Universidade de São Paulo – USP
Instituto de Química de São Carlos - IQSC
Grupo de Química Medicinal e Biológica - NEQUIMED

Análise quimiométrica de seletividade nos sítios ativos de
cisteíno proteases da família da papaína para o estudo de
substâncias antiparasitárias e antineoplásicas

Dissertação apresentada ao Programa de Pós-Graduação
em Química do Instituto de Química de São Carlos da
Universidade de São Paulo como parte dos requisitos para
a obtenção do título de mestre em química com ênfase em
Química Orgânica e Biológica.

Candidato: Felipe Marçal Morgantini
Bolsista CNPq, processo 130469/2018-01
Orientador: Prof. Dr. Andrei Leitão

São Carlos, S.P.

2020

FELIPE MARÇAL MORGANTINI

Exemplar revisado

O exemplar original encontra-se em
acervo reservado na Biblioteca do IQSC-USP

Análise quimiométrica de seletividade nos sítios ativos de
cisteíno proteases da família da papaína para o estudo de
substâncias antiparasitárias e antineoplásicas

Dissertação apresentada ao Programa de Pós-Graduação
em Química do Instituto de Química de São Carlos da
Universidade de São Paulo como parte dos requisitos para
a obtenção do título de mestre em química com ênfase em
Química Orgânica e Biológica.

Orientador: Prof. Dr. Andrei Leitão

São Carlos, maio de 2021

AGRADECIMENTOS

Minha gratidão:

Ao aconchegante Instituto de Química de São Carlos (IQSC);

Aos órgãos governamentais responsáveis por administrar os orçamentos às pesquisas no país: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); e no estado de São Paulo: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP);

Ao orientador professor Andrei Leitão;

Aos dirigentes do grupo NEQUIMED, professores Andrei Leitão, Fernanda Canduri e Carlos Alberto Montanari, à funcionária Fabiana Rosini e a todos discentes;

Aos membros da Comissão da Pós-Graduação (CPG): Gislei Aparecida Alchangelo de Oliveira, Daniele Decicino Paiutta e Andreia Cristina Cardozo de Moraes;

Aos membros do Departamento de Química e Física Molecular (DQFM): Claudia Nascimento de Almeida e Veroneide de Maria da Silva Athayde;

Aos membros da Seção Técnica de Informática (STI): Oscar Carlson Gasparetto, Irineu Lourenço da Silva, Flávio Oliveira e Eduardo Zanollo Junior.

Aos professores membros das bancas avaliadoras: Albérico Borges Ferreira da Silva, Magaly Girão Albuquerque e Silvana Giuliatti.

Amigos do ambiente da universidade: Carolina Bianco, Chubraider Xavier, Luiz Antonio Abondancia, Marcio Costa, Mariano Valim e Wilneide Maiorano.

Professores os quais participei de disciplinas: Andrei Leitão em duas disciplinas: “Chemical Aspects of the Cell” e “A Química Medicinal e o Planejamento de Fármacos”, a qual o professor Carlos Alberto Montanari também ministrou aulas. Rafael Guido ministrou a disciplina “Química Medicinal: Fundamentos do Planejamento de Fármacos”. Júlio César Borges em “Bioquímica Avançada”. Edénir Rodrigues Pereira Filho em “Quimiometria 1: planejamento de experimentos” e Renato Lajarim Carneiro em “Quimiometria 2: análise multivariada.”

A todos os amigos, destacando-se Jadson Luan dos Santos Marcelino, Breno Souza e em especial Guilherme Barsoti pela contribuição expressiva no progresso da dissertação.

Obviamente à família.

EPÍGRAFE

“Quando Fraser e Brown propuseram, em 1869, uma definição de relações entre estrutura-química e atividade eles, na realidade, estabeleceram o exemplo pioneiro da colaboração entre farmacologistas e químicos.”¹

RESUMO

Entre as enzimas pesquisadas pelo Grupo de Química Medicinal e Biológica NEQUIMED estão as cisteíno proteases (CP), especialmente as envolvidas em doenças negligenciadas, e.g. doença de Chagas, endêmica na América Latina, e Leishmaniose, endêmica em 83 países (forma visceral) incluindo o Brasil. As CPs envolvidas neste trabalho pertencem à família da papaína: cruzaina do *Trypanosoma cruzi* e catepsinas humanas: B, K, L e S. Por serem alvos terapêuticos de diversas patologias a inibição dessas proteases é bastante pesquisada. A alta similaridade na sequência de aminoácidos torna a seletividade um desafio. Considerando a grande disponibilidade de estruturas cristalográficas depositadas, a abordagem computacional adotada foi a utilização de descritores tridimensionais: campos de interação molecular, onde o mapeamento energético é obtido através da interação intermolecular entre sondas químicas e proteína no sítio ativo, que é usado como variáveis na análise de componentes principais, o qual destacou regiões seletivas, racionalizando a busca de grupos químicos com propriedades similares ao modelo de seletividade. O espaço químico de busca utilizado foi delimitado por inibidores miméticos de dipeptidil nitrila com elevada potência disponíveis na base de moléculas ChEMBL. A complementaridade química dos compostos selecionados foi analisada após docagem molecular para a seleção de substituintes químicos que foram utilizados como blocos construtores no planejamento de estruturas baseado no ligante denominado hibridização molecular, buscando aliar seletividade e potência inibitória nos compostos planejados.

Palavras-chave: campos de interação molecular (MIF), quimioinformática, planejamento de ligantes baseado no alvo macromolecular (SBLD), cisteíno catepsinas, cruzaina, análise de componentes principais (PCA), docagem molecular, hibridização molecular (HM)

ABSTRACT

Among the enzymes researched by the Medicinal and Biological Chemistry Group NEQUIMED are cysteine proteases (CP), especially those involved in neglected diseases, e.g. Chagas disease, endemic in Latin America, and Leishmaniasis, endemic in 83 countries (visceral form) including Brazil. The CPs involved in this work belong to the papain family: cruzain from *Trypanosoma cruzi* and human cathepsins: B, K, L and S. Since they are therapeutic targets for several pathologies, the inhibition of these proteases is intensively researched. The high similarity in the amino acid sequence makes selectivity a challenge. Considering the great availability of deposited crystallographic structures, the in silico approach adopted was the use of three-dimensional descriptors: molecular interaction fields, where the energy mapping is obtained through the intermolecular interaction between chemical probes and protein in the active site, that is used as variables in principal component analysis, which highlighted selective regions, rationalizing the search for chemical groups with properties similar to the selectivity model. The chemical search space used was limited by dipeptidyl nitrile inhibitors with high potency available on ChEMBL database. The chemical complementarity of the selected compounds was analyzed after molecular docking for the selection of chemical substituents that were used as building blocks in ligand-based drug design approach called molecular hybridization, seeking to combine selectivity and pharmacodynamic potency in the planned compounds.

Key-words: molecular interaction field (MIF), chemoinformatics, structure-based ligand design (SBLD), cysteine cathepsin, cruzain, partial components analysis (PCA), molecular docking, molecular hybridization (MH)

LISTA DE ILUSTRAÇÕES

Figura 1 — Ângulos que caracterizam a geometria da ligação de hidrogênio através da direcionalidade do par de elétrons.	22
Figura 2 — Diagrama esquemático das interações entre cadeias laterais do substrato peptídico e subsítios da proteína com nomenclatura de Schechter e Berger ⁴⁹	26
Figura 3 — Estrutura tridimensional em estilo <i>cartoon</i> destacando os motivos (<i>motifs</i>) da enzima cruzaína, com adição da volta ocludente da proteína catepsina B destacada em preto.	27
Figura 4 — Singularidade da catepsina B: alça batizada de volta ocludente.	28
Figura 5 — Estrutura de substrato ou inibidor no sítio ativo da catepsina B com a conformação fechada.	28
Figura 6 — Casos confirmados da doença de Chagas aguda entre no período 2006-2019.	30
Figura 7 — Protótipo mimético de dipeptidil nitrila estudado no grupo NEQUIMED ao centro da imagem e moléculas que inspiraram a respectiva série congênere localizadas nas laterais.	32
Figura 8 — Reação catalítica de cisteíno proteases com composto dipeptidil (identificada genericamente pela esfera preta) funcionalizada com grupo reativo nitrila (cinza).	33
Figura 9 — Definição de ângulos utilizados como parâmetros na função energética referente a ligação de hidrogênio do campo de força GRID.	36
Figura 10 — Método de análise gráfica do mapeamento energético feita pelo programa GRID através da visualização dos campos de interação molecular com respectiva estrutura molecular.	38
Figura 11 — Modificação estrutural de hidroxila (composto A) por grupo guanidina (composto B) guiada através de análise no modo de interação e mapeamento energético com campos de interação molecular. A imagem C mostra o composto B cristalizado com a enzima neuraminidase.	39
Figura 12 — Resultados gráficos da PCA: gráfico de <i>scores</i> no lado esquerdo e gráfico de <i>loadings</i> no lado direito, em estudo que analisou vinhos de 4 países com relação a 14 propriedades físico-químicas.	43
Figura 13 — Exemplo do planejamento de híbrido molecular (ao centro) utilizando subestruturas dos fármacos Raloxifeno e Sorafenib através de estudo que utilizou dentre	

as ferramentas a hibridização molecular.....	46
Figura 14 — Estudo de hibridização molecular que resultou no fármaco Benorilato.	47
Figura 15 — Fluxograma das técnicas computacionais utilizadas no trabalho.	50
Figura 16 — Matriz de dados dos valores energéticos dos campos de interação molecular ilustrada de forma representativa	59
Figura 17 – Arcabouço molecular das dipeptidil nitrilas selecionadas para o estudo à esquerda e respectivo arcabouço molecular modificado para a forma ligada, preparado o procedimento de redocagem, à direita.....	62
Figura 18 — Subestrutura molecular do complexo PDB 1MS6 utilizada em restrições na docagem das moléculas planejadas.	67
Figura 19 — Imagem à esquerda: interações conservadas de átomos do protótipo dipeptidil nitrila com respectivos átomos da cadeia principal da proteína assim como a nomenclatura dos grupos substituintes com respectivos subsítios em que ocupam na enzima. Imagem ao centro: boxplot de distâncias entre átomos mencionados do complexo proteína ligante. Imagem ao centro: boxplot de ângulos entre os átomos.	70
Figura 20 — Sobreposição de subestruturas moleculares das moléculas miméticos de dipeptidil nitrilas cocrystalizas com as enzimas estudadas (cruzaína, catepsinas humanas: B, K, L e S) destacando a similaridade conformacional.	72
Figura 21 — Sítio ativo das enzimas cisteíno proteases divididas em subsítios.....	73
Figura 22 — Diagrama de Venn representando relações físico-químicas entre os 20 aminoácidos principais.....	74
Figura 23 — Subsítio S1' das enzimas estudadas (cruzaína e catepsinas humanas B, K, L e S) com destaque para os aminoácidos localizados na mesma região (rotulados) que não são semelhantes.	75
Figura 24 — Sítio ativo da catepsina K (PDB 1AU3 à esquerda e PDB 1MS6 à direita) com destaque para resíduo não conservado no subsítio S1' (em relação às proteínas estudadas, catepsinas: B, L, S e cruzaína) de glutamina, rotulados com distâncias de átomos de inibidores cocrystalizados.	76
Figura 25 — Sítio ativo do complexo inibidor-cruzaína (PDB 1F29) destacando distância entre átomo do ligante em P1 alcançando resíduo de aspartato em S1'.....	77
Figura 26 — Resíduo de arginina no subsítio S1' não conservado na região estrutural da catepsina S (a partir da imagem à esquerda: PDB ID 1MS6, 2R9O e 2R9N) e ligantes	

cocrystalizados com átomos próximos rotulados com distâncias.	77
Figura 27 — Subsítio S1 das enzimas estudadas (catepsinas B, K, L, S e cruzaina) com destaque para resíduo de glutamato único da catepsina B.....	78
Figura 28 — Catepsina B (complexo 6AY2) e seu resíduo não conservado (em relação às proteínas estudadas: catepsinas K, L, S e cruzaina) de glutamato no subsítio S1 mostrada através da única estrutura cristalográfica de CatB que acessa o resíduo.....	78
Figura 29 — Subsítio S2 das proteínas estudadas (catepsinas humanas B, K, L, S e cruzaina) separadas em grupos, 7 ao 14, com aminoácidos não conservados destacados.....	79
Figura 30 — Glutamato do subsítio S2 da cruzaina interagindo com P2 do inibidor: cadeia lateral da tirosina à esquerda e guanidina na imagem à direita, através de ligação de hidrogênio e possível ponte salina , respectivamente.	80
Figura 31 — Possível interação hidrofóbica π -stacking no canto inferior de S2 (F205) e na saliência inferior de S3 (F67) da proteína catepsina S.....	81
Figura 32 — Complexo inibidor-CatL (PDB 3HWN) possui P2 que alcança resíduos de glutamato de difícil alcance em S2 profundo.	82
Figura 33 — Subsítio S3 das proteínas estudadas (catepsinas B, K, L, S e cruzaina) superpostas, separadas por grupos, 15 ao 20, para visualização de resíduos não conservados.....	83
Figura 34 — Estruturas cristalográficas da cruzaina destacando resíduo não conservado de serina em S3 e grupos próximos mostrando a possibilidade de interações seletivas.	84
Figura 35 — Subsítio S2' das proteínas estudadas (catepsinas B, K, L, S e cruzaina) superpostas, separada em grupos, 21 ao 24, para visualização de resíduos não conservados.....	85
Figura 36 — Imagem do sítio catalítico da proteína catepsina B (PDB 1CSB) com destaque ao subsítio S2' onde está localizado o par de histidinas seletivas (em relação às outras proteínas do trabalho: CatK, CatL, CatS e Cz) que participam de pontes salinas feita com o carboxilato do inibidor cocrystalizado.....	86
Figura 37 — Exemplos de critérios utilizados para escolha dos monômeros que compuseram o conjunto de proteínas do trabalho.....	88
Figura 38 — Ilustração de superposição estrutural com duas enzimas diferentes da mesma família.	90
Figura 39 — Valor de RMSD por classe enzimática representado por gráfico diagrama de caixas.....	91
Figura 40 — Diferenças estruturais na catepsina B em relação às outras enzimas estudadas	

(catepsinas: K, L, S e cruzaína).	92
Figura 41 — Ilustração das estruturas antes e após modelagem do loop lateral exposto ao solvente.....	94
Figura 42 — Caixa com malha delimitando as coordenadas do cálculo com intersecções entre as coordenadas x, y e z igualmente espaçadas em 1 Å.	95
Figura 43 — Gráfico de barras mostrando porcentagens de variância explicada pelas componentes principais de 1 à 10.	96
Figura 44 — Gráfico de pontuação da PCA com as variáveis PC1, PC2 e PC3 realizada com o conjunto de proteínas do trabalho.....	97
Figura 45 — Gráfico de pontuação da PCA com as variáveis PC1, PC2 e PC4 realizada com o conjunto de proteínas do trabalho.....	98
Figura 46 — Pontos encontrados pela utilização sequencial MIF-PCA que representam interações intermoleculares entre as sondas químicas específicas e o sítio ativo que apresentam seletividade dentre o grupo de proteínas estudados.....	100
Figura 47 — Estruturas dos compostos dipeptidil nitrilas cocrystalizados com as enzimas estudadas: catepsinas B, K, L, S e cruzaína.	103
Figura 48 — Gráfico de barras relacionando-se o RMSD entre as estruturas redocadas com os programas HYBRID e FRED e os respectivos ligantes cognatos.	104
Figura 49 — Conjunto de inibidores da base de dados ChEMBL obtido após seleção virtual.	106
Figura 50 — Dipeptidil Nitrila CHEMBL3640727 com P3 carregado positivamente e pontos seletivos que representam sondas de cargas positivas docado em receptor catepsina K (PDB 1VSN).	110
Figura 51 — Proposta de inibidores da catepsina K com P3 selecionado a partir da aplicação do mapeamento de seletividade (CHEMBL360727 e CHEMBL371064 (balicatib)) e hibridização molecular com o composto Neq0409 previamente descrito como inibidor de catepsina L.	111
Figura 52 — Híbridos moleculares 1 (hm_1) e 2 (hm_2) docados covalentemente com o programa GOLD visualizados com os pontos seletivos encontrados no modelo MIF-PCA para a catepsina K.	112
Figura 53 — Estudo de docagem associado aos pontos de seletividade obtidos pela PCA sugere que a utilização de um grupo espaçador modifica o composto Neq0682 um candidato mais adequado para a catepsina K segundo hipótese encontrada no modelo	

seletivo MIF-PCA.	113
Figura 54 — Dipeptidil Nitrila CHEMBL3640716 e pontos seletivos que representam sondas de cargas negativas docado em receptor catepsina S (PDB 1MS6).	114

LISTA DE TABELAS

Tabela 1 — Esquema das interações intermoleculares em química medicinal.....	19
Tabela 2 — Raios atômicos em uma interação de van der Waals, unidades em angstrom.	20
Tabela 3 — Parâmetros de série de análogos de estudo (arcabouço na parte inferior direita) que investigou ligações de halogênio em compostos dipeptidil nitrila ^{32,33} à esquerda, e gráfico de cálculo quântico energético ³² à direita.	24
Tabela 4 — Exemplo de doenças relacionadas às cisteíno proteases estudadas.....	31
Tabela 5 — Estudo com inibidores parasitário e humano da enzima N-miristoil transferase utilizando hibridização molecular confere seletividade ao composto 3.	48
Tabela 6 — Esqueletos moleculares miméticos de dipeptidil nitrilas cocrystalizados com as proteínas estudadas (catepsinas B, K, L, S e cruzaina).	52
Tabela 7 — Código PDB ID dos complexos que compuseram a análise do modo de interação do sítio ativo.....	53
Tabela 8 — Valores da caixa em que foi realizado o cálculo dos Campos de Interação Molecular.	56
Tabela 9 — Sondas químicas virtuais utilizadas no programa GRID para o cálculo dos campos de interação molecular.....	58
Tabela 10 — Sondas químicas utilizadas na PCA.	60
Tabela 11 — Critérios de seleção de variáveis da matriz dos pesos (<i>loadings</i>) adotados após análise dos gráficos de <i>scores</i>	61
Tabela 12 — Ligantes utilizados para delimitar região de pontos no sítio ativo.	61
Tabela 13 — Informações do uso do programa OMEGA.....	62
Tabela 14 — Informações sobre a utilização do Make Receptor e do FRED no procedimento de	63
Tabela 15 — Parâmetros referentes a redocagem com o programa FRED e HYBRID.	63
Tabela 16 — Proteínas utilizadas como alvos na docagem.	64
Tabela 17 — Moléculas selecionadas para docagem covalente, exceto pelo Neq409.....	66
Tabela 18 — Distância e ângulos de átomos do arcabouço principal dos ligantes do conjunto de proteínas do trabalho cocrystalizados com dipeptidil nitrilas.	70
Tabela 19 — Códigos PDB ID do conjunto de proteínas do trabalho: catepsinas B, K, L, S e cruzaina.....	89

Tabela 20 — Dados estatísticos de valores do desvio da raiz quadrada média (RMDS) do conjunto de proteínas do trabalho: catepsinas B, K, L, S e cruzaina.....	91
Tabela 21 — Proteínas que tiveram seu loop exposto ao solvente modelado pelo programa FREAD por haver resíduos não resolvidos nas estruturas cristalográficas.	93
Tabela 22 — Legenda de cores relacionando a natureza das interações das sondas químicas virtuais.	99
Tabela 23 — Conjunto de dipeptidil nitrilas cocrystalizadas nas proteínas do estudo: catepsinas B, K, L, S e cruzaina.	102
Tabela 24 — Valores de pK_i dos inibidores resgatados no filtro realizado na base de dados ChEMBL	108
Tabela 25 — Dados do alinhamento tridimensional realizado para as cisteíno proteases da base de dados baseado na estrutura cristalográfica da cruzaina 4QH6 pelo programa PyMol.	138
Tabela 26 — Informações relativas aos parâmetros das sondas químicas virtuais do programa GRID utilizadas no cálculo dos campos de interação molecular.	141
Tabela 27 — Pontuação da docagem com a série de inibidores virtuais.....	143

