª edição. Porto Editora: 2006. ISBN:978-972-0-06029-7.