LISTA DE ABREVIATURAS E SIGLAS

3D tridimensional
ACC atom-centered charges
ADME absorção, distribuição, metabolismo e excreção
ASM active site mapping
CAS Chemical Abstract Service
CADD computer-aided drug design
CatB catepsina B
CatK catepsina K
CatL catepsina L
CatS catepsina S
CoMFA Comparative Molecular Field Analysis
CP Cisteíno Protease
CPCA consensus partial analysis
csv comma-separated values
CPU central processing unit
Cz Cruzaína
FDA Food and Drug Administration
HM Hibridização Molecular
HTS High throughput screening
LBDD Ligand-based drug design
MEP Molecular Electrostatic Potential
MIF Molecular Interaction Fields
MIP Molecular Interaction Potentials
MOB Mode of Binding
NEQ Nova Entidade Química
PCA Partial Component Analysis
PDB Protein Data Bank
PLS Partial least squares
QSAR Quantitative structure activity relationship
QSPR Quantitative property activity relationship

RMSD Root Mean Square deviation

RCSB - Research Collaboratory for Structural Bioinformatics

SAR Structure-activity Relationship

SBDD Structure-based drug design

SIMCA Independent modeling of class analogies

SVD Singular Value decomposition

VSEPR Valence Shell Electron Pair Repulsion

SUMÁRIO

AGRADECIMENTOS	3
1 INTRODUÇÃO	17
1.1 INTERAÇÕES INTERMOLECULARES	19
1.2 CISTEÍNO PROTEASES	25
1.3 CAMPOS DE INTERAÇÃO MOLECULAR	34
1.4 ANÁLISE DE COMPONENTES PRINCIPAIS	40
1.5 DOCAGEM MOLECULAR	44
1.6 HIBRIDIZAÇÃO MOLECULAR	46
2 OBJETIVOS	49
2.1 OBJETIVOS ESPECÍFICOS	49
3 MATERIAIS E MÉTODOS	50
3.1 O MODO DE INTERAÇÃO NO SÍTIO ATIVO E A SELETIVIDADE	51
3.2 CAMPOS DE INTERAÇÃO MOLECULAR	53
3.3 ANÁLISE DE COMPONENTES PRINCIPAIS	59
3.4 REDOCAGEM DO CONJUNTO DE DIPEPTIDIL NITRILAS COCRISTALIZADAS	61
3.5 DOCAGEM DO CONJUNTO DE INIBIDORES DO FILTRO VIRTUAL	64
3.6 HIBRIDIZAÇÃO MOLECULAR	65
4 RESULTADOS E DISCUSSÃO	69
4.1 O MODO DE INTERAÇÃO NO SÍTIO ATIVO E A SELETIVIDADE	69
4.2 CAMPOS DE INTERAÇÃO MOLECULAR	86
4.3 ANÁLISE DE COMPONENTES PRINCIPAIS	96
4.4 REDOCAGEM DO CONJUNTO DE DIPEPTIDIL NITRILAS COCRISTALIZADAS	101
4.5 DOCAGEM DO CONJUNTO DE INIBIDORES DO FILTRO VIRTUAL	105
4.6 HIBRIDIZAÇÃO MOLECULAR	109
5 CONCLUSÃO E PERSPECTIVAS	115
REFERÊNCIAS	116
ANEXO	138

1 INTRODUÇÃO

A química medicinal é uma área que contempla a química dos processos fisiopatológicos em nível atômico. O principal objetivo é o desenvolvimento de novos fármacos (princípio ativo do medicamento, pode ser definido como substância química que causa um efeito farmacológico no sistema vivo que foi administrado¹), que entre os possíveis modos de ação, podem atuar através da formação de um complexo inibidor-alvo biomacromolecular (e.g. pequena molécula orgânica e enzima, situação encontrada nesta dissertação) modelando uma resposta farmacológica desejada. A diferença de energia livre entre o complexo e as moléculas separadas é chamada de afinidade, que está relacionada com a potência inibitória do ligante frente ao alvo, propriedade que pode ser medida através do valor de K_i (constante de equilíbrio de dissociação termodinâmico do complexo inibidor-enzima) que por sua vez é determinada pelas ligações (covalentes e não covalentes) que realiza com o alvo macromolecular no sítio de interação². Observa-se que essas energias não são aditivas e sim sinérgicas, dificultando a decomposição das contribuições das interações como termos individuais e consequentemente adicionando erros à predição da afinidade de ligação (também chamado de afinidade de interação) entre inibidor-alvo biomacromolecular³. Forças que protagonizam a interação ligante-enzima serão descritas na primeira subseção.

Outra propriedade importante a ser considerada no planejamento de novos ligantes além da potência é a seletividade: a diferença de afinidade de interação de um ligante perante diferentes alvos (e.g. enzima), responsável por modular diferentes respostas biológicas, portanto interferindo em diferentes vias metabólicas, determinando os efeitos colaterais do fármaco (produto final do planejamento, fora do escopo deste trabalho). Os alvos enzimáticos deste estudo apresentam um desafio com relação à seletividade pelo alto grau de homologia: 5 classes de proteínas cisteíno proteases (CP) da família da papaína, que serão abordados na segunda subseção.

A disponibilidade de estruturas com coordenadas atômicas dos complexos disponíveis auxiliam o planejamento de novos ligantes baseado no alvo (*structure-based ligand design*, SBDD) por sua simples inspeção visual. Além dessa análise feita e descrita na primeira subseção de Resultados e Discussão, ferramentas computacionais que auxiliem na compreensão do reconhecimento molecular entre ligante e alvo biomacromolecular são utilizadas. O planejamento de fármacos com auxílio computacional (*computer-aided drug*

design, CADD) — termo utilizado no planejamento de novos candidatos a fármaco utilizando programas computacionais⁴ — possui abordagens que auxiliem no estudo da complementaridade química entre a pequena molécula e o alvo macromolecular, incluindo aplicações da modelagem molecular⁵, onde sistemas pequenos podem utilizar cálculos quânticos, mas em sistemas maiores o custo computacional implica na utilização de métodos híbridos (Nobel de química em 2013⁶) entre mecânica molecular (utiliza-se da mecânica clássica) e mecânica quântica (considera os elétrons no sistema), onde interações moleculares ao redor do sítio ativo são tratadas de forma quântica. Programas livres ou comerciais de CADD cada vez mais possibilitam o distanciamento da química teórica e da linguagem de programação do químico medicinal, possibilitando desta forma um aumento na especialização da mão-de-obra e o aumento da necessidade de se integrar as áreas de conhecimento para fazer química medicinal, integração que cresce desde o nascimento do planejamento racional⁷. Abordagens computacionais baseadas na mecânica molecular para estudar a complementaridade química no sítio ativo do complexo (ligante-enzima) serão descritos em subseções específicas: campos de interação molecular (*molecular interaction fields*, MIF) e docagem molecular.

Avanços nas últimas décadas como por exemplo: técnicas de ensaio biológico automatizado em alta escala (*high throughput screening*, HTS), química combinatória e robótica, aumentou expressivamente o espaço químico a ser explorado⁸. Além disto, aproximadamente 200 estruturas são depositadas semanalmente na base de dados PDB⁹ (Protein Data Bank) do Research Collaboratory for Structural Bioinformatics RCSB¹⁰. Duas abordagens são indispensáveis para explorar a crescente quantidade de informações:

- a) Quimiometria¹¹, que reduz a dimensionalidade do sistema, e entre as técnicas está a análise de componentes principais (PCA), que será descrita em subseção específica desta introdução e auxiliará no estudo da seletividade enzimática no sítio ativo.
- b) Quimioinformática, que codifica as estruturas das moléculas utilizando a teoria dos grafos e relaciona suas propriedades (descritores) através de vetores. Assim torna-se possível navegar no espaço químico de forma eficiente¹². Por essas características que a quimioinformática está inserida em praticamente todo o contexto da química computacional medicinal, já que traduz a química para ser utilizada por métodos computacionais.

1.1 Interações Intermoleculares

As ligações não covalentes são todas de natureza eletrostática¹³. Porém diferentes características entre a natureza dos átomos possibilitam muitas maneiras de classificá-las. Tenta-se abordar nesta seção uma classificação no escopo da química medicinal, reunidas na Tabela 1, em que as interações no sítio ativo enzimático entre ligante e biomacromolécula precisam ser modeladas em funções para sua utilização em modelos computacionais. Mesmo inserido neste contexto há divergências em conceitos e o texto não tem o objetivo de ser exaustivo, nem de comparar os métodos, mas sim de exemplificar considerações gerais das forças que atuam na interação ligante-biomacromolécula e são utilizadas por algoritmos para prever afinidade de interação, conformação bioativa do ligante, etc.

Tabela 1 — Esquema das interações intermoleculares em química medicinal.

Tipo de interação	Desenhos esquemáticos
van der Waals	$X \cdots Y$
Interação eletrostática	$\oplus \cdots \ominus$
Ponte salina	$D^{\oplus} \cdots A^{\ominus}$
Ligação de hidrogênio	$D-H \cdots A-B$
Interação de halogênio	$D-X \cdots A-B$
Transferência de carga
Interação hidrofóbica

Notas: Interações entre átomos polarizados ($\delta^+ \cdots \delta^-$) também podem ser classificadas como interações eletrostáticas. Linhas pontilhadas representam o contato intermolecular. O esquema da linha Interação hidrofóbica almeja mostrar o efeito hidrofóbico.

Fonte: autoria própria.

Conceitualmente, interações de van der Waals são interações entre espécies químicas neutras que possuem dipolo: dipolo permanente (espécie polar), dipolo induzido (espécie

apolar em contato com espécie polar que induz a formação de um dipolo) ou dipolo instantâneo (espécies apolares, também chamada de interações de London)¹⁴. O raio de van der Waals pode ser considerado o tamanho efetivo do átomo¹⁵, ou em uma definição clássica: raio de um átomo em uma interação intermolecular do tipo de van der Waals¹⁶. Uma forma de se definir um contato de van der Waals, considerando a natureza do átomo, é mostrada na equação 1¹⁷.

$$C = D_{xy}/(R_x + R_y) \quad 1$$

Onde D_{xy} é a distância entre os átomos hipotéticos X e Y. Os símbolos R_x e R_y são os raios de van der Waals dos átomos X e Y. Um valor na relação acima de 1,3 Å não é considerado contato. Até 0,9 ainda é considerado um bom contato, porém abaixo desse valor o contato já se torna energeticamente desfavorável, repulsivo, e é considerado um impedimento estérico¹⁷.

Os valores dos raios de van der Waals utilizados nos algoritmos encontrados tanto comercialmente quanto em código aberto não divergem consideravelmente do trabalho clássico de Bondi¹⁸, que podem ser consultados na Tabela 2.

Tabela 2 — Raios atômicos em uma interação de van der Waals, unidades em angstrom.

Átomo	H	F	Cl	Br	I	O	S	N	C
Raio de van Der Waals	1,2	1,47	1,75	1,85	1,98	1,52	1,80	1,55	1,70

Fonte: Bondi, A. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, v.68, n. 3, p. 441–451, 1964¹⁸.

As funções energéticas do tipo Lennard-Jones modelam a interação de van der Waals levando em conta a natureza do átomo, um exemplo de função LJ_{12-6} ¹⁹ é exibida na equação 2:

$$E_{LJ} = 4\varepsilon([\sigma r^{-1}]^{12} - [\sigma r^{-1}]^6) \quad 2$$

O parâmetro ε é a energia em que a atração entre as espécies é máxima, r é a distância interespecies e σ é a distância interespecies em que $E_{LJ} = 0$ (zero). O termo de grau 12 modela a energia repulsiva — interação de importância para a geometria em que o confôrmero será estável²⁰ — e o termo de grau 6 a energia atrativa (energia dispersiva de London).

Conceitualmente interação eletrostática ocorre entre espécies em que ao menos uma não é neutra. As abordagens computacionais em que campos de força molecular (mecânica

molecular) são utilizados para modelar as energias de interações eletrostáticas consideram a carga elétrica parcial dos átomos pontual e fixa, muitas delas utilizando-se da lei de Coulomb como base teórica²¹, equação 3, assim, efeitos de polarizabilidade (perturbação da densidade eletrônica ocasionada por um campo elétrico) e de transferência de cargas são negligenciados ou aproximados com a inclusão de parâmetros.

$$E \propto \epsilon^{-1} q_1 q_2 r^{-1}$$

3

Na relação acima a energia potencial é proporcional às cargas das espécies (q_1 , q_2) e inversamente proporcional a: ϵ , a constante dielétrica do meio e a r , a distância interespecies.

Melhores são as previsões da função quântica, que modelam a carga elétrica através da distribuição da densidade eletrônica da espécie química, com eventuais aproximações que diminuam o custo computacional do cálculo quântico²².

Ligação de hidrogênio pode ser considerada uma específica interação dipolo-dipolo em que o polo positivo é um átomo de hidrogênio²³. A distância entre o átomo aceitador e o hidrogênio de forma geral é menor que a soma dos respectivos raios de van der Waals. Além da distância, a direcionalidade da interação é importante, geralmente a maior força de interação é atingida com a linearidade do átomos doador, o hidrogênio e o átomo aceitador²⁴. A direcionalidade do ângulo HÂB deve considerar a densidade eletrônica de B, sendo a geometria do par de elétrons no modelo de repulsão eletrônica entre os pares de elétrons na camada de valência (VSEPR) uma aproximação utilizada²⁵. Para critérios que definem a interação serem estabelecidos é necessário ângulos que indiquem sua direcionalidade, considerando a simetria do grupo do sistema. As diferenças angulares no plano em que está o par de elétrons (*in-plane*) e a diferença angular do plano que está o par de elétrons (*out-of-plane*) descrevem a direcionalidade, ilustrada na Figura 1.

Figura 1 — Ângulos que caracterizam a geometria da ligação de hidrogênio através da direcionalidade do par de elétrons.

Notas: A imagem pictórica à esquerda ilustra o ângulo que desvia do par de elétrons (P.E.) projetados em seu plano (plano em que estão contidos os pontos q e r) α (*in-plane*) e o ângulo que desvia do plano (*out-of-plane*), β . Imagem central exemplifica utilização do ângulo para o grupo funcional carbonila e imagem à direita para grupo funcional hidroxila. Ângulos na imagem esquerda mostram como eixo de referência o próprio par de elétron, em algoritmos são utilizados os eixos de simetria do sistema (imagem central e direita).

Fontes: Imagem esquerda: autoria própria. Imagem central e direita: AHMED, M. e colaboradores.

Relationship between Stereochemistry and Charge Density in Hydrogen Bonds with Oxygen Acceptors. *Crystal Growth & Design*, v. 13, n. 1, p. 315–325, 2013²⁶.

A dificuldade, no entanto, está em atribuir a localização dos pares de elétrons, que não podem ser assumidas levando em conta apenas a hibridização do átomo aceitador de elétrons, já que a formação da ligação de hidrogênio perturba o orbital 1s. Dessa forma é necessário criar modelos os quais considerem o ambiente ao redor do local que a ligação de hidrogênio seja formada²⁷.

Entre os critérios que definem ligações de hidrogênio estão os que fundamentam-se em resultados de estudos quânticos e os que utilizam estudos estatísticos em coordenadas atômicas de modelos empíricos (e.g. cristalográficos, espectroscópicos). Os grupos químicos possuem particularidades que implicam em diferentes critérios geométricos.

De forma grosseira, utilizando a nomenclatura da Tabela 1, os limites inferiores do ângulo do doador (D $\hat{H}$ A) é de 110°²⁸ e o ângulo do aceitador (H $\hat{A}$ B) é de 90°¹⁷. A distância máxima entre os átomos (H---A) é de 2,8 Å¹⁷. Os limites angulares propostos são de ligações intermoleculares entre as espécies do complexo ligante-alvo macrobiológico, mais brandos por levar em consideração os múltiplos contatos que o ligante realiza²⁹. Tratando-se de ligações envolvendo o enxofre, menos eletronegativo que oxigênio e nitrogênio, esses são fracos aceitadores e moderados doadores, portanto diferenciando-se quanto a distância e

geometria de ligação³⁰.

Pontes salinas (*salt bridges*) em química medicinal é considerado uma ligação de hidrogênio entre espécies ionizadas, e.g. cadeias laterais da arginina e glutamato³¹.

Compostos contendo o halogênio X, onde X = Cl, Br ou I; ligando-se na forma D-X---A-B; o átomo X agindo como ácido de Lewis e A como uma base de Lewis, podem realizar interação de halogênio caso a distância e a geometria dos átomos estiverem dentro de limites específicos. Parâmetros utilizados no programa Maestro (Schrödinger Release 2020-4: Maestro, Schrödinger, LLC, New York, NY, 2020) são mostrados como referência da geometria e distância deste tipo de interação: a distância máxima considerada é de 3,5 Å, o limite inferior do ângulo D $\hat{X}$ A é de 140° e o limite inferior para X $\hat{A}$ B é 90°. A base de Lewis mais comum é o oxigênio carbonílico da cadeia principal dos aminoácidos.

A Tabela 3 sumariza o trabalho realizado para explorar a interação de halogênio entre substituintes de inibidores dipeptidil miméticos na ramificação da posição P3 e o subsítio S3 da enzima catepsina L, através de um estudo de relação estrutura-atividade (*structure-activity relationship*, SAR) utilizando-se a análise de estruturas de experimentos cristalográficos de raio-X e dados de atividade biológica, suportados por cálculo computacional.

Tabela 3 — Parâmetros de série de análogos de estudo (arcabouço na parte inferior direita) que investigou ligações de halogênio em compostos dipeptidil nitrila^{32,33} à esquerda, e gráfico de cálculo quântico energético³² à direita.

A interação entre os aromáticos halogenados e o oxigênio da cadeia principal do resíduo G61 em S3 condizem com o resultado de potência dos compostos cristalizados. A sequência decrescente de atividade segue a força de interação de halogênio: iodo, bromo e cloro³⁵. O composto com flúor apresentou pior atividade que o substituinte metila, resultado que pode ser racionalizado, já que o flúor não realiza interações de halogênio³², ao menos através do modo tradicional. O composto com substituinte trifluorometil apresentou atividade inexplicada, enquanto que o substituinte tiofeno clorado parece realizar interação de halogênio, considerando-se sua distância, mas não tem um valor esperado de atividade em relação a sequência de moléculas. É proposto pelo artigo³² que tensões intramoleculares no inibidor interfiram no valor da atividade e que realmente ocorra a interação de halogênio. O

gráfico à direita na Tabela 3 mostra o resultado de estudo quântico suportando os resultados experimentais do estudo.

O elétron interespecies transferido de um orbital HOMO para um orbital LUMO é a racionalização da teoria de orbitais de fronteira para a interação intermolecular conhecida como transferência de carga. Quanto maior a energia do orbital HOMO maior o caráter doador e o contrário é válido para o orbital LUMO³⁶. Classifica-se de interação π -stacking quando as duas espécies envolvidas são aromáticas, e as subclassificações *face to face* ou *edge to face* são apenas considerações quanto à direcionalidade do ângulo entre os grupos aromáticos. Na interação π -stacking *face to face*, o limite máximo do ângulo formado entre os planos dos grupos aromáticos é de 30°. Na interação π -stacking *edge to face*, o limite inferior do ângulo é de 60°. Ou seja, a interação π -stacking não pode possuir um ângulo entre os planos formados pelos grupos aromáticos entre 30 e 60°. A distância máxima entre os centroides dos grupos aromáticos é de 4,4 Å para a primeira e 5,5 Å para a segunda. A interação π -cátion acontece entre um grupo aromático e uma carga positiva. O ângulo máximo formado entre a carga positiva e o vetor normal do plano formado pelo grupo aromático tem limite superior de 30°. A distância máxima entre a carga positiva e o centroide do grupo aromático não deve ultrapassar 6,6 Å.¹⁷ Ainda é verificada interações dessa natureza entre os elétrons delocalizados da amida e sistemas aromáticos³⁷.

Funções de interações hidrofóbicas são as mais difíceis de serem modeladas por levarem em consideração o efeito hidrofóbico, portanto a entropia de solvatação do sistema. De uma forma simplificada acontece uma interação hidrofóbica em um ambiente lipofílico, e.g. a cavidade de uma proteína, onde sua formação através do deslocamento de moléculas de água é entropicamente dirigido³⁸.

As interações não convencionais, CH/X, onde X = π , O, N, halogênios; vem recebendo crescente atenção no reconhecimento molecular entre ligante e alvo biomolecular³⁹.

1.2 Cisteíno proteases

MEROPS^{40,41} é uma base de dados que contém informações relacionadas às proteínas classificadas como peptidases (também chamadas de proteases, proteinases ou enzimas proteolíticas)⁴². Proteínas homólogas (alta similaridade da estrutura primária⁴³) são agrupadas em famílias⁴⁴. A similaridade da estrutura terciária agrupa famílias em clãs⁴⁵. Subfamílias existem para classificar grupos de proteínas que mostram divergência evolucionária antiga

dentro da família⁴⁴. Proteínas que possuem a estrutura terciária como a da proteína papaína são classificados na família C1 (“C” de “cysteine” e 1 é um caractere sequencial) dentro do clã CA (a primeira letra indica o resíduo catalítico novamente e o segundo caractere tem caráter sequencial). As enzimas, cisteíno proteases, estudadas nesta dissertação — cruzaína do protozoário *Trypanosoma cruzi* (Cz), catepsinas humanas: B (CatB), K (CatK), L (CatL) e S (CatS) — são classificadas como sendo da subfamília C1A, comumente denominada família C1A.

Cisteíno proteases são enzimas que catalisam a hidrólise de ligações peptídicas⁴⁶ (assim como outras proteases⁴⁷), com mecanismo covalente, através da cisteína catalítica. Seus substratos são polipeptídeos⁴⁸. Um esquema do substrato com a nomenclatura cunhado pelo artigo de Schechter e Berger⁴⁹ está na Figura 2.

Figura 2 — Diagrama esquemático das interações entre cadeias laterais do substrato peptídico e subsítios da proteína com nomenclatura de Schechter e Berger⁴⁹.

Notas: Ligação a ser clivada está destacada na cor vermelho. No lado esquerdo está o lado *unprime* e no lado direito o lado *primed*⁴⁸. Letra “S” significa subsítio e letra “P” refere-se às cadeias laterais dos aminoácidos. A nomenclatura também é utilizada para interações entre enzima e inibidor peptídico.

Fonte: Adaptado de: POREBA, M.; DRAG, M. Current Strategies for Probing Substrate Specificity of Proteases. *Current Medicinal Chemistry*, v. 17, n. 33, p. 3968–3995, 2010⁵⁰.

São enzimas globulares e monoméricas. Na forma madura (ativa, ou funcional) a estrutura das proteínas pode ser dividida em dois domínios, N-terminal do lado direito, onde predominam estruturas secundárias folhas-beta, e C-terminal do lado esquerdo, onde predominam estruturas secundárias hélices. Entre os dois domínios encontra-se o sítio ativo em forma da letra V⁵¹. Essa representação está ilustrada na Figura 3. Possui entre 215⁵² (cruzaína) a 261⁵³ aminoácidos (catepsina B).

Figura 3 — Estrutura tridimensional em estilo *cartoon* destacando os motivos (*motifs*) da enzima cruzaina, com adição da volta ocludente da proteína catepsina B destacada em preto.

Notas: Estrutura da cruzaina de código PDB 4QH6 representada através do estilo *cartoon* pelo programa PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Estruturas supersecundárias (motivos ou *motifs*): hélices estão na cor ciano, folhas-beta na cor magenta e as voltas estão coloridas de rosa. A cadeia lateral da cisteína catalítica está representada na forma de bastões (*sticks*) assim como a cadeia principal dos resíduos C-terminal (ao lado esquerdo) e N-terminal (lado direito). A alça colorida em preto é estrutura presente apenas na proteína catepsina B e é representada através da superposição da CatB de código PDB 1GMY⁵³.

Fonte: autoria própria.

A particularidade estrutural da CatB com relação a outras proteases da família é conhecida e amplamente descrita^{54–58} que é denominada de volta ocludente (*occluding looping*)⁵⁹, destacada na cor preto na Figura 3 para comparação com as outras proteínas da família da papaína e mostrada em detalhes na Figura 4.

Figura 4 — Singularidade da catepsina B: alça batizada de volta ocludente.

Notas: Estrutura de código PDB ID 1GMV foi utilizada para visualizar o *loop* localizado na parte superior da catepsina B, localizado no subsítio S2', colorido em preto, em que fazem partes os resíduos I105–T125. Destaca-se as pontes salinas em que os pares de resíduos ionizados: H110-D22 e R116-D224 formam a configuração denominada de fechada, responsável pela função exopeptidase da catepsina B, que aproxima os resíduos H110 e H111 ionizados do carboxilato C-terminal do substrato peptídico, propiciando a ligação ponte salina.

Fonte: autoria própria.

Essa volta exposta ao solvente possui aproximadamente vinte aminoácidos (I105 – T125)⁵⁴. Com o pH próximo de 5^{59,48} forma duas ligações iônicas entre os pares de resíduos carregados: D22-H110 e D224-R116. Esse modo de interação aproxima a volta ocludente do sítio ativo e bloqueia parte do acesso ao lado *primed* conferindo a função catalítica à enzima de exopeptidase (Figura 5), a única proteína da família que não é exclusivamente endopeptidase⁴⁰, através dos resíduos H110-H111 protonados que ligam-se ao carboxilato do resíduo C-terminal de substratos⁶⁰ assim como de inibidores.

Figura 5 — Estrutura de substrato ou inibidor no sítio ativo da catepsina B com a conformação fechada.

Notas: ligação em vermelho é a que será rompida na reação catalítica.

Fonte: adaptado de: SCHMITZ, J. e colaboradores. Cathepsin B: Active site mapping with peptidic substrates and inhibitors. *Bioorganic & Medicinal Chemistry*, v. 27, n. 1, p. 1–15, 2019.

A família C1A participa de diferentes vias metabólicas⁵⁰, como a ativação de pró-enzimas ou hormônios⁶⁰ e o processo de apoptose. O funcionamento anormal (e.g. aumento de

expressão, diferentes localizações)⁶⁰ implica no seu envolvimento em diversos processos fisiopatológicos nas células de animais vertebrados (especialmente em células cancerosas)⁶¹ e invertebrados como no caso dos tripanossomatídeos^{62,63}. A catepsina B está envolvida no rompimento da matriz extracelular e sua função no desenvolvimento de tumores é reconhecida, inclusive é utilizada como marcador biológico em seu desenvolvimento e vem sendo validado como alvo de inibição⁶⁴. Já a catepsina S possui função no sistema imunológico atuando no sistema antígeno-complexo principal de histocompatibilidade II (MHC II), promovendo o contato entre antígeno e células T CD4+⁶⁵, enquanto que a catepsina K é secretada, dentre outras células, no osteoclasto, degradando proteínas no processo de reabsorção óssea⁶⁶, portanto alvo em destaque na disfunção osteoporose.

Como catepsinas B, L e S estão associadas ao mecanismo de migração e metástase das células cancerosas^{67,68,69}, o estudo de inibidores destas enzimas tem sido realizado, inclusive em nosso grupo, com atividade citostática observada para células de câncer de pâncreas⁷⁰.

A cruzipaina é uma cisteíno protease amplamente estudada como alvo macromolecular para o estudo de novas substâncias bioativas antiparasitárias, uma vez que ela é ubiquamente expressa em todos os estágios do agente etiológico, protozoário flagelado, *Trypanosoma cruzi* (formas epimastigota, tripomastigota, amastigota) e tem um papel fundamental na sua sobrevivência e invasão na célula hospedeira, sendo um alvo macromolecular para o estudo de inibidores com atividade antichagásica⁷¹. A doença de chagas ou tripanossomíase americana, uma infecção parasitária mortal (média de 4663 morte/ano no Brasil, superando as hepatites virais) é classificada como negligenciada, o que a relaciona com condições de pobreza e intervenções terapêuticas ineficazes⁷². Os tratamentos existentes baseiam-se principalmente em dois medicamentos: nifurtimox e benznidazol, os quais têm a eficácia diminuída com o avanço de infecção⁷³. Tem como vetor os triatomíneos (inseto conhecido como barbeiro), possui 2 fases, aguda e crônica e é endêmica (delimitada por uma região e com taxa estável em número de casos⁷⁴) em 21 países das Américas⁷⁵. A Figura 6 mostra o número de casos confirmados no Brasil assim como evidencia a prevalência da região Norte (o estado do Pará possui 68% dos casos registrados no Brasil no ano de 2019) seguida da região Nordeste, as duas regiões de menor renda per capita do país⁷⁶, onde a maior incidência de transmissão é por via oral através principalmente do consumo de suco de açaí de cana-de-açúcar. O segundo maior número de casos não registrados são os que não contém relatado a via de transmissão e o terceiro é a transmissão vetorial⁷⁷.

Figura 6 — Casos confirmados da doença de Chagas aguda entre no período 2006-2019.

Notas: Gráfico mostra que a predominância em novos casos confirmados da doença de chagas na fase aguda concentra-se predominantemente na região Norte.

Fonte: Dados retirados de Tabnet-datasus:

(<http://tabnet.datasus.gov.br/cgi/defthtm.exe?sinannet/cnv/chagasbr.def>). Acesso em 15 dez. 2020.

A forma truncada da sequência usualmente estudada é a cruzaina, que contém o domínio catalítico da cruzipaina sem o domínio de 130 aminoácidos C-terminal⁷⁸.

A Tabela 4 sumariza algumas doenças em que as enzimas cisteíno proteases estão relacionadas em estudos publicados.

Tabela 4 — Exemplo de doenças relacionadas às cisteíno proteases estudadas.

Proteína	Enfermidades relacionadas
Catepsina B	Câncer pancreático e cervical ⁷⁹ Artrite reumatoide, distúrbios neurodegenerativos ⁵³ Inflamações respiratórias ⁶⁰ Alzheimer ⁸⁰
Catepsina K	Osteoporose ⁴⁸
Catepsina L	Diabete, aterosclerose, aneurisma da aorta abdominal ⁸¹ Cânceres ^{69,82,83} Alzheimer ⁸⁴ Disfunções renais, infecções virais ⁶ Doenças autoimunes, osteoporose ⁸⁵
Catepsina S	Artrite, inflamações alérgicas, câncer, diabetes ⁸⁶

Fonte: autoria própria.

As catepsinas estudadas se encontram no lisossomo principalmente⁸⁷, mas podem ser secretadas como forma de pró-enzima em outras células⁴⁶. O pH no lisossomo é considerado próximo de 5⁸⁸.

Dentre os derivados de dipeptidil nitrila estudados no grupo está o de esqueleto no centro da Figura 7, que contém o arcabouço privilegiado do inibidor K777 (cujo testes clínicos estão em andamento⁸⁹), mas sem apresentar a irreversibilidade na ligação, efeito da substituição do grupo reativo (*warhead*⁹⁰) vinil sulfona pela nitrila, mesmo grupo reativo do composto Odanacatib⁹¹, que se liga de forma covalente e reversível nas cisteíno proteases⁴⁶ e também é investigado como inibidor de catepsinas, especificamente catepsina K na disfunção osteoporose⁹², inibidor desta protease que mais avançou nas etapas clínicas⁹³ e que foi descontinuado somente após um longo estudo clínico fase 3⁹⁴.

De fato, a maioria dos inibidores da família são miméticos de dipeptidil nitrilas funcionalizados com um grupo eletrofílico⁵³. O grupo nitrila é seletivo comparando-se cisteíno proteases com serino proteases^{95,46}, além do mais se liga de forma reversível (que em contraste com inibidores irreversíveis) possui menos efeitos colaterais no organismo em sistemas *in vivo*⁹².

Figura 7 — Protótipo mimético de dipeptidil nitrila estudado no grupo NEQUIMED ao centro da imagem e moléculas que inspiraram a respectiva série congênere localizadas nas laterais.

Notas: À esquerda K-777, à direita Odanacatib. Ao centro encontra-se o esqueleto molecular de dipeptidil nitrilas estudadas no grupo com grande diversidade de grupos P1, P2 e P3.

Fonte: autoria própria.

Na reação com o inibidor, Figura 8, o carbono da nitrila (grupo reativo eletrofílico das moléculas do trabalho) está na no mesmo “nível de oxidação” (termo de uso particular pela fonte consultada) — aproximando as características de reatividade — do carbono carbonílico da amida⁹⁶. Possuindo eletrofilicidade semelhantes na reação catalítica são atacados nucleofilicamente pelo tiolato da cisteína catalítica — formado através da abstração do próton pelo nitrogênio da histidina catalítica, resíduo que, através de seu outro nitrogênio, faz interação do tipo ligação de hidrogênio com o oxigênio amídico da cadeia lateral da asparagina catalítica, desta forma participando da estabilização do nucleófilo tiolato. Esses três resíduos mencionados são conhecidos como tríade catalítica das cisteíno proteases⁹⁷.

Figura 8 — Reação catalítica de cisteíno proteases com composto dipeptídil (identificada genericamente pela esfera preta) funcionalizada com grupo reativo nitrila (cinza).

Notas: numeração dos resíduos citados nas notas específicas são da proteína Cz PDB 4QH6.

Notas específicas:

Imagem A) Interações do tipo ligação de hidrogênio no resíduo catalítico H162 com N182 e com C25 resultam na formação de um par iônico entre H162 e C25⁹⁸.

Imagem B) Estabilidade da histidina na forma catiônica (+1) tem auxílio do oxigênio da cadeia lateral de N182 participando como aceitador de ligação de hidrogênio. O enxofre carregado negativamente torna-se nucleofílico o suficiente para um ataque ao carbono eletrofílico da nitrila, nesta imagem é mostrada as etapas do mecanismo de forma concertada, desta maneira não é mostrado o estado de transição tetraédrico carregado negativamente, que interage com a cadeia lateral do resíduo de Q19³² que ajuda sua estabilização⁹⁹. O nitrogênio da nitrila abstrai um próton do nitrogênio da H182 voltando assim os resíduos a ficarem neutros com a formação da ligação covalente tioimidato.

Imagem C) Formação do complexo covalente no sítio catalítico.

Fonte: Adaptado de: SILVA, J.R.A. e colaboradores. Assessment of the Cruzain Cysteine Protease Reversible and Irreversible Covalent Inhibition Mechanism. *Journal of Chemical Information and Modeling*, v. 60, n. 3, p. 1666–1677, 2020.¹⁰⁰.

A formação da ligação covalente reversível⁵³ entre o enxofre e o carbono acontece após a formação do intermediário carregado negativamente (após ataque do tiolato ao grupo nitrila) que possui a estabilidade auxiliada pela cadeia lateral de um resíduo de glutamina (Q19, PDB 4QH6) que é denominado juntamente com o nitrogênio da cadeia principal da

cisteína catalítica, que também participa da estabilização através de ligações de hidrogênio, de bolsão do oxiânion⁵³ (*oxyanion hole*).

As dipeptidil nitrilas sintetizadas no grupo NEQUIMED, vem apresentando atividade inibitória para a cruzaína, além de terem apresentados resultados promissores em ensaios celulares associados^{52,101,102,103,104}.

1.3 Campos de Interação molecular

Quimioinformática pode ser definida como a área da química que codifica dados estruturais moleculares e suas propriedades para manipulação em algoritmos¹². O conceito de propriedade química em quimioinformática se expande para o termo descritor molecular¹⁰⁵, que além de propriedade molecular, relaciona propriedade de subestruturas e propriedades de coleções de compostos¹⁰⁶. Exemplo de classe de descritor bidimensional (2D) é o topológico, que utiliza teoria do grafo (objeto de estudo é o grupo e suas conexões) para retirar informações relacionadas a conectividade da molécula¹⁰⁷. O descritor MIF é um descritor tridimensional (3D), pois associa-se à estruturas 3D, que faz o mapeamento energético na molécula objeto de estudo (pequena molécula ou biomacromolécula) através do cálculo de energia de interações intermoleculares com sondas químicas — átomo, grupo químico ou ainda um pseudo-átomo com propriedades físico-químicas parametrizadas — resultando em uma matriz energética que pode ser utilizada para: visualizar mapas de potencial em conjunto com a estrutura, podendo guiar de forma direta a modificação de um composto³⁹ ou podem ser utilizados em análises quimiométricas¹⁰⁸.

O GRID¹⁰⁹ é um dos principais programas que utiliza o descritor MIF. Um prisma retangular (definido através das coordenadas dos vértices), usualmente chamado de caixa, delimita a área do cálculo. Dentro do volume especificado são definidas coordenadas igualmente espaçadas (padrão de 1 Å) onde as sondas químicas selecionadas são posicionadas para o cálculo da energia de interação intermolecular com a molécula alvo.

O termo GRID também se refere ao campo de força do programa, calibrado com dados experimentais¹⁰⁹. Átomos de hidrogênios apolares, portanto que não participam de ligações de hidrogênio, estão contidos em unidades com um átomo pesado, chamado na teoria de campo de força de átomo-unidos (*extended atom* ou *united-atom*)¹¹⁰, e.g. grupo metileno, conceito também usado no campo de força CHARMM¹¹¹. Porém, quando átomos da estrutura alvo não são reconhecidos como um átomo parametrizado pelo método, os parâmetros dos

hidrogênios são considerados explicitamente. O átomo principal da sonda química tem livre rotação para encontrar as interações intermoleculares mais favoráveis para as ligações de hidrogênio (caso a sonda virtual possua essa propriedade) em função da distância e ângulo. As funções de energia são calibradas para interações que geralmente ocorrem no sítio ativo entre ligante e alvo bioreceptor, descritos a seguir de acordo com 3 artigos publicados no Journal of Medicinal Chemistry por Rebecca Wade e Peter Goodford^{112–114}, a fonte mais atual que descreva as funções energéticas do campo de força GRID.

A função de van der Waals 12-6 ou Lennard-Jones utilizada é representada pela equação 4:

$$E_{LJ} = Ar_{it}^{-12} - Br_{it}^{-6} \quad 4$$

Onde r_{it} é a distância entre o ponto i e a unidade atômica t . Os parâmetros A e B são parâmetros que levam em conta a natureza da unidade atômica: raio de van der Waals, polarizabilidade e o número efetivo de elétrons.

A função de energia eletrostática está representada pela equação 5:

$$E_{el} = \frac{pq}{KM} \left(\frac{1}{d} + \frac{M-W}{(M+W)\sqrt{d^2+4PQ}} \right) \quad 5$$

As cargas eletrostáticas da sonda química e da unidade atômica são representadas por p e q , respectivamente, e estão separadas por uma distância d . A variável K considera fatores geométricos. As variáveis M e W representam constantes dielétricas da cavidade da proteína e da água (superfície da proteína) e P e Q representam a profundidade das cargas p e q quando estas se encontram na cavidade da macromolécula.

A função energética da ligação de hidrogênio pode ser descrita pela equação 6¹¹⁴:

$$E_{HB} = E_r E_t E_p \text{ para } 0 < E_t < 1 \text{ e } 0 < E_p < 1 \quad 6$$

Onde As funções E_r , E_t e E_p são funções relacionadas aos parâmetros geométricos r , t e p que simbolizam respectivamente: distância entre átomo pesado da espécie a ser mapeada (*target*) e a sonda química virtual (*probe*); ângulo relacionado a *target*; ângulo relacionado a *probe*. A equação 6 tem o suporte da Figura 9, que suas notas tem descrições mais

detalhadas dos ângulos t e p quando *target* doa e aceita elétrons.

Figura 9 — Definição de ângulos utilizados como parâmetros na função energética referente a ligação de hidrogênio do campo de força GRID.

Legendas: T simboliza *target*, o átomo da espécie a ser mapeada. A letra H simboliza o átomo de hidrogênio. A letra P simboliza *probe*, a sonda química virtual. A letra r simboliza a distância entre T e P. Na imagem A, *target* atua como doador de ligação de hidrogênio e a letra t simboliza o ângulo suplementar a T $\hat{H}$ P. A letra p simboliza o ângulo entre a reta que contém a direção do par de elétrons da sonda e a reta formada pelo hidrogênio e por P. Na imagem b, T (*target*) atua como aceitador de ligação de hidrogênio, e o ângulo t é formado entre a direção das retas TP e a direção do par de elétrons. A letra p é o ângulo suplementar a P $\hat{H}$ T. No caso em que é possível estabelecer o plano em que os pares de elétrons se encontram, como no exemplo da imagem C, define-se dois ângulos, t_o e t_i : *out-of-plane* e *in-plane* (consultar a seção introdução para mais informações).

Fonte: BOOBYER, D.N.A. e colaboradores. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *Journal of Medicinal Chemistry*, v. 32, n. 5, p. 1083–1094, 1989.

A função de distância E_R , variável na função de ligação de hidrogênio (equação 6), é dependente da distância entre a sonda e o átomo *target* r e é mostrada na equação 7.

$$E_R = C/r^8 - D/r^6 \quad 7$$

A função E_R também depende dos parâmetros C e D, que são dependentes da natureza do átomo, são mostradas nas equações 8 e 9:

$$C = -3E_m r_m^8 \text{ kcal}\text{\AA}^8/\text{mol} \quad 8$$

$$D = -4E_m r_m^6 \text{ kcal}\text{\AA}^6/\text{mol} \quad 9$$

Onde E_m é o mínimo energético na distância ótima interatômica de interação.

A função E_t quando o átomo do alvo investigado (*target*) é doador só depende do valor de ângulo t , com restrições impostas ao ângulo t_i . Quando o átomo é aceitador, a função E_t pode depender de t_i , t_o e t , dependendo da natureza do átomo aceitador e assume diversas formas.

Já a função angular E_p , quando faz somente uma ligação de hidrogênio, recebe o valor 1, pois a sonda tem livre rotação para orientar seu par de elétrons ou hidrogênio, direcionando-o para possuir ângulo nulo. Quando mais de uma ligação de hidrogênio é realizada a função assume diferentes formas.

A energia da interação de um par sonda-unidade atômica é mostrada na equação 10.

$$E_{\text{pair}} = E_{\text{VDW}} + E_{\text{EL}} + E_{\text{HB}} \quad 10$$

A soma energética total para uma sonda específica em uma coordenada específica é discriminada na equação 11, onde todas interações estabelecidas entre os pares sonda-unidade atômica são consideradas.

$$E_{\text{TOTAL}} = \sum_{j=1}^{j=t} (E_{\text{LJ}} + E_{\text{EL}} + E_{\text{HB}}) + S \quad 11$$

O termo S é um valor que modela a entropia, dependente apenas da coordenada, que considera a hidrofobicidade do ambiente. O índice j representa a j -ésima unidade atômica do alvo e t a unidade atômica do alvo mais distante dentro do limite de cálculo estabelecido pelos limites estabelecidos pela caixa.

Os parâmetros atribuídos aos átomos são: raio de van der Waals em Å, carga atômica — que utiliza do conceito carga centrada no átomo (ACC - *atom-centered charges*) — em unidades u.c.a. (unidade de carga atômica), número efetivo de elétrons, polarizabilidade do átomo em Å³, além dos parâmetros de ligação de hidrogênio: (i) energia mínima da ligação de hidrogênio considerando a distância ótima de interação em unidade kcal.mol⁻¹; (ii) números possíveis de ligação de hidrogênio (aceitadora e doadora) e (iii) geometria que faz com o hidrogênio ligado covalentemente e geometria da ligação de hidrogênio¹¹⁵.

Os valores de energia obtidos podem ser visualizados por meio de interface gráfica. A escolha de valores de corte para a energia destacam regiões de interesse, ilustrado na imagem da Figura 10.

Figura 10 — Método de análise gráfica do mapeamento energético feita pelo programa GRID através da visualização dos campos de interação molecular com respectiva estrutura molecular.

Notas: Visualização de campos de interação molecular por interface gráfica para o aminoácido leucina com as sondas: hidrofóbica (amarela - A), amida (vermelha - B) e NH_3^+ (azul - C).

Fonte: WADE, R.C. Calculation and Application of Molecular Interaction Fields. In: Cruciani, G. (Ed.); *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*. Weinheim: Wiley-VCH Verlag, 2006, p. 27–38¹¹⁶.

Essas imagens tridimensionais ajudam a elucidar sítios energéticos favoráveis a interações intermoleculares específicas, possibilitando então o estudo do complexo ligante-alvo macromolecular no sítio ativo, podendo ser aplicada em planejamentos de fármacos baseado no alvo (*structure-based ligand design*, SBDD), a primeira e a ainda utilizada função do software¹⁰⁹. Dessa forma é um método utilizado no planejamento de novos ligantes, seja na fase inicial de descoberta, assim como em sua otimização, auxiliando na gênese planejada de fármacos⁷. O programa foi o pioneiro em utilizar uma maior diversidade de sondas químicas que mimetizam as propriedades físico-químicas de grupos químicos usados para o cálculo das energias de interações intermoleculares¹¹⁶.

Apesar do termo MIF (conhecido também por *molecular interaction potentials*, MIP¹¹⁷) não ter sido definido no artigo que apresenta o software GRID e o respectivo campo de força, esta é a publicação pioneira que apresenta a estratégia do método e sua aplicação¹⁰⁹. Neste período já havia trabalhos voltados a desenvolvimentos de moléculas bioativas utilizando-se as estruturas cristalográficas de alvos macromoleculares, porém a visualização simultânea da estrutura tridimensional e dos campos de interação provê um método alternativo para descrição das interações entre as espécies. Neste artigo é feita uma comparação entre as interações intermoleculares das estruturas cristalográficas de proteínas com respectivos ligantes no sítio de ligação, que são visualmente inspecionadas e com o cálculo energético do GRID, promovendo de certa forma os resultados de predição das interações pelo software, onde o mapeamento energético com sondas específicas coincidiram com fragmentos do inibidor complexado no sítio de ligação da proteína.

Um caso de sucesso de utilização do software GRID foi publicado na revista *Nature* em

1993¹¹⁸, em que substituições no já potente inibidor de sialidase Neu5Ac2en, Figura 11A, resultou no análogo 4-guanidino-Neu5Ac2en, Figura 11B, mais potente, posteriormente aprovado como fármaco, ZANAMIFIR¹¹⁹ (medicamento RELENZA¹²⁰), para a prevenção e tratamento para Influenza a e b. A substituição de um grupo hidroxila por uma guanidina foi racionalizada através do método utilizando-se como sonda virtual uma amina primária protonada, que resultou em uma superfície de contorno energética favorável na região do sítio de ligação da enzima. Este resultado levou a cristalização do complexo ZANAMIFIR-neuraminidase¹²¹, Figura 11C.

Figura 11 — Modificação estrutural de hidroxila (composto A) por grupo guanidina (composto B) guiada através de análise no modo de interação e mapeamento energético com campos de interação molecular. A imagem C mostra o composto B cristalizado com a enzima neuraminidase.

Notas específicas: C) ZANIMIFIR complexado com enzima neuraminidase, PDB 1NNC¹²¹, onde o grupo guanidina realiza ligações de hidrogênio (W178, D151) e possíveis interações iônicas: E227 e E119.

Fonte: adaptado de: ITZSTEIN, M. VON e colaboradores. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, v. 363, n. 6428, p. 418–423, 1993.

A estrutura cristalográfica confirmou a predição feita pelo software, que identificou a região no sítio catalítico propícia a serem ocupadas por nitrogênios básicos.

Recentemente, 2017, Tosco e Mackey escreveram um capítulo de livro² dedicado ao tópico MIF e destacam tanto aplicações clássicas quanto inovações em sua aplicação, tanto em LBDD — similaridade, predições de conformação, *virtual screening*, *scaffold hopping*, 3D-QSAR e propriedades farmacocinéticas de absorção, distribuição, metabolismo e excreção (ADME), quanto em SBDD: identificação e similaridade de sítios de ligação e identificação de moléculas de água provenientes de estrutura cristalográficas são exemplos de tópicos abordados, sendo um material de consulta não exaustivo porém atualizado.

Será descrito na próxima seção o mapeamento energético em análises quimiométricas.

1.4 Análise de componentes principais

A química medicinal está na era da *big data*¹²². O espaço químico tanto real quanto virtual está distante de limites. Patrick Walters em seu artigo de revisão intitulado “Virtual Chemical Libraries”¹²³ apresenta as fundamentações para essa afirmação: em 2015 a empresa *Chemical Abstract Service* (CAS) já registrava 100 milhões de compostos registrados. A base de dados ChEMBL¹²⁴ (<https://www.ebi.ac.uk/chembl/>), especializado em informações para área de química medicinal, possuía nessa data mais de dois milhões de compostos. Números pequenos comparados com bibliotecas de compostos virtuais disponíveis para compra, ZINC¹²⁵ com aproximadamente 980 milhões de compostos. Dados curados do genoma humano apresentam um acelerado crescimento, impactando na quantidade de estruturas biomacromoleculares depositadas em repositórios para consulta (que também deve-se a outros fatores, e.g. tecnológico), ou na modelagem de macromoléculas ainda não resolvidas. A abundância de informação indica que a habilidade de manipular os dados é essencial¹⁰⁶.

A quimiometria, onde a grande quantidade de informação pode ser manipulada através de modelos estatísticos-matemáticos multivariáveis que reduzem a dimensionalidade do sistema, culminando na interpretação de sistemas químicos complexos, pode ser aplicada em praticamente todas as áreas da química, com grande relevância para a química medicinal. Modelos multivariados preditivos são utilizados em relações quantitativas de estrutura-atividade (*quantitative structure-activity relationship*, QSAR) e relações quantitativas de estrutura-propriedade (*quantitative property activity-relationship*, QSPR). O nascimento de QSAR¹²⁶ na década de 60 é um marco histórico na química medicinal¹²⁷, pois foi o início do uso de modelos matemáticos para a racionalização da compreensão da atividade biológica por meio das propriedades físico-químicas dos compostos relacionando atividade biológica com descritor hidrofóbico, que ficou conhecido como método extratermodinâmico¹²⁶. O software *Comparative Molecular Field Analysis* (CoMFA)¹²⁸ é um método de QSAR 3D aplicado em planejamento de fármaco baseado no ligante (*ligand-based drug design*, LBDD) usando o método dos mínimos quadrados parciais (*partial least squares*, PLS) a partir dos valores de energia obtidos com funções da mecânica clássica. Neste método, os principais descritores tridimensionais foram provenientes das energias resultantes da função van der Waals 12-6 (Lennard-Jones) e o potencial eletrostático, onde a sonda química que interage com a

molécula é uma carga positiva, gerando o descritor 3D: potencial eletrostático molecular (*Molecular Electrostatic Potential*, MEP), ainda em amplo progresso na modelagem de interações eletrostáticas intermoleculares²². Além do modelo quantitativo é possível ainda visualizar superfícies de energia através do MEP em conjunto com a estrutura molecular.

A análise de componentes principais (*principal component analysis*, PCA) é uma análise estatística multivariável, classificatória, utilizada para se reduzir a dimensionalidade do sistema, onde as variáveis no espaço amostral original são usadas para a composição de vetores para cada componente principal que, por definição, são ortogonais entre si^{129,130}. A ortogonalidade entre as novas variáveis (denominadas componentes principais) elimina o problema da covariância que pode existir nas variáveis originais¹³¹ (no caso deste trabalho, as energias das interações intermoleculares calculadas no programa GRID).

A PCA é comumente aplicada para estudar a similaridade de amostras e influência das variáveis do modelo em tal separações. A matriz original de dados X contém as amostras organizadas em linhas e as variáveis organizadas nas colunas. O modelo estatístico decompõe ou fatora a matriz de dados Y em duas matrizes conforme a equação 12¹³¹:

$$X = X_2 + E = PT \quad 12$$

Em que X é a matriz de dados, P é a matriz de pesos (*loadings*) e T é a matriz de pesos (*scores*), X_2 é matriz truncada sem resíduos (erros aleatórios) e E é a matriz de resíduos.

Existem vários algoritmos para cálculo de PCA¹³². Um dos mais utilizados envolve a técnica de decomposição vetorial oriunda da álgebra linear denominada decomposição de valores singulares (*singular value decomposition*, SVD), onde uma matriz é decomposta em três fatores (matrizes) segundo a equação 13¹³³:

$$X = VDU \quad 13$$

Onde V e U são matrizes de vetores ortogonais classificadas como autovetor esquerdo e direito respectivamente. A matriz D é diagonal e classificada como de valores singulares da matriz X.

A matriz V contém as observações da matriz X em um sistema de base referencial de vetores rotacionados para o maior valor de variância dos dados que é resultado da combinação linear dos vetores do sistema original¹³¹. Na correspondência com a

nomenclatura de matrizes da PCA a matriz V corresponde a matriz P e é chamada matriz de pesos (*loadings*).

A matriz diagonal D de valores singulares contém as variâncias dos autovetores, que compõem o novo sistema de coordenadas, e o produto das matrizes D e U é igual a matriz T na nomenclatura da PCA, chamada matriz de *scores*. A matriz de *scores* contém os dados da amostra combinadas linearmente na nova base de coordenadas; e o gráfico de seus vetores produz pontos de coordenadas que representam as amostras. A partir da visualização dos gráficos de seus vetores é possível verificar o agrupamento das amostras e a similaridade que é resultado das contribuições dos *loadings*. Com a interpretação conjunta dos gráficos/valores das matrizes dos *loadings* e *scores* é feita a atribuição de quais variáveis do modelo são responsáveis pelas separações das amostras.

Ao se utilizar uma nova base de coordenadas rotacionadas decrescentemente com relação a variância das observações, torna-se possível fazer a redução de dimensionalidade do sistema, aumentando o poder de interpretação dos dados através da análise gráfica. Observando a matriz D é possível escolher a quantidade de vetores que representará uma porcentagem de variância adequada para o estudo do modelo, na maioria das vezes escolhidos estrategicamente com duas ou três dimensões para interpretação gráfica que os gráficos de *scores* e *loadings* proporcionam de forma conjunta, ilustrados na Figura 12, retirada de uma didática revisão do método¹³⁰, aplicado em química analítica.

Figura 12 — Resultados gráficos da PCA: gráfico de *scores* no lado esquerdo e gráfico de *loadings* no lado direito, em estudo que analisou vinhos de 4 países com relação a 14 propriedades físico-químicas.

Notas: Legenda sobre origem do vinho está no canto superior direito da imagem à esquerda.

Fonte: BRO, R.; SMILDE, A.K. *Principal component analysis*. Analytical Methods, v. 6, n. 9, p. 2812–2831, 2014.¹³⁰

Vinhos de diferentes países foram analisados quanto a diferentes propriedades. Como um exemplo de direta interpretação está a constatação que vinhos chilenos possuem valores de PC2 grandes e negativos através do gráfico de *scores*, implicando que amostras de vinhos chilenos em geral possuem grandes quantidades de ácido málico e baixas quantidades de ácido láctico, ácidos totais, etanol e uma grande volatilidade, propriedades identificadas no gráfico de *loadings*.

O estudo de similaridade entre sítios catalíticos de alvos biomacromoleculares é qualidade requerida no desenvolvimento de fármacos: objetiva a minimização de efeitos colaterais ou a racionalização da reatividade-cruzada no planejamento de novos candidatos a fármacos¹³⁴, e pode ser analisado através da PCA. A análise de seletividade no sítio de ligação de proteínas foi estudada¹³⁵ com a utilização de valores de energia de MIF como variáveis independentes de uma PCA usando estruturas cristalográficas da enzima diidrofolato redutase ortólogas humana e bacteriana *Escherichia coli*. Os resultados apresentam locais no sítio e possíveis grupos moleculares importantes para obtenção de seletividade requerida para as enzimas.

Campos de interação molecular foram utilizados em serino proteases da família da tripsina e analisadas através do método *consensus partial component analysis* (CPCA). O algoritmo escalona os blocos de variáveis, possibilitando assim que sondas químicas com

distintas magnitudes energéticas tenham o mesmo peso para o modelo. interações íon-dipolo e ligações de hidrogênio possuem valores de energia que sobrepõem a importância as interações mais fracas, como as de van der Waals.¹³⁶ Uma revisão do método foi feita por Montanari e colaboradores¹³⁷, que já havia utilizado o método para destacar pontos favoráveis de interação no subsítio S2 da cruzaina¹³⁸.

Os campos de interação molecular em conjunto com quimiometria foram usadas em diversos trabalhos realizados por membros do grupo de pesquisa, como o estudo de receptores de estrogênio e seus ligantes^{139,140}, o qual foram estudadas as isoformas alfa e beta explorando a seletividade dos sítios destas enzimas e também dos compostos bioativos.

1.5 Docagem molecular

O modo de interação (*mode of binding*, MOB) entre ligante e alvo biomacromolecular no sítio ativo é estudada quando se deseja planejar uma nova molécula. Se disponível a estrutura do alvo biomacromolecular (depositada ou modelada com uma estrutura de preferência com alta similaridade sequencial¹⁴¹), este sendo um pré-requisito para qualidade do modelo¹⁴², uma alternativa é a docagem molecular¹⁴³, técnica computacional que busca prever o modo e a afinidade de ligação do complexo¹⁴⁴. O encaixe do ligante ao alvo realizada com movimentos translacionais na molécula e movimentos rotacionais nas ligações, manualmente, observando-se a complementaridade química com posterior cálculo de energia do complexo, seria o uso não automatizado da técnica⁷. A definição atual do método computacional em CADD é de dois algoritmos que respectivamente geram e pontuam as *poses* — estruturas únicas de ligantes que são definidas por duas variáveis: vetor posição (com referencial na macromolécula alvo) e conformação^{145,146} — no sítio ativo do alvo macromolecular¹⁴⁷.

De modo pictórico pode-se caminhar da não flexibilidade do alvo biológico, assemelhando-se a teoria enzimática clássica da chave-fechadura¹⁴⁸, aumentando-se o modo de se considerar a flexibilidade no alvo, aproximando-se do modelo bioquímico do encaixe induzido¹⁴⁹. A flexibilidade pode ser aumentada relaxando parâmetros estéricos; utilizando mais estruturas da mesma enzima ou até mesmo incluindo graus de liberdade da macromolécula no cálculo¹⁵⁰.

O algoritmo de busca (que geram as *poses*)¹⁴², adotam diferentes estratégias para realizar a tarefa, mas todos objetivam encontrá-las através da minimização energética¹⁵¹. Os

algoritmos de busca tipicamente são classificados¹⁵⁰ em: busca sistemática — em que todo espaço químico é explorado; busca estocástica — onde as mudanças nos graus de liberdade da molécula são randômicas¹⁵²; método determinístico — simulações para encontrar confôrmeros de menor energia¹⁵³, e.g. métodos de minimização energética¹⁵³.

As funções de pontuação baseadas em campo de força (*force field*) utilizam funções da mecânica molecular, especificamente termos da função que descrevem a interação de átomos não ligados covalentemente. Quando os coeficientes das funções são dados empíricos e modelados através de regressão linear dá-se o nome de algoritmo empírico. Caso as funções e parâmetros utilizados para modelar o algoritmo forem distribuições estatísticas obtidas de dados experimentais diz-se que o algoritmo é baseado no conhecimento (*knowledge-based*). Essa classificação apesar de tradicional pode ser considerada obsoleta por tornar confuso a classificação de métodos híbridos e não incluir algoritmos no estado da arte do tópico, e.g. aprendizado de máquina para modelar as funções¹⁵⁴.

O ranqueamento dos complexos docados podem visar tanto ligantes diferentes (método inicial em triagem virtual¹⁵⁵) assim como *poses* de um mesmo ligante (predição de *poses*) e são tarefas com características distintas¹⁵⁶, ambas ainda permanecem em desenvolvimento e ainda não são métodos exatos, ou acurados^{152,157}, em partes pelas várias aproximações necessárias para manter o bom custo computacional do método¹⁵⁴. A preparação do sistema aumenta a predição: e.g. conhecimento do sítio ativo, estados de protonação do alvo enzimático¹⁵⁸, restrições conformacionais aplicadas aos ligantes¹⁵⁹. Assim como utilizar pontuação consenso, que combina resultados provindos de mais de uma função de pontuação, otimizando a classificação das moléculas docadas^{158,160}.

O principal programa de docagem utilizado nesta dissertação, FRED^{161,162}, possui um sistema de busca sistemático classificado como exaustivo, percorrendo assim todo espaço químico da molécula a ser docada variando os graus de liberdade translacional e rotacional da estrutura rígida dos confôrmeros pré-gerado, desta maneira, implicitamente considera a flexibilidade do ligante¹⁶¹. A função de pontuação utiliza funções gaussianas com termos que calculam a energia de van Der Waals, ligações de hidrogênio e interação com metal¹⁶³.

Já o programa GOLD, também utilizado neste estudo, possui o sistema de busca estocástico, subclassificado como algoritmo genético, o qual inspira-se em eventos da genética. Cada cromossomo codifica parâmetros do par ligante-receptor possibilitando sua pontuação por um conjunto de funções, a qual busca refletir a energia de interação do

complexo formado. Na busca por otimizar as soluções, modificações nos cromossomos são realizadas de forma randômica porém ponderadas, exemplos são: troca entre as moléculas dos cromossomos (*crossing-over*) e modificações (mutações) nos parâmetros específicos do cromossomo^{164,165}.

1.6 Hibridização molecular

Uma das estratégias em QM que pode ser utilizada no desenho de novos precursores é a hibridização molecular (HM), técnica que utiliza partes de moléculas para formar uma nova entidade química (NEQ)¹⁶⁶. “As partes” são fragmentos estruturais¹⁶⁷, ou expandindo a definição, grupos farmacofóricos¹⁶⁸ — que podem ser definidos não apenas a partir da estrutura mas a partir das propriedades químicas atribuídas às interações com o alvo macromolecular¹⁶⁹.

Sorafenib, inibidor de proteínas quinases, possui o fragmento ureia diarílica. Sua aprovação como fármaco pela agência *Food and Drug Administration* (FDA) em 2005 para o tratamento de câncer¹⁷⁰ vem encorajando a utilização desta estrutura em NEQs com atividade antitumoral¹⁷¹. Raloxifeno, fármaco aprovado para o tratamento de osteoporose pós-menopausa com efeito preventivo no câncer de mama¹⁷² possui o fragmento benzotiofeno utilizado em compostos para modular a atividade biológica em várias disfunções, sendo considerado uma estrutura privilegiada no planejamento de fármacos¹⁷³. O composto na Figura 13 é produto de SAR que utilizou do conceito de HM nos fragmentos privilegiados resultando em melhores resultados *in vitro* do que Sorafenib, quando testados em linhagens de células de câncer: HT-29¹⁶⁸.

Figura 13 — Exemplo do planejamento de híbrido molecular (ao centro) utilizando subestruturas dos fármacos Raloxifeno e Sorafenib através de estudo que utilizou dentre as ferramentas a hibridização molecular.

Fonte: adaptado de ZAREI, O. e colaboradores. Design, synthesis, and biological evaluation of novel benzo[b]thiophene-diaryl urea derivatives as potential anticancer agents. *Medicinal Chemistry Research*, v. 29, n. 199, p. 1438–1448, 2020¹⁶⁸.

Um exemplo clássico, utilizado inclusive de forma didática para mostrar o conceito de HM, é a combinação de dois fármacos, acetaminofeno (paracetamol): analgésico e antipirético¹⁷⁴ e ácido acetilsalicílico: anti-inflamatório, analgésico e antipirético¹⁷⁴, que serviram como estratégia para o planejamento do fármaco Benorilato¹⁷⁵, abordagem mostrada na Figura 14.

Figura 14 — Estudo de hibridização molecular que resultou no fármaco Benorilato.

Fonte: adaptado de: ARAÚJO, C.R.M. e colaboradores. Desenvolvimento de fármacos por hibridização molecular: uma aula prática de química medicinal usando comprimidos de paracetamol e sulfadiazina e a ferramenta virtual. *Química Nova*, v. 38, n.6, p. 868-873, 2015¹⁶⁷.

A Tabela 5 ilustra recorte de trabalho realizado em 2018¹⁷⁶, o qual pesquisadores reportaram estudo de inibidores da enzima N-miristoiltransferase, presente tanto no protozoário *Trypanosoma Brucei* — alvo da doença Tripanossomíase africana (também conhecida como doença do sono) — quanto nos humanos. Entre os resultados alcançados está o inibidor 3, que através da hibridização molecular, aliou a potência do inibidor 1 à seletividade do inibidor 2.

Tabela 5 – Estudo com inibidores parasitário e humano da enzima N-miristoil transferase utilizando hibridização molecular confere seletividade ao composto 3.

			ID	IC ₅₀ _{Tb} (nM)	IC ₅₀ _{Hs} (nM)
1	< 5	62
			2	1.10 ³	>1.10 ⁵
			3	< 5	2,3.10 ⁵

Legenda: No cabeçalho IC₅₀ é metade da concentração inibitória máxima, Tb significa o organismo protozoário *Trypanosoma brucei* e Hs o organismo *Homo sapiens*.

Fonte: HARRISON, J.R.e colaboradores. A Molecular Hybridization Approach for the Design of Potent, Highly Selective, and Brain-Penetrant N-Myristoyltransferase Inhibitors. *Journal of Medicinal Chemistry*, v. 61, n. 18, p. 8374–8389, 2018¹⁷⁶.

2 OBJETIVOS

Utilizar hibridização molecular no planejamento de inibidores de cisteíno proteases da família C1A (cruzaína do protozoário *Trypanosoma cruzi* e catepsinas humanas: B, K, L e S), considerando a seletividade inibitória dentro do grupo a propriedade a ser otimizada, através de estudo da complementaridade química entre proteína (i) e ligante (ii).

(i) Mapear energeticamente os sítios ativos das enzimas por meio de estudo de descritores moleculares MIF e posterior análise quimiométrica (PCA), identificando regiões que possuem seletividade para tipos específicos de interações intermoleculares.

(ii) Docar inibidores miméticos de dipeptidil nitrilas com elevada potência reportados na literatura (ChEMBL) o que possibilita o estudo da complementaridade química entre inibidor e sítio ativo.

2.1 Objetivos específicos

Cálculo de campos de interação molecular no sítio ativo das enzimas: cruzaína e catepsinas B, L, K e S;

Encontrar regiões seletivas no sítio ativo para o grupo de enzimas estudadas utilizando os dados energéticos em uma análise de PCA;

Estudar os inibidores ($pK_i > 7$) miméticos de dipeptidil nitrilas disponíveis na base de dados virtual ChEMBL utilizando docagem molecular;

Relacionar os resultados encontrados através de análise da complementaridade química entre enzima e inibidor;

Utilizar hibridização molecular para o planejamento de novos inibidores.

3 MATERIAIS E MÉTODOS

O trabalho computacional foi feito em dois servidores Linux. O primeiro possui 12 unidades centrais de processamento (*central processing unit*, CPU)) Intel® Xeon® 5680 3,3 GHz com arquitetura x86_64 com 16 GB de memória de ram e memória de armazenamento de 4 TB. Possui uma placa gráfica Matrox Electronics Systems Ltd. MGA G200eW WPCM450 (rev 0a) e uma placa gráfica NVIDIA Corporation GF100GL [Tesla C2050 / C2070] (rev a3).

O segundo servidor possui 16 CPU Intel® Xeon® E5620 2,4 GHz com arquitetura x86_64 com 56 GB de memória ram e memória de armazenamento de 6 TB. Possui 4 placas gráficas “NVIDIA Corporation GK110 [GeForce GTX 780] (rev all)” e uma placa gráfica “Matrox Electronics System Ltd. MGA G220eW WPCM450 (rev 0a)”.

O fluxograma da Figura 15 mostra as principais etapas do estudo entre enzimas CPs e inibidores dipeptidil nitrilas convergindo para o planejamento de novos ligantes.

Figura 15 — Fluxograma das técnicas computacionais utilizadas no trabalho.

Notas: As técnicas utilizadas estão dentro das caixas. A base de dados de proteínas foi o PDB e a base de dados de pequenas moléculas foi o ChEMBL.

No estudo com as enzimas (parte esquerda do fluxograma) cisteíno proteases do conjunto de proteínas do trabalho (cruzaína do *Trypanosoma cruzi* e catepsinas humanas B, K, L e S) foram selecionadas da base de dados do PDB para cálculo dos descritores energéticos MIF o qual mapeou o sítio ativo com sondas químicas. Os valores de energia foram organizados em uma matriz e interpretada através da PCA.

No estudo com os ligantes (parte direita do fluxograma) as moléculas com grupo

funcional nitrila e com arcabouço dipeptídico que possuem dados de inibição na forma de K_i menor que 100 nM em relação às classe de enzimas do trabalho (CatB, CatK, CatL, CatS e Cz) foram selecionadas do site ChEMBL^{177–179}, formando o conjunto de inibidores do filtro virtual e utilizadas para análise estrutural através da docagem molecular.

O modo de interação (*mode of binding*, MOB), parte central do fluxograma, no sítio ativo do complexo: enzima-inibidor (compostos miméticos de dipeptidil nitrilas) será descrito para as estruturas envolvidas no estudo.

A seleção de substituintes conecta as três etapas, à esquerda baseado no alvo (SBDD), à direita baseada no ligante (LBDD) e a central baseada na interface do complexo ligante-proteína. Para isso é utilizado o conceito de hibridização molecular (HM), onde grupos moleculares dos compostos docados similares quimicamente ao modelo seletivo obtido através da PCA foram selecionados para o planejamento de novas moléculas buscando a otimização da seletividade e da potência.

3.1 O modo de interação no sítio ativo e a seletividade

Os complexos proteína-ligante, em que proteína pertence a classe CatB, CatK, CatL, CatS ou Cz e o ligante cocrystalizado é mimético de dipeptidil nitrila foram analisados, com auxílio de programas de visualização gráficas, quanto ao posicionamento do ligante com relação ao sítio ativo e a natureza atômica do ambiente. A análise foi dividida em duas partes: interações conservadas do arcabouço molecular das moléculas dipeptidil nitrilas e a descrição dos grupos substituintes P1, P2 e P3 interagindo com aminoácidos pertencentes aos respectivos subsítios da enzima.

Três diferentes arcabouços moleculares foram identificados como miméticos de dipeptidil nitrilas e são mostrados na Tabela 6:

Tabela 6 — Esqueletos moleculares miméticos de dipeptidil nitrilas cocrystalizados com as proteínas estudadas (catepsinas B, K, L, S e cruzaina).

	CatB	CatK	CatL	CatS	Cz
1GMY	4X6H_3xt ^b 4X6I ^{*,b} 4X6J ^b	2XU1 ^a 2XU3 ^a 2XU4 ^a 2XU5 ^a 2YJ2 ^a 2YJ8 ^a 2YJB ^{a,#} 2YJ9 ^{a,#} 2YJC ^a 3HHA 3HWN 5F02 ^a 6EZP ^{b,o} 6EZ ^{b,o} 6F06 ^{b,*}	1MS6 2FUD 2G7Y [*] 2R9M ^c 2R9N ^c 2R9O ^c	4QH6
1VSN 5TDI		2FQ9 2FRA 2FT2
2FRQ	

Notas: Duas estruturas não foram incluídas apesar de possuírem ligantes que são similares a miméticos de dipeptidil nitrilas: 4DMX e 4DMY, por não compartilharem das interações do arcabouço principal (comentada nesta seção) assim como 6AY2 não foi incluído por não possuir N2. O ligante de 4X6H tem conformações alternativas de ligantes, o ligante selecionado é o 4X6H_3XT, pois está ligado covalentemente na proteína. Durante o texto os ligantes são referenciados pelo código PDB ID das respectivas proteínas onde estão cocrystalizados.

Notas específicas: a, b e c) identificam compostos que fazem parte de uma mesma série de moléculas respectivamente.

Notas específicas: O átomo N1 pode ser considerado como parte do primeiro peptídeo assim como N2 e O2 pode ser considerado do segundo peptídeo da estrutura. O átomo N0 é o nitrogênio do grupo reativo nitrila (*warhead*) na complexação proteína-ligante.

#) são proteínas com ligantes cocrystalizados iguais.

*) significa que o inibidor não se ligou covalentemente à proteína.

o) ligantes cíclicos cocrystalizados

Fonte: autoria própria.

As estruturas cristalográficas na forma halo, em que o ligante não é um polipeptídeo, das proteases CatB, CatK, CatL, CatS e Cz foram selecionadas da base de dados PDB para compor o espaço químico da segunda fase da análise visual, Tabela 7, para que a região do sítio ativo do complexo fosse descrito através do posicionamento do ligante com relação ao respectivo sítio ativo destacando-se os possíveis contatos intermoleculares em S1-P1/P1'/P2', S2-P2 e S3-P3.

Tabela 7 — Código PDB ID dos complexos que compuseram a análise do modo de interação do sítio ativo.

Enzima	"PDB ID" _ "Monômero selecionado"
CatB	1CSB_A, 1GMV_A, 3AI8_B, 6AY2_A
CatK	1ATK_A, 1AU0_A, 1AU2_A, 1AU3_A, 1AU4_A, 1AYU_A, 1AYV_A, 1AYW_A, 1BGO_A, 1MEM_A, 1NL6_A, 1NLJ_A, 1Q6K_A, 1SNK_A, 1TU6_B, 1U9V_A, 1U9W_A, 1U9X_A, 1VSN_A, 1YK7_A, 1YK8_A, 1YT7_A, 2ATO_A, 2AUX_A, 2AUZ_A, 2BDL_A, 2R6N_A, 3C9E_A, 3H7D_A, 3KW9_A, 3KWB_X, 3KWZ_A, 3KX1_A, 3O0U_A, 3O1G_A, 3OVZ_A, 4DMX_A, 4DMY_A, 4X6H_A, 4X6I_A, 4X6J_A, 4YV8_A, 5TDI_A, 6HGY_A, 6QBS_B, 6QL8_A, 6QLM_A, 6QLW_A
CatL	1MHW_A, 2NQD_A, 2XU1_B, 2XU3_A, 2XU4_A, 2XU5_A, 2YJ2_A, 2YJ8_A, 2YJ9_A, 2YJB_A, 2YJC_A, 3BC3_A, 3H89_A, 3H8B_A, 3H8C_A, 3HHA_A, 3HWN_C, 3K24_B, 3OF8_A, 3OF9_A, 4AXM_L, 5F02_A, 5MAE_A, 5MAJ_A, 5MQY_A, 6EZP_A, 6EZX_A, 6F06_A
CatS	1MS6_A, 1NPZ_A, 1NQC_A, 2F1G_A, 2FQ9_A, 2FRA_B, 2FRQ_B, 2FT2_A, 2FUD_B, 2FYE_A, 2G6D_A, 2G7Y_A, 2H7J_A, 2HH5_B, 2HHN_B, 2HXZ_A, 2OP3_A, 2R9M_A, 2R9O_B, 3IEJ_B, 3N3G_A, 3N4C_A, 3OVX_B, 4P6E_A, 4P6G_C, 5QBU_B, 5QBV_A, 5QBW_A, 5QBX_A, 5QBY_B, 5QBZ_A, 5QC0_B, 5QC1_A, 5QC2_B, 5QC3_A, 5QC4_B, 5QC5_B, 5QC6_A, 5QC7_A, 5QC8_A, 5QC9_A, 5QCA_B, 5QCB_A, 5QCC_A, 5QCD_A, 5QCE_B, 5QCF_A, 5QCG_B, 5QCH_A, 5QCI_A, 5QCJ_B
Cruzaína	1AIM_A, 1EWL_A, 1EWM_A, 1EWO_A, 1EWP_A, 1F29_B, 1F2A_A, 1F2B_A, 1F2C_A, 1ME3_A, 1ME4_A, 1U9Q_X, 2AIM_A, 2OZ2_C, 3HD3_A, 3I06_A, 3IUT_A, 3KKU_A, 3LXS_A, 4KLB_C, 4PI3_B, 4QH6_A, 4W5B_A, 4W5C_B, 4XUI_A, 4W5C_B, 4XUI_A,

Notas: Seguido do código PDB ID da estrutura, separado por *underline*, está o monômero usado.

Fonte: autoria própria.

3.2 Campos de Interação Molecular

As próximas alíneas, de “a” a “j”, detalharam os critérios utilizados no filtro utilizado para seleção do conjunto de proteínas do trabalho que formaram o banco de dados de estruturas para ser feita a posterior análise PCA.

- a) A fonte de obtenção dos arquivos das estruturas tridimensionais das enzimas foi a base de dados PDB⁹;
- b) As classes de enzimas utilizadas são catepsinas humanas e cisteíno protease do *Trypanosoma cruzi*;
- c) O limite superior da resolução das estruturas tridimensionais das enzimas foi definido a 3,00 Å;
- d) Apenas proteínas na forma holo, cocrystalizados com ligante, foram usadas no conjunto de proteínas;
- e) O limite inferior da massa molecular dos ligantes no sítio de ligação das estruturas das enzimas foi definido com valor de 285 g.mol⁻¹ evitando-se assim compostos com a massa molecular na ordem de fragmentos moleculares (abaixo de 300 g.mol⁻¹);
- f) Foram selecionadas apenas pequenas moléculas como ligantes cocrystalizados, sendo excluídas as estruturas com polipeptídeos no sítio ativo;
- g) Estruturas na forma de pró-enzimas foram excluídas do conjunto;
- h) Proteínas com mutação na cisteína catalítica foram removidas;
- i) Nas estruturas em que há mais de um monômero na forma holo foi feita a seleção do melhor monômero para uso neste trabalho. A fonte de informação para escolha do melhor monômero foi o relatório de validação dos arquivos provenientes da cristalografia de raio-X¹⁸⁰ a partir do site PDB (<https://www.rcsb.org/>). A análise de escolha da cadeia focou nos resíduos presentes no sítio ativo da enzima, verificando-se a menor quantidade de *outliers* relacionados com questões geométricas e de densidade eletrônica;
- j) Foram analisadas as estruturas depositadas até março de 2019.

O programa PyMOL¹⁸¹, versão 2.1.0, foi utilizado para o alinhamento das estruturas tridimensionais usando a opção *align*¹⁸² utilizando os átomos de carbono α , tendo como padrão a estrutura de Código PDB ID 4QH6⁵², sendo que os ligantes permaneceram no arquivo durante o alinhamento.

Os arquivos com extensão PDB foram separados em duas partes após o alinhamento: (i) arquivo contendo informações relativas à proteína e (ii) arquivo contendo informações do seu respectivo ligante. Os próximos parágrafos descrevem a edição desses arquivos.

Foram excluídos dos arquivos informações de toda espécie química que não faziam

parte da estrutura da proteína. As espécies químicas excluídas foram:

- a) Moléculas de água;
- b) Contra-íons;
- c) Demais moléculas utilizadas no processo de cristalização da proteína (e.g. tampão, traços de outros solventes, precipitantes¹⁸³);
- d) Informações relativas aos átomos de hidrogênios, quando presentes;
- e) Informações sobre conformações alternativas de átomos da proteína, quando presentes (*alternate location indicator*, coluna 17 do formato da extensão PDB). O critério principal da seleção da conformação que permaneceu no arquivo foi estabelecido pelo maior valor de ocupância. O critério secundário em caso de valores idênticos de ocupância para as duas conformações foi o parâmetro fator B, sendo selecionada a conformação que possuiu menor valor.

O programa DeepView/Swiss-PdbViewer¹⁸⁴ foi utilizado para realizar edições nos arquivos das estruturas cristalográficas tridimensionais das proteínas com os seguintes objetivos:

- a) Adição de átomo de oxigênio ao carboxi-terminal da proteína (nomenclatura OXT no formato PDB) quando necessário;
- b) Adição de átomos não descritos nos resíduos de aminoácidos (*missing atoms*) da proteína;
- c) Em alguns casos em que o(s) átomo(s) faltante(s) pertence(m) a uma sequência inserida posteriormente (*Code for insertions of residues*, coluna 27 do formato PDB), fez-se necessário completar manualmente com o respectivo *iCode*¹⁸⁵ após a inclusão automática do átomo faltante pelo programa DeepView.

Há estruturas em que aminoácidos pertencentes a domínios estruturais classificados como *loop* estavam ausentes, apesar de estarem descritos na sequência primária dos aminoácidos. No entanto, o arquivo PDB não continha as coordenadas tridimensionais (*missing residues*). Nestes casos, o programa de modelagem de loops FREAD¹⁸⁶ (<http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/fread/>) foi usado. Segue a relação dos Códigos PDB ID das proteínas alvo do procedimento: 2XU5, 2YJ8, 2YJC, 5F02, 2YJ2, 4AXM, 2XU4, 6EZP, 5MQY, 5MAJ, 2YJ9, 3H8B, 3H89 e 5MAE.

O programa PROPKA^{187,188} foi usado para analisar o sítio ativo com relação aos estados de protonação dos resíduos do sítio ativo.

Utilizando-se as rotinas do programa GRID (MINIM e FILMAP) através da interface gráfica GREATER foi efetuada a adição de contra-íons até a neutralização das proteínas, adicionando-se os íons cloreto e sódio em locais calculados pelo programa com a opção: "íons móveis" assinalada.

As etapas para o cálculo de energia dos campos de interação molecular (MIF) entre os sítios ativos das enzimas e as sondas químicas virtuais foram executadas pelo programa GRID v. 22d^{109,112,189}, distribuído pela empresa Molecular Discovery Ltd. (<https://www.moldiscovery.com/>). Há duas principais rotinas computacionais utilizadas no cálculo das energias de interação intermolecular: GRIN e GRID. As definições dos parâmetros das duas rotinas computacionais são feitas por arquivos chamados grin.in e grid.in, respectivamente.

A etapa inicial do estudo computacional é a utilização da rotina GRIN para checar erros nos arquivos de entrada e atribuir parâmetros para o cálculo das energias. Todas as diretrizes (parâmetros de cálculo) utilizadas no arquivo grin.in, arquivo que contém os parâmetros do cálculo da rotina GRIN, foram utilizadas com a configuração padrão.

O programa GRID foi utilizado para o cálculo dos MIF a partir do arquivo grid.in que contém as diretrizes para a execução da rotina computacional. O programa GRID automaticamente insere as coordenadas que serão utilizadas no cálculo de energia, com a caixa englobando todo o alvo, que é a proteína no caso deste estudo. Porém optou-se por modificar o volume da caixa, especificado através dos vértices, Tabela 8, que englobou o sítio ativo das enzimas previamente alinhadas. Para a determinação deste volume foi necessário sobrepor o conjunto de ligantes cocrystalizados em um arquivo PDB a parte, sendo usada uma distância de 5 Å entre os átomos localizado nas extremidades de cada eixo de coordenada.

Tabela 8 — Valores da caixa em que foi realizado o cálculo dos Campos de Interação Molecular.

Limite da caixa	Coordenada X	Coordenada Y	Coordenada Z
Limite Inferior	-13,00	20,00	18,00
Limite superior	26,00	44,00	43,00

Notas: a caixa foi criada a partir da sobreposição de todos ligantes do conjunto de ligantes cocrystalizados com a distância de 5 Å dos átomos localizados nas extremidades de cada coordenada.

Fonte: autoria própria.

De posse das coordenadas necessárias para o cálculo, utilizou-se a diretriz POSI no arquivo grid.in de modo que foi necessário inserir todas as coordenadas em que o cálculo foi feito (dentro do volume especificado). A lista de coordenadas foi feita de modo a englobar toda a caixa com espaçamentos de 1 Å entre os pontos. A escolha da diretriz POSI para gerar a caixa 3D é artifício operacional que facilitou a manipulação dos arquivos gerados para as etapas posteriores, contornando restrições de conversão das extensões de arquivos provenientes do programa GRID¹¹⁵.

É necessário especificar no arquivo de diretrizes grid.in qual a sonda química a ser utilizada para o cálculo dos MIF. As sondas virtuais selecionadas para o cálculo de energia estão ilustradas na Tabela 9.

Tabela 9 — Sondas químicas virtuais utilizadas no programa GRID para o cálculo dos campos de interação molecular.

Descrição da sonda	Símbolo no GRID
CH ₃	C3
CH vinílico ou aromático	C1=
N sp ³ com par de elétrons livres	N:
NH sp ³ com par de elétrons livres	N1=
NH ₂ sp ³ com par de elétrons livres	N2:
Trimetil amônio	NM3
NH sp ³ com carga positiva	N1+
NH ₃ cátion amina	N3+
NH de amida	N1
NH ₂ de amida	N2
Nitrogênio sp ² com par de elétrons livres	N:=
NH sp ² com par de elétrons livres	NH=
N aniônico de tetrazol	N-:
NH ₂ sp ² cátion amina	N2=
N sp com par de elétrons livre	N:#
NH sp	N1#
Água	OH2
O sp ³ de éster	OES
O de éter ou furano	OC2
OH de grupo alquílico	O1
OH de fenol ou grupo carboxílico	OH
O de carbonila	O
O sp ² de grupo carboxílico (carbonila)	O::
O de sulfona ou sulfóxido	OS
O de grupo nitro	ON
O de sulfato ou sulfonamida	O=
O sp ² de ânion fenolato	O-
F em composto orgânico	F
Sonda hidrofóbica	DRY
Sonda anfifílica	BOTH
CL ⁻	CL-
Na ⁺	NA+

Notas: A sonda virtual BOTH não deve ser utilizada juntamente com a opção de mobilidade das cadeias laterais da proteína¹¹⁵.

Fonte: Adaptado do manual do programa GRID¹¹⁵.

A diretriz MOVE tem como padrão o valor booleano 0, que determina que os átomos das cadeias laterais dos aminoácidos devem ficar rígidos. Esse parâmetro foi configurado para

o valor 1, permitindo a mobilidade nas posições dos grupos funcionais nas cadeias laterais dos resíduos de aminoácidos com o objetivo de otimizar a energia de interação com a sonda.

O programa GRID foi usado para que todas as sondas químicas fizessem o mapeamento energético do conjunto de proteínas do trabalho. Assim, foram gerados 3648 arquivos (32 sondas em 114 estruturas cristalográficas (3 CatB, 42 CatK, 23 CatL, 23 CatS, 23 Cz). Foram feitos os cálculos para 26.000 coordenadas (todas diretrizes POSI) para cada arquivo, totalizando 94.848.000 valores de energias de interações.

As energias de todos arquivos gerados foram organizadas na forma de matriz, Figura 16: proteínas dispostas em linhas e sondas químicas atuando em coordenadas específicas nas colunas.

Figura 16 — Matriz de dados dos valores energéticos dos campos de interação molecular ilustrada de forma representativa.

	GRUPO METILA				CARBONILA					CLORETO			
	(X1, Y1, Z1)	(X2, Y2, Z2)	...	(XN, YN, ZN)	(X1, Y1, Z1)	(X2, Y2, Z2)	...	(XN, YN, ZN)		(X1, Y1, Z1)	(X2, Y2, Z2)	...	(XN, YN, ZN)
1AIM									• • •				
1ATK													
.													
.													
.													
.													
6AY2													
6EZP													
6EZX													

Fonte: autoria própria.

3.3 Análise de componentes principais

A matriz de energias descrita na seção anterior, composta por dados energéticos provindos de ligações não covalentes entre sondas químicas e aminoácidos no sítio ativo do conjunto de proteínas do trabalho — composta por 114 estruturas: Cz (23), CatB (3), CatK (42), CatL (23) e CatS (23) — foi utilizada para a construção de um modelo através da seleção de sondas químicas que representam as principais interações intermoleculares no sítio ativo, estas relacionadas na Tabela 10.

Tabela 10 — Sondas químicas utilizadas na PCA.

Descrição da sonda	Símbolo no GRID
Grupo metil	C3
Nitrogênio primário amídico	N2
Água	OH2
Oxigênio de carbonila	O
Íon cloreto	CL-
Cátion sódio	NA+

Notas: descrição dos parâmetros das sondas estão na Tabela 26.

Fonte: autoria própria.

A matriz de dados energéticos foi então carregada no programa R v. 3.5.1¹⁹⁰. O valor da função `options(max.print=value)` foi modificado para 1.10^9 e o pacote `data.table` foi usado para fazer a transferência da matriz de dados compilada em arquivo de texto para o ambiente do programa R. A matriz energética foi escalonada por blocos de variáveis correspondentes a cada sonda utilizando-se os respectivos valores de desvio padrão. Atribui-se o valor 0 (zero) para valores de energia positivos (interações desfavoráveis), as colunas com valores constantes foram removidas com o pacote `janitor` e efetuou-se a operação de multiplicação da matriz pela constante -1 (permitindo desta forma a relação direta entre as matrizes de *loadings* e *scores*).

A PCA foi calculada com o pacote `prcomp`, atribuindo-se ao argumento `center` o valor `FALSE`. Para dar suporte a separação encontrada no gráfico de *scores* foi utilizada a função `kmeans()` do programa R, v-4.0.4, na matriz energética, atribuindo-se 5 para o argumento `centers` e 10 para `nstart`. A componente denominada cluster foi utilizada para verificar a entidade de cada amostra.

A Tabela 11 é resultado da análise dos gráficos de pontuação (*scores*), Figura 44 e Figura 45, gerados no programa Gnuplot¹⁹¹. A interpretação desses gráficos foi utilizada para a seleção de pontos (elementos) da matriz de *loadings*. Os valores foram filtrados utilizando-se o programa Microsoft Excel 2010 formando arquivos de 9999 pontos para cada proteína e os arquivos salvos em extensão `csv` e posteriormente modificados para o formato `pdb`, para poderem serem visualizados com o programa PyMOL.

Tabela 11 — Critérios de seleção de variáveis da matriz dos pesos (*loadings*) adotados após análise dos gráficos de *scores*.

Classe de proteína	Componente Principal	Critério classificatório
Catepsina B	PC1	Decrescente
Catepsina K	PC2	Decrescente
Catepsina L	PC4	Crescente
Catepsina S	PC2	Crescente
Cruzaína	PC3	Decrescente

Notas: Critérios baseados na análise do gráfico de *scores*.

Fonte: autoria própria.

Como teste de hipótese do modelo, a análise da PCA e os procedimentos posteriores foram aplicados para as proteínas CatK, CatS e Cz.

No programa PyMOL foi feito o descarte de pontos distantes mais de 4 Å de um conjunto de ligantes cocrystalizados, selecionados por possuírem grupos de maior extensão em cada subsítio, descritos na Tabela 12.

Tabela 12 — Ligantes utilizados para delimitar região de pontos no sítio ativo.

Subsítio que motivou a escolha do ligante	Código PDB do ligante cocrystalizado
S2'	3AI8
S1'	1MS6
S1	6AY2
S2	3HWN
S3	5TDI

Fonte: autoria própria.

Para cada proteína selecionou-se os dez pontos de maior valor segundo critério descrito na Tabela 11 que foram utilizados para a seleção de grupos funcionais após docagem do conjunto de inibidores do filtro virtual.

3.4 Redocagem do conjunto de dipeptidil nitrilas cocrystalizadas

Foi realizado experimento de redocagem não covalente com o propósito de se avaliar a performance dos algoritmos frente ao conjunto de proteínas do trabalho com ligantes dipeptidil nitrilas cocrystalizados, e assim posteriormente usar o de melhor performance no conjunto de inibidores do filtro virtual (próxima seção). O grupo de ligantes precisaram atender a dois critérios: (i) ter o arcabouço de dipeptidil nitrila (estrutura à esquerda na Figura 17), que é o espaço químico usado para avaliação neste estudo; (ii) ligar-se covalentemente com a cisteína catalítica em ao menos um dos monômeros do cristal de tal forma que o grupo reativo, que é a nitrila neste caso, esteja próxima à cisteína catalítica.

Figura 17 – Arcabouço molecular das dipeptidil nitrilas selecionadas para o estudo à esquerda e respectivo arcabouço molecular modificado para a forma ligada, preparado o procedimento de redocagem, à direita.

Notas: A nomenclatura das estruturas auxilia a interpretação da Tabela 14. A seta não indica uma reação química e sim a edição dos ligantes para a forma imina, função química similar após ligação covalente com a enzima (tioimidato).

Fonte: autoria própria.

As nitrilas foram modificadas para a forma imina (lado direito na Figura 17) e hidrogênios foram adicionados aos ligantes 2D (formato SMILE) calculado pelo programa Open Babel¹⁹² com posterior análise visual. Os arquivos dos ligantes das estruturas: 2XU3, 2XU4, 2YJ2, 2YJ8, 2YJB, 3HWN e 5F02 continham coordenadas de hidrogênio e estavam em seu estado neutro. Para estes e o restante dos ligantes foi utilizado o ambiente com pH 5,5. A análise visual resultou na retirada da carga positiva ao carbono do ligante de 2FRQ, com código

smile:

(C1(CCCC1)C)[C@H](N/[C+](=N/S(=O)(=O)c1ccccc1)/C)C(=O)NC1(CCS(=O)(=O)CC1)C=N.

As conformações dos ligantes foram geradas pelo programa OMEGA v. 3.0.1.2¹⁹³: (OpenEye Scientific Software, Santa Fe, NM, USA, <https://www.eyesopen.com/>), cujo parâmetros utilizados são apresentados na Tabela 13.

Tabela 13 — Informações do uso do programa OMEGA.

Formato de entrada	Parâmetros
2D, SMILE	-maxconfs 50000, -mpi_np 12, -strictstereo false, -rms 0.3, -maxtime inf

Notas: O parâmetro `-maxconfs` determina o máximo de conformações que o cálculo irá gerar. O parâmetro `-strictstereo` barra as estruturas que não estiverem com sua estereoquímica absoluta determinada, está atribuído `false` ao parâmetro. O parâmetro `-rms` determina qual o limite inferior de RMSD entre os conformeros formados na saída do programa. O parâmetro `-mpi_np` se refere ao número de processadores usados para os cálculos em paralelo. Parâmetro `-maxtime` é o tempo dedicado à busca de ângulos torcionais, a atribuição `inf` retira o limite de tempo.

Fonte: autoria própria

As proteínas foram preparadas — pH 5,5 para escolhas do estado de protonação de ácido aspártico, ácido glutâmico e histidinas — com o programa Protein Preparation Wizard¹⁹⁴ (Schrödinger Release 2020-1: Protein Preparation Wizard. Schrödinger, LLC, New York, NY)

através do módulo PROPKA^{187,188}. O resíduo catalítico, C25 na CatK, CatL, CatS e Cz e C29 na CatB, foi modificado para glicina com o programa Maestro¹⁴⁵.

Os alvos enzimáticos foram preparados com os programas Make Receptor v. 3.3.1.2^{143,162} e Receptor setup v. 3.2.0.2: OpenEye Scientific Software, Santa Fe, NM.

Os alvos preparados com o programa Make Receptor utilizaram a caixa padrão calculada pelo algoritmo através do ligante cognato cocrystalizado. Uma série de restrições personalizadas foram adicionadas, podem ser consultadas na Tabela 14, e seguem a nomenclatura da Figura 17.

Tabela 14 — Informações sobre a utilização do Make Receptor e do FRED no procedimento de redocagem mostrando restrições de localização.

Átomo	Raio / Å	Código em linguagem SMARTS das restrições usadas
N ⁰	2	[$\$([C;\$(C=[N;!$(N\sim[\$([!#6;\$(N,O,S,F,Cl,Br,I,P)]))])])])]$]
C ¹	2	[$\$([C;!$(C=[!#6])])]$]
N ¹	3	[$\$([\$(C;\$(C=[\$([O;D1;\$(O=C)])])]);\$(CN);!\$(C(N)(=O)[!#6])])]$]
O ²	3	[$\$([C;\$(C=[N;!$(N\sim[\$([!#6;\$(N,O,S,F,Cl,Br,I,P)]))])])])]$]
N ²	3	[$\$([\$(C;\$(C=[\$([O;D1;\$(O=C)])])]);\$(CN);!\$(C(N)(=O)[!#6])])]$]
C=N ⁰	5	C=N

Notas: nomenclatura segue a adotada na Figura 17. Código SMARTS gerado pelo programa Make Receptor.

Fonte: autoria própria.

Após a preparação dos ligantes e proteínas a docagem foi feita com programas do pacote OEDOCKING: FRED v. 3.2.0.2^{161,162} e HYBRID v. 3.2.0.2^{161,162}, sumarizadas na Tabela 15.

Tabela 15 — Parâmetros referentes a redocagem com o programa FRED e HYBRID.

Formato de entrada	Parâmetros
Estruturas 3D geradas no OMEGA em extensão oeb.gz	-num_poses 3, -mpi_np 16, -out: *.sdf

Notas: Nos parâmetros utilizados para docagem: -num_poses é o número máximo de posições (*poses*) a serem salvas após a docagem de cada composto. O parâmetro -mpi_np se refere ao número de processadores usados para os cálculos em paralelo. O “*” na última linha é um caractere curinga que são as *poses* docadas em formato SDF.

Fonte: autoria própria.

Após dividir o arquivo de saída da docagem em arquivos separados para cada *pose* em arquivos com formato SDF, o módulo Obrms¹⁹⁵ do programa Open Babel foi usado para calcular a medida de RMSD entre as estruturas das *poses* das moléculas docadas e a molécula na conformação nativa encontrada no arquivo PDB cocrystalizado no sítio ativo da enzima,

usando-a como medida de sucesso da docagem.

Comparações entre os dados estão na parte de Resultados e Discussões para a seleção do melhor programa para a base de dados de complexos utilizados.

3.5 Docagem do conjunto de inibidores do filtro virtual

Da base de dados do site ChEMBL foi feito o *download* das estruturas em formato SMILES dos compostos que continham dados de potência reportadas em função de K_i para ao menos alguma proteína do estudo (CatB, CatK, CatL, CatS, Cz). O valor de corte considerado de K_i foi de 100 nM ($pK_i > 7$). Desse grupo formado selecionou-se as estruturas que possuem a nitrila como grupo reativo e que fossem miméticos de dipeptidil, ou seja, miméticos de dipeptidil nitrilas. Estes dados compõem a base de dados intitulada conjunto de inibidores do filtro virtual.

Foram realizados procedimentos de docagem com o programa FRED da OEDocking nos ligantes do conjunto de inibidores do filtro virtual. Os códigos das proteínas utilizadas como alvos biológicos estão na Tabela 16.

Tabela 16 — Proteínas utilizadas como alvos na docagem.

Proteína	Código PDB ID
catepsina B	1GMY
catepsina K	1VSN
catepsina L	2YJ2
catepsina S	1MS6
Cruzaína	4QH6

Fonte: autoria própria

O método utilizado é idêntico ao procedimento de docagem apresentado na seção anterior. Apenas houve o acréscimo de um procedimento na verificação da carga formal das moléculas através do programa Quacpac v. 1.7.0.2 (OpenEye Scientific Software, Santa Fe, NM), que gera tautômeros considerando uma larga faixa de pH, mantendo-se o resultado encontrado pelo programa Open Babel com pH de 5,5 exceto pela estrutura ChEMBL2153167, que foi docada com o anel tetrazol neutro. Na caixa utilizada o tamanho do padrão foi modificado, em direção ao subsítio S2', S3 e ao solvente, levando em conta a maior variedade estrutural das moléculas do conjunto de inibidores virtual em relação ao conjunto de moléculas dipeptidil nitrilas cocrystalizadas com as proteínas do trabalho (que foi utilizada na redocagem na seção anterior).

3.6 Hibridização molecular

Com as conformações (*poses*) geradas na docagem do conjunto de inibidores do filtro virtual juntamente com respectivos valores de potência escolheu-se por substituintes com características similares ao modelo seletivo obtido com o procedimento MIF-PCA. Foram selecionadas moléculas do banco de dados ChEMBL com seletividade para CatK (ChEMBL CHEMBL371064 e CHEMBL364727) e CatS (ChEMBL3640716).

O procedimento de hibridização molecular foi feito apenas para a CatK. Os híbridos moleculares foram formados a partir do grupo P3 das moléculas selecionadas (ChEMBL364727 e CHEMBL371064 – denominado Balicatib) e do arcabouço do composto Neq409⁵² (PubChem CID 137348120), Tabela 17.

A molécula *in-house* Neq0682¹⁰⁰ já está sendo estudada pelo grupo e possui as características em P3 desejáveis para a CatK, também selecionada para análise. Assim como um derivado planejado que contém um grupo espaçador em P3 (grupo metileno entre o nitrogênio do anel pirazina e o carbono carbonílico), denominado nesta dissertação de neq0682_v2. Os compostos são mostrados na Tabela 17.

Tabela 17 — Moléculas selecionadas para docagem covalente, exceto pelo Neq409.

Estrutura	Nome	Origem
Neq409	In-house
Neq0682	Molécula <i>in-house</i>
Neq0682_v2	Neq0682
hm_1	CHEMBL3640727 e neq409
hm_2	CHEMBL371064 (Balicatib) e neq409

Fonte: autoria própria.

Nesse conjunto de compostos da Tabela 17 (exceto pela primeira linha) foi feito um procedimento de docagem covalente tendo como alvo a CatK (PDB ID 1VSN). Iniciou-se pela utilização do programa AVOGADRO¹⁹⁶ (Avogadro: an open-source molecular builder and visualization tool. Version 1.2., <http://avogadro.cc/>): o ligante cocristalizado de código PDB 1MS6 tiveram os grupos P1, P2, P3 retirados (exceto pelos átomos pesado β de P1, P2 e/ou P3 quando presentes) e o grupo nitrila modificado para o grupo tioimidato. Esta última condição é artefato operacional para a posterior docagem covalente com o programa GOLD¹⁶⁵. Uma versão deste arcabouço foi salva para posterior uso na docagem, Figura 18.

Figura 18 — Subestrutura molecular do complexo PDB 1MS6 utilizada em restrições na docagem das moléculas planejadas.

Notas: A subestrutura completa foi considerada para modificação da molécula neq0682 e neq0682_v2, enquanto que para hm_1 e hm_2 o C β foi retirado. As ligações fixadas como rígidas na docagem foram N1-C2, C2-Ca2, Ca2-N2. A subestrutura composta pelos átomos SG, C1, NO, Ca1, C β 1 (quando presente), N1, C2 e O2 foi utilizada como restrição conformacional na docagem em todos derivados.

Os átomos foram fixados para posterior otimização geométrica. Os grupos P1, P2 e P3 das moléculas foram inseridas no arcabouço e hidrogênios (pH=5,5) foram adicionados as estruturas que foram minimizadas geometricamente através do campo de força MMFF94¹⁹⁷ com 9999 iterações. Os arquivos com formato mol2 foram editados manualmente (outro artefato operacional para a docagem covalente com o programa GOLD): Ao campo subst_id foi atribuído o valor 1 e ao restante dos átomos da molécula o valor 2. O campo subst_name foi modificado para CYS25.

Foram adicionados hidrogênios com o programa Hermes (*Cambridge Crystallographic Data Centre*) à proteína CatK de código PDB 1VSN, que foi utilizada como alvo enzimático.

As moléculas minimizadas no AVOGRADO foram então docadas covalentemente utilizando o algoritmo de pontuação ChemPLP. Foram fixadas as rotações de algumas ligações, descritas nas notas da Figura 18.

Foi usado como referência para o tamanho do sítio ativo o ligante cocrystalizado com o PDB 1MS6 (Figura 47).

Uma subestrutura do arcabouço molecular anteriormente preparado foi utilizado como restrição conformacional (scaffold constraint), Figura 18, com o campo “Never dock a ligand when a constraint is physically impossible” marcado. Na aba “Ligand Flexibility” foi habilitada as caixas: Flip pyramidal N, Flip amide bonds, Detect internal H bonds, flip ring corners e Match template conformations. Na aba “Fitness & Search Options” a opção “Allow early termination” foi desabilitada.

Uma restrição farmacofórica de localização para anéis foi utilizada no subsítio S3 com raio de 1 Å.

Nos compostos com P1', neq0682 e neq0682_v2, foi adicionada uma restrição para anéis aromáticos com raio de 1 Å.

Distâncias entre 2,5 e 3,5 Å foram usadas como restrições entre átomos da proteína ASN158:O, A:GLY66:N, A:GLY66:O aos átomos N1,O2,N2, Figura 19.

A flexibilidade da docagem foi configurada para o valor máximo (search efficiency = 200%). Análise visual foi feita nas 10 primeiras poses de cada ligante docado para a seleção da conformação.

4 RESULTADOS E DISCUSSÃO

Os títulos das seções secundárias estão organizadas como na seção métodos.

4.1 O modo de interação no sítio ativo e a seletividade

O modo de interação (*mode of binding*, MOB) no sítio ativo do complexo: enzima-inibidor (compostos miméticos de dipeptidil nitrilas) é um tipo de análise possível pela disponibilidade de coordenadas atômicas de estruturas tridimensionais resolvidas, arquivos de estruturas depositadas na base de dados de proteínas PDB (exceto as estruturas de empresas privadas, e.g. farmacêuticas e biotecnológicas¹⁹⁸), sendo utilizada em abordagens baseadas no alvo (*structure-based drug design*, SBDD). Apesar de indispensável, a análise do modo de interação não é ferramenta independente capaz de racionalizar a afinidade da interação entre ligante e alvo macrobiológico. Fatores entrópicos, a dinâmica (ou mobilidade) conformacional do complexo e a solvatação no ambiente biológico e consequentemente variação da constante dielétrica entre superfície da proteína exposta ao solvente e o ambiente hidrofóbico de bolsões no sítio ativo são apenas alguns exemplos de fenômenos não observáveis pelo MOB.

A base de dados PDB possui 171916 estruturas depositadas¹⁹⁹. Dessas, 80% são provenientes de experimento de difração por raio-X²⁰⁰ (cristalografia de raio-X), que mesmo sendo um experimento com grandes incertezas incluídas durante a transformação da densidade eletrônica em um modelo de coordenadas atômicas²⁰¹, onde até mesmo ligações entre os átomos apresentam dificuldades para serem modeladas (em resoluções médias)²⁰, as estruturas geradas são de grande importância para planejamento de inibidores baseado no alvo e deve ser inspecionada cuidadosamente.

A análise do modo de interação nos sítios ativos é dividida em duas partes: A análise nos átomos do esqueleto molecular (imagem à esquerda da Figura 19), e a análise realizada nos substituintes: os grupos químicos P1, P2, P3.

Assim através da análise das estruturas constata-se que as possíveis interações de ligação de hidrogênio que os átomos N1, O2 e N2 fazem são constantes em todas as estruturas cristalográficas verificadas, ou melhor, a distância e o ângulo de ligação são propícios para que ocorram (considerando limites mais brandos de interação entre proteína-ligante). A Tabela 18 reúne estes dados acompanhado do esquema da Figura 19 (imagem à

esquerda) como suporte ilustrativo.

Figura 19 — Imagem à esquerda: interações conservadas de átomos do protótipo dipeptidil nitrila com respectivos átomos da cadeia principal da proteína assim como a nomenclatura dos grupos substituintes com respectivos subsítios em que ocupam na enzima. Imagem ao centro: boxplot de distâncias entre átomos mencionados do complexo proteína ligante. Imagem ao centro: boxplot de ângulos entre os átomos.

Legenda: S1, S1' e S2'; S2 e S3, referem-se aos subsítios das enzimas que interagem respectivamente com os substituintes do inibidor: P1, P2 e P3. Linhas tracejadas representam as possíveis interações entre os átomos do inibidor e átomos da cadeia principal do subsítio.

Notas: Figura pode ser utilizada como suporte para leitura da Tabela 18.

Fonte: autoria própria.

Tabela 18 — Distância e ângulos de átomos do arcabouço principal dos ligantes do conjunto de proteínas do trabalho cocristalizados com dipeptidil nitrilas.

PDB	N1 (S1)			O2 (S3)			N2 (S3)		
	(NH--O)/º	(H--OC)/º	(N--O)/Å	(NH--O)/º	(H--OC)/º	(O--N)/Å	(NH--O)/º	(H--OC)/º	(N--O)/Å
Catepsina B									
1GMY ⁵³	G198-O			G74-N			G74-O		
	172,4	142,5	3,0	152,3	129,5	3,0	177,1	141,5	2,8
Catepsinas K									
1VSN ²⁰²	D158-O			G66-N			G66-O		
	167,2	127,6	2,9	150,3	133,9	2,9	129,3	118,9	3,2
5TDI ⁹⁷	N161-O			G66-N			G66-O		
	161,6	120,0	2,8	151,1	136,8	3,0	149,2	121,6	3,0
4X6H ²⁰³	N161-O			G66-N			G66-O		
	149,3	107,3	3,1	152,5	129,9	3,1	125,7	123,1	3,0
4X6I ²⁰³	N161-O			G66-N			G66-O		
	132,3	104,4	3,2	148,9	129,5	2,9	157,5	114,7	3,2
4X6J ²⁰³	N161-O			G66-N			G66-O		
	150,5	106,5	3,0	151,0	131,4	3,1	126,9	121,8	2,9
Catepsinas L									
2XU1 ³²	D162-O			G68-N			G68-O		
	168,5	126,6	3,0	143,2	141,6	3,2	N terciário		3,7
2XU3 ³²	D162-O			G68-N			G68-O		
	164,9	129,2	2,9	145,0	137,2	3,2	N terciário		3,6
2XU4 ³²	D162-O			G68-N			G68-O		

Tabela 18 — Distância e ângulos de átomos do arcabouço principal dos ligantes do conjunto de proteínas do trabalho cocrystalizados com dipeptidil nitrilas.

PDB	N1 (S1)			O2 (S3)			N2 (S3)		
	(NH--O)/°	(H--OC)/°	(N--O)/Å	(NH--O)/°	(H--OC)/°	(O--N)/Å	(NH--O)/°	(H--OC)/°	(N--O)/Å
	164,7	121,2	2,9	143,1	140,2	3,2	N terciário		3,9
2XU5 ³²	D162-O			G68-N			G68-O		
	175,1	122,3	2,9	150,8	144,8	3,3	N terciário		3,9
2YJ2 ³³	D162-O			G68-N			G68-O		
	174,5	124,1	2,9	146,8	142,5	3,2	N terciário		3,7
2YJ8 ³³	D162-O			G68-N			G68-O		
	176,0	123,8	2,9	144,7	140,4	3,2	N terciário		3,7
2YJ9 ³³	D162-O			G68-N			G68-O		
	160,8	122,1	2,9	143,3	141,9	3,3	N terciário		3,8
2YJB ³³	D162-O			G68-N			G68-O		
	154,5	116,1	3,0	154,2	149,9	3,2	N terciário		3,7
2YJC ³³	D162-O			G68-N			G68-O		
	168,3	125,5	2,9	148,0	139,9	3,2	N terciário		3,7
3HHA ²⁰⁴	D162-O			G68-N			G68-O		
	174,1	138,1	2,9	149,4	137,1	3,0	158,7	137,1	2,9
3HWN ²⁰⁵	D162-O			G68-N			G68-O		
	160,6	137,3	2,8	131,2	138,0	3,1	152,9	119,7	3,1
5F02 ²⁰⁶	D162-O			G68-N			G68-O		
	170,2	130,2	2,9	142,9	135,9	3,2	N terciário		3,6
6EZP ²⁰⁷	D162-O			G68-N			G68-O		
	168,7	134,3	2,8	146,6	136,9	2,9	163,0	134,5	2,9
6EZ ²⁰⁷	D162-O			G68-N			G68-O		
	169,6	153,5	2,5	146,2	132,2	2,7	158,4	128,5	2,8
6F06 ²⁰⁷	D162-O			G68-N			G68-O		
	134,0	153,4	3,4	153,5	133,9	3,1	172,3	146,2	2,7
Catapsinas S									
1MS6 ²⁰⁸	N163-O			G69-N			G69-O		
	169,4	148,0	3,0	156,8	134,6	2,9	131,9	131,0	3,0
2FQ9 ²⁰⁹	N158-O			G66-N			G66-O		
	174,9	150,7	2,8	156,6	136,9	2,9	131,4	124,1	3,0
2FRA ²¹⁰	N158-O			G66-N			G66-O		
	167,9	148,9	2,9	157,0	141,5	3,0	129,1	124,3	3,1
2FRQ ²¹¹	N158-O			G66-N			G66-O		
	171,0	156,9	3,0	159,4	141,7	3,0	166,5	136,0	2,8
2FT2 ²¹²	N158-O			G66-N			G66-O		
	173,4	144,5	2,9	157,9	140,8	3,0	149,8	134,2	3,1
2FUD ²¹³	N158-O			G66-N			G66-O		
	168,3	161,0	2,8	153,8	130,2	3,2	138,7	133,6	3,1
2G7Y ²¹⁴	N158-O			G66-N			G66-O		

Tabela 18 — Distância e ângulos de átomos do arcabouço principal dos ligantes do conjunto de proteínas do trabalho cocrystalizados com dipeptidil nitrilas.

PDB	N1 (S1)			O2 (S3)			N2 (S3)		
	(NH--O)/°	(H--OC)/°	(N--O)/Å	(NH--O)/°	(H--OC)/°	(O--N)/Å	(NH--O)/°	(H--OC)/°	(N--O)/Å
	156,8	144,7	3,1	151,1	129,6	3,0	N terciário		4,5
2R9M ²¹⁵	N163-O			G69-N			G69-O		
	172,6	162,1	2,8	150,2	126,3	3,2	139,8	136,5	3,0
2R9N ²¹⁵	N163-O			G69-N			G69-O		
	171,2	147,3	2,7	152,4	126,0	3,3	131,9	135,0	3,0
2R9O ²¹⁵	N163-O			G69-N			G69-O		
	177,1	158,3	2,7	150,8	123,5	3,2	133,9	125,2	3,0
4QH6 ⁵²	D161-O			G66-N			G66-O		
	135,4	143,5	3,4	145,1	142,2	3,0	134,8	123,1	3,9

Notas: A nomenclatura do átomo do ligante no cabeçalho da tabela refere-se ao esqueleto da estrutura da Figura 19. Em parênteses está o subsítio ao qual pertence o átomo. Exemplificando, o ângulo doador de ligação de hidrogênio entre N1 do ligante do complexo 4QH6 e o oxigênio amídico do resíduo D161 é 135,4°. Os ligantes dos complexos 2G7Y, 6F06 e 4X6I não estão ligados covalentemente à proteína. O ligante do complexo 4X6H possui mais de uma conformação depositada, a utilizada é a de código 4X6H_3XT.

O resíduo catalítico H162 (numeração da Cz de PDB ID 4QH6) é semelhante nas família de enzimas estudadas, localizada no subsítio S1, está aproximadamente a distância de 5 Å dos inibidores estudados, podendo formar interações polares.

A regularidade das ligações de hidrogênio encontradas no arcabouço molecular do conjunto de miméticos de dipeptidil nitrilas pode ser um dos fatores que implica na regularidade espacial encontrada nos ligantes cocrystalizados deste tipo de estrutura, como é possível verificar na Figura 20, possibilitando o uso de restrições, espaciais ou de contato, em simulações entre ligante com a proteína.

Figura 20 — Sobreposição de subestruturas moleculares das moléculas miméticos de dipeptidil nitrilas cocrystalizadas com as enzimas estudadas (cruzaína, catepsinas humanas: B, K, L e S) destacando a similaridade conformacional.

Notas: Na imagem é apresentada a maioria das estruturas do conjunto de dipeptidil nitrilas cocrystalizadas com as enzimas estudadas (CatB, CatK, CatL, CatS e Cz), onde apresentam RMSD de aproximadamente (dependente da estrutura de referência) 1,3 Å.

Fonte: autoria própria.

Após descrição das interações que o esqueleto molecular faz com a proteína a atenção será direcionada aos subsítios do sítio ativo: S2', S1', S1, S2 e S3, apresentados na Figura 21.

Figura 21 — Sítio ativo das enzimas cisteíno proteases divididas em subsítios.

Notas: macromoléculas representadas através da representação em superfície molecular (estilo de representação *surface* no programa PyMOL) e *cartoon* (representa a estrutura supersecudária ou motivos) ao fundo.

Notas específicas:

A) Visão do subsítio S2' da CatB com os resíduos H110 e H111 em bastões. Cisteína catalítica em amarelo. O fragmento cocrystalizado é da estrutura PDB 3AI8.

B) Subsítios S1', S1, S2 e S3 (PDB 4QH6) coloridos com o ligante cocrystalizado mostrado na forma de linha.

Fonte: autoria própria.

A família cujo grupo de proteínas foi estudada nesta dissertação apresenta um alto grau de homologia⁴⁶, portanto é um desafio fazer o modificações estruturais em compostos que resultem em seletividade dentre as diferentes classes do grupo de enzimas, principalmente interações no subsítio S3, mais exposto ao solvente²¹⁶.

No restante desta seção serão descritas diferenças estruturais encontradas através da análise do modo de interação de complexos enzimáticos CatB, CatK, CatL, CatS e Cz, com enfoque principal aos que tenham dipeptidil nitrilas cocrystalizadas. Serão destacados

aminoácidos equivalentes estruturalmente (com as proteínas superpostas) não idênticos: conservados (preserva propriedades físico-químicas⁴³, e.g carga elétrica, polaridade e volume²¹⁷) e não conservados. A Figura 22 ilustra propriedades a serem analisadas em substituições conservadas dos aminoácidos²¹⁸.

Figura 22 — Diagrama de Venn representando relações físico-químicas entre os 20 aminoácidos principais.

Notas: O diagrama mostra relação de propriedades físico-químicas dos aminoácidos mostrando diferentes perspectivas de substituições conservadas.

Fonte: EXARCHOS, K. P. e colaboradores. Detection of discriminative sequence patterns in the neighborhood of proline cis peptide bonds and their functional annotation. *BMC Bioinformatics*, v. 10, n. 113, 2009²¹⁸.

A análise será dividida em subsítios, e seguirá a sequência: S1', S1, S2, S3 e S2'.

Os ligantes que atingem S1' (e.g. PDB 1MS6) são ancorados por três resíduos (numeração de PDB 4QH6), que limitam a profundidade de alcance do grupo P1 neste subsítio. Dois resíduos — W184 e H162 — são conservados (e semelhantes) no conjunto de proteínas do trabalho, portanto não importantes para explorar-se a seletividade. Já o resíduo de ancoramento A141 (grupo 3) não é igual em todas as proteínas. A Figura 23 mostra o subsítio S1' do conjunto de proteínas do trabalho superpostas destacando os resíduos não idênticos, separados por grupos, com a finalidade de encontrar aminoácidos que possam contribuir com a modificação racional de ligantes buscando seletividade. Figuras e análises semelhantes serão feitas ao longo do texto com os subsítios S1, S2 e S3.

Figura 23 — Subsítio S1' das enzimas estudadas (cruzaína e catepsinas humanas B, K, L e S) com destaque para os aminoácidos localizados na mesma região (rotulados) que não são semelhantes.

Notas: A legenda dos resíduos em cada local (numerada em vermelho) está descrita na tabela da imagem. As estruturas das proteínas CatB, CatK, CatL, CatS e Cz utilizados possuem respectivamente código PDB ID: 6AY2, 1VSN, 2YJ2, 2G7Y e 4QH6. Quando não há “equivalente tridimensional” do resíduo em alguma proteína, a legenda é representada pelo caractere hífen na respectiva linha da tabela. A cisteína catalítica é exibida na forma de bastões e está rotulada em vermelho para auxiliar na orientação espacial (parte inferior esquerda) da enzima. Com opacidade reduzida estão representados em estilo *cartoon* partes das estruturas supersecudárias onde localizam-se os resíduos destacados com propósito de auxiliar na visualização da perspectiva 3D, e.g. o grupo 2 está localizado em uma região mais exposta ao solvente que o grupo 5.

Fonte: autoria própria.

No grupo 1 a Cz possui um resíduo não conservado de metionina, porém não há estruturas em que o ligante alcance o resíduo. A CatK contém um resíduo polar glutamina e pode realizar ligação de hidrogênio. A Figura 24 mostra na imagem à esquerda que o ligante cocrystalizado de 1AU3²¹⁹, uma cetona cíclica peptídeo mimética ligada covalentemente a C25, alcança sua cadeia lateral. A imagem à direita mostra que o complexo 1MS6 possui S1'-P1 próximos.

Figura 24 — Sítio ativo da catepsina K (PDB 1AU3 à esquerda e PDB 1MS6 à direita) com destaque para resíduo não conservado no subsítio S1' (em relação às proteínas estudadas, catepsinas: B, L, S e cruzaina) de glutamina, rotulados com distâncias de átomos de inibidores cocrystalizados.

Notas: As imagens não mostram interações e sim distâncias, com intuito de guiar modificações em inibidores que almejam alcançar os resíduos não conservados no subsítio.

Fonte: autoria própria.

No grupo 2 há um resíduo de metionina exclusivo da CatB, porém não há pequenas moléculas cocrystalizadas dessa enzima (humana) que explorem o subsítio S1'.

No grupo 3 apresentam-se resíduos não semelhantes mas conservados em relação à natureza apolar, destaca-se o único grupo volumoso do resíduo de valina da CatB, mas não há estruturas depositadas que explorem o contato.

No grupo 4 a CatB possui uma glicina, de menor volume do grupo. A substituição em P1 de um grupo ciclopropil por um grupo isopropil-metil-benzeno em SAR explorando interações S1'-P1 em miméticos de dipeptidil nitrilas apresentou um acréscimo de 0,4 e 0,6 unidades em escala logarítmica em seu pK_i , em dois protótipos com combinações de P2 e P3 diferentes, enquanto a substituição em outras proteínas não apresentou um aumento na potência¹⁰⁴, a substituição em mais de um composto ameniza o efeito da não aditividade dos substituintes na SAR, hipotetizando-se que a seletividade seja resultado ao maior espaço da glicina não conservada na região, que acomodaria melhor o grupo isopropil-metil-benzeno (cadeia lateral do resíduo modificado Thr-O-Bn), juntamente com a especificidade do grupo 3 apresentada em parágrafo anterior, ou seja, a valina da CatB, que pode fazer contatos apolares com o benzeno do grupo substituído.

Ainda no grupo 4 CatK e CatS possuem resíduos de asparagina, de muito difícil acesso, enquanto CatL e Cz possuem resíduos de aspartato. Não há estruturas cristalográficas da CatL e nem da Cz que possuam interações intermoleculares com esse resíduo de S1'. Porém há estruturas que encorajam o planejamento para tal contato, como o inibidor ligado covalentemente vinil sulfona da estrutura de raio-X PDB 1F29 mostrada na Figura 25, destacando a distância o subsítio-grupo molecular S1'-P1.

Figura 25 — Sítio ativo do complexo inibidor-cruzaína (PDB 1F29) destacando distância entre átomo do ligante em P1 alcançando resíduo de aspartato em S1'.

Notas: As imagens não mostram interações e sim distâncias, com intuito de guiar modificações em inibidores que almejam alcançar os resíduos no subsítio.

Fonte: autoria própria.

O grupo 5 possui um resíduo de arginina na CatS não conservado. Três estruturas de CatS do conjunto de proteínas do trabalho com ligantes dipeptidil nitrilas cocrystalizados possuem P1 prolongado e alcançam o subsítio S1': 1MS6, 2R9O E 2R9N, estes ilustrados na Figura 26.

Figura 26 — Resíduo de arginina no subsítio S1' não conservado na região estrutural da catepsina S (a partir da imagem à esquerda: PDB ID 1MS6, 2R9O e 2R9N) e ligantes cocrystalizados com átomos próximos rotulados com distâncias.

Notas: As imagens não mostram interações e sim contatos entre átomos mais próximos, com intuito de guiar modificações racionais baseadas na estrutura.

Fonte: autoria própria.

A cadeia lateral de arginina está em direção contrária do subsítio S1', e uma possível interação parece improvável. O mesmo acontece nos resíduos de serina da CatK e da Cz.

O subsítio S1 já apresenta uma maior similaridade nos aminoácidos (é onde localiza-se a tríade catalítica) onde é possível encontrar apenas um resíduo de glutamato na CatB não conservado, a ilustração está na Figura 27.

Figura 27 — Subsítio S1 das enzimas estudadas (catepsinas B, K, L, S e cruzaína) com destaque para resíduo de glutamato único da catepsina B.

Notas: As estruturas utilizadas foram: 6AY2, 1VSN, 2YJ2, 2G7Y e 4QH6. Não há “equivalente tridimensional” do resíduo nas enzimas CatK, CatL, CatS e Cz. O rótulo (número 6) em vermelho auxilia na descrição do grupo durante o texto. O resíduo seletivo E122 da proteína catepsina B tem sua cadeia lateral com os átomos de carbono ilustrado na forma de bastões verdes. A superfície molecular (*surface* no PyMOL) é da Cz. A cisteína catalítica está rotulada e representada na forma de bastões. Fonte: autoria própria.

É necessário um inibidor que se estenda em P1 para alcançar esse resíduo. O mimético de dipeptidil nitrila cocrystalizado com a CatB de código 6AY2 possui um P1 com esqueleto da cadeia lateral de arginina, porém ao invés de um grupo guanidina há um grupo ureia, considerado estrutura privilegiada em química medicinal por ser biossintetizado e eliminado no organismo humano. Mesmo fora do ângulo de uma interação de ligação de hidrogênio está a distância de 3,1 Å, interessante de ser notada para planejamentos que almejam alcançar esse resíduo em P1. A Figura 28 ilustra a descrição feita.

Figura 28 — Catepsina B (complexo 6AY2) e seu resíduo não conservado (em relação às proteínas estudadas: catepsinas K, L, S e cruzaína) de glutamato no subsítio S1 mostrada através da única estrutura cristalográfica de CatB que acessa o resíduo.

Notas: A ligação de hidrogênio não é realizada, o hidrogênio do nitrogênio aponta para a superfície da cavidade em S1, e o carboxilato do resíduo glutamato está direcionado ao solvente. Porém a falta de interações observadas em estruturas com o resíduo de difícil acesso torna esse complexo importante para análise.

Fonte: autoria própria.

O subsítio S2 é hidrofóbico³³, possui uma cavidade maior em relação aos outros subsítios. É preenchido com grupos químicos volumosos²⁰⁸, possui a forma convexa e é

considerado o principal subsítio com relação a especificidade²²⁰ na família de proteínas estudadas. De forma geral a CatB possui o S2 mais largo e a CatK o mais estreito em comparação às outras enzimas estudadas (Cz, CatL, CatS)²¹⁶.

É proposta uma divisão em três partes: P2, P2 estendido e P2 profundo (nomenclatura não convencional) para melhor descrição dos resíduos que possam contribuir para a seletividade na interação com um ligante mais extenso. A Figura 29 mostra o subsítio S2 das enzimas do trabalho superpostas.

Figura 29 — Subsítio S2 das proteínas estudadas (catepsinas humanas B, K, L, S e cruzaína) separadas em grupos, 7 ao 14, com aminoácidos não conservados destacados.

Notas: A legenda dos resíduos em cada local (numerada em vermelho) está descrita na tabela da imagem. As estruturas das proteínas CatB, CatK, CatL, CatS e Cz utilizados possuem respectivamente código PDB ID: 6AY2, 1VSN, 2YJ2, 2G7Y e 4QH6. Quando não há “equivalente tridimensional” do resíduo em alguma proteína, a legenda é representada pelo caractere hífen na respectiva linha da tabela. A cisteína catalítica está rotulada em vermelho para auxiliar na orientação espacial (parte superior esquerda) da enzima. Com opacidade reduzida estão representados em estilo *cartoon* partes das estruturas supersecundárias onde localizam-se os resíduos destacados com propósito de auxiliar na visualização da perspectiva 3D. O subsítio foi dividido (nomenclatura não convencional) em três partes: S2, S2 estendido e S2 profundo. A cisteína catalítica está rotulada e representada na forma de bastões. Fonte: autoria própria.

Os grupos 7-11 fazem uma região conservada em relação à natureza hidrofóbica dos resíduos, dificultando alguma racionalização em relação a propriedade de seletividade. Os resíduos do grupo 8 estão localizados em uma alça, diferenciando-se estruturalmente da CatB, que prolonga em direção ao solvente em motivo folha beta, implicando em uma largura maior em S2 e também conferindo característica a S2 de maior profundidade. No grupo 10 a CatB possui o resíduo não conservado de prolina. O grupo 10 e 11 podem ser considerados superfícies de ancoramento de P2; P2 estendido mais P2 profundo respectivamente, pois limitam a profundidade do subsítio.

No grupo 12 destaca-se o resíduo de glutamato que faz parte tanto da CatB quanto da Cz. Estudos de mapeamento do sítio ativo (*active site mapping*, ASM) com peptídeos do sítio ativo da CatB encontraram como resultado a preferência por resíduos hidrofóbicos na região⁶⁴ e não resíduos básicos, contrariando a racionalidade da inspeção visual. Não há nenhuma estrutura cristalográfica da CatB onde o glutamato do S2 faça uma interação porém há 5 estruturas da cruzaina onde observa-se a formação de ligação de hidrogênio ou pontes salinas: PDB 1AIM, 2AIM, 3LXS, 4PI3 e 4XUI. As duas primeiras são mostradas na Figura 30.

Figura 30 — Glutamato do subsítio S2 da cruzaina interagindo com P2 do inibidor: cadeia lateral da tirosina à esquerda e guanidina na imagem à direita, através de ligação de hidrogênio e possível ponte salina, respectivamente.

Notas: à esquerda: complexo PDB 1AIM, à direita: complexo PDB 2AIM.

Autoria: própria.

Ainda no grupo 12 é interessante notar que o resíduo de alanina da CatL é o menor, facilitando o acesso de grupos ao P2 estendido e profundo, ver Figura 32. Por outro lado o resíduo de leucina da CatK bloqueia esse acesso, que juntamente com o resíduo de tirosina (classificado como S3, localizado no grupo 20, ver Figura 33) na parede oposta, torna o S2 da CatK o menor (em extensão) do conjunto de proteínas do trabalho.

Em quase a totalidade de estruturas do conjunto de proteínas do trabalho com ligantes dipeptidil nitrilas cocrystalizados em que o grupo P2 é um anel este se encontra na posição vertical no subsítio (paralela às “paredes laterais” do subsítio). Especificamente na CatS onde um resíduo não conservado de fenilalanina faz parte de S2, é possível encontrar o grupo

aromático aparentemente realizando duas interações π -stacking do tipo *edge-to-face*, ilustrada na Figura 31. Uma constatação que talvez possa contribuir para a seletividade da CatS frente às outras proteínas do grupo estudado.

Figura 31 — Possível interação hidrofóbica π -stacking no canto inferior de S2 (F205) e na saliência inferior de S3 (F67) da proteína catepsina S.

Notas: Complexo utilizado na imagem é o PDB 2FRA.

Fonte: autoria própria.

No grupo 13 há um resíduo de serina da CatB, portanto com caráter de doador de ligação de hidrogênio, característica não presente nos resíduos do restante do grupo de proteínas, considerando os grupos ácidos desprotonados. Apenas um P2 bem estendido poderia interagir com este resíduo, não há estruturas cristalográficas em que possa se verificar tal contato.

No grupo 14 há um resíduo de aspartato da CatL não conservado entre o grupo de proteínas estudadas. Ele está localizado no fundo do subsítio, e só um P2 bem extenso para alcançar esse resíduo. O ligante cocrystalizado na CatL de código 3HWN²⁰⁵ consegue, com a benzila ligada em meta ao grupo oxidiazol funcionalizada com metilimina em P2. O nitrogênio sp^2 fica a 2,4 Å de um dos oxigênios do carboxilato, contato mostrado na Figura 32.

Figura 32 — Complexo inibidor-CatL (PDB 3HWN) possui P2 que alcança resíduos de glutamato de difícil alcance em S2 profundo.

Notas: Destaque em amarelo para o resíduo reativo cisteína 25 para auxiliar na orientação espacial da enzima.

Fonte: autoria própria.

Ainda no grupo 14 é interessante notar o resíduo de alanina da CatB, o menor no grupo. Ele contribui (juntamente com o resíduo de P76) para a CatB possuir o P2 mais extenso das proteínas.

A substituição em P2 de leucina por meta-clorofenilalanina em SAR explorando interações, entre quais, S2-P2, em miméticos de dipeptidil nitrilas apresentou um acréscimo de 0,9 e 1,0 unidades em escala logarítmica de pK_i na CatB e CatL respectivamente, enquanto que foi deletério em outras proteínas¹⁰⁴. Não é possível racionalizar a seletividade encontrada através da análise das estruturas.

Todas estruturas do conjunto de proteínas cristalizadas com dipeptidil nitrilas possuem anel em P3. Diferentemente do grupo substituinte P2, os grupos P3 assentam de forma paralela ao chão (superfície da cavidade) do subsítio S3, formado por um par de glicinas (65-66 na Cz PDB 4QH6). No S3 estendido esse ancoramento é feito pela cadeia principal do par de resíduos D60-S61 (PDB 4QH6). A Figura 33 mostra o subsítio S3 e destaca grupos de resíduos não semelhantes.

Figura 33 — Subsítio S3 das proteínas estudadas (catepsinas B, K, L, S e cruzaina) superpostas, separadas por grupos, 15 ao 20, para visualização de resíduos não conservados.

Notas: A legenda dos resíduos em cada local (numerada em vermelho) está descrita na tabela da imagem. As estruturas das proteínas CatB, CatK, CatL, CatS e Cz utilizados possuem respectivamente código PDB ID: 6AY2, 1VSN, 2YJ2, 2G7Y e 4QH6. Quando não há “equivalente tridimensional” do resíduo em alguma proteína, a legenda é representada pelo caractere hífen na respectiva linha da tabela. A cisteína catalítica está rotulada em vermelho para auxiliar na orientação espacial (parte central direita da imagem) da enzima. Com opacidade reduzida estão representados em estilo *cartoon* partes das estruturas supersecudárias onde localizam-se os resíduos destacados com propósito de auxiliar na visualização da perspectiva 3D. O subsítio foi dividido (nomenclatura não convencional) em três partes: S3, S3 estendido e S3 profundo.

Fonte: autoria própria.

No grupo 15 destaca-se um resíduo de glicina na CatK, sendo o menor resíduo do grupo e o único apolar.

O grupo 16 apresenta uma possibilidade de seletividade através da serina 61 na cruzaina. O inibidor não covalente hidroximetil cetona cocrystalizado na estrutura PDB 1ME3 aparentemente realiza uma ligação de hidrogênio através do nitrogênio da piridina em P2 com o oxigênio da serina em S2, assim como o nitrogênio da quinolina do grupo P3 do inibidor covalente irreversível oxiarilcetona da estrutura 3IUT²²¹ da cruzaina indica a possibilidade da interação de ligação de hidrogênio com a o oxigênio da serina 61 em S2. Imagens são

mostradas na Figura 34.

Figura 34 — Estruturas cristalográficas da cruzaina destacando resíduo não conservado de serina em S3 e grupos próximos mostrando a possibilidade de interações seletivas.

Notas: à esquerda estrutura de código PDB ID 1ME3 e à direita 3IUT.

Fonte: autoria própria.

Ainda no grupo 16 a CatS possui um resíduo de lisina não conservado mas os ângulos e distâncias do resíduo encontrados nas estruturas depositadas não são favoráveis a interações. Esse resíduo limita o alcance do grupo P3 dos inibidores até parte estendida e profunda do subsítio S3, particularidade dessa proteína com relação ao conjunto de proteínas do trabalho (CatB, CatK, CatL, CatS, Cz).

No grupo 17 a CatS possui um resíduo glutamato, único com propriedades ácidas entre as proteínas, mas ele é bem no fundo de S3. Nenhuma estrutura cristalográfica contém ligantes com P3 que se aproximem do resíduo.

No grupo 18 a CatK tem um glutamato, como no grupo 17, o único ácido, mas ele é bem no fundo de S3 com sua cadeia lateral exposta ao solvente, de difícil acesso.

No grupo 19 CatB tem um glutamato não conservado entre o grupo de proteínas. Nenhum ligante cocrystalizado da CatB explorou o P3 profundo, além de difícil acesso está direcionado ao solvente.

O grupo 20 possui resíduos hidrofóbicos, acessíveis, porém de natureza similar, dificultando a busca por interações seletivas. Tanto a CatK quanto a CatB possuem resíduos de tirosina possibilitando uma ligação de hidrogênio apesar de não haver estruturas depositadas que possuam essa interação. Mas há estruturas, como de código 5TDI (ligante Odanacatib), que possuem interação π -stacking edge-to-face com o resíduo aromático.

A Figura 35 mostra o subsítio S2' separado por grupos para visualizar a possibilidade de interações seletivas.

Figura 35 — Subsítio S2' das proteínas estudadas (catepsinas B, K, L, S e cruzaína) superpostas, separada em grupos, 21 ao 24, para visualização de resíduos não conservados.

Notas: A legenda dos resíduos em cada local (numerada em vermelho) está descrita na tabela da imagem. As estruturas das proteínas CatB, CatK, CatL, CatS e Cz utilizados possuem respectivamente código PDB ID: 6AY2, 1VSN, 2YJ2, 2G7Y e 4QH6. Quando não há “equivalente tridimensional” do resíduo em alguma proteína, a legenda é representada pelo caractere hífen na respectiva linha da tabela. A cisteína catalítica está rotulada em vermelho para auxiliar na orientação espacial (parte central direita da imagem) da enzima. Com opacidade reduzida estão representados em estilo *cartoon* partes das estruturas supersecudárias onde localizam-se os resíduos destacados com propósito de auxiliar na visualização da perspectiva 3D. Os resíduos idênticos triptofano e glutamina na superfície da cavidade são mostrados apenas para orientação e estão exibidos na forma de linhas.

Fonte: autoria própria

Não há estruturas cristalográficas de raio-X que com ligantes dipeptidil nitrilas em que se alcance subsítio S2' dentro da classe de enzimas estudadas. A Figura 21A (imagem superior) mostra a estrutura de um fragmento (massa molecular inferior a 300 Da e o número de átomos pesados entre 2 e 18¹⁸³) nitroxolina cocrystalizado com a CatB de código PDB ID 3A18²²², interagindo com os 2 resíduos seletivos da volta ocludente (*occluding looping*): H110 e H111 (grupos 21 e 22), ou seja, não são conservados com as outras classes de proteínas do estudo (CatK, CatL, CatS e Cz). Já a Figura 36 ilustra a primeira estrutura cristalográfica, CatB de código 1CSB, a interagir no subsítio S2' através de um ligante mimético de dipeptidil

epóxido, ligado covalentemente e irreversivelmente²²³.

Figura 36 — Imagem do sítio catalítico da proteína catepsina B (PDB 1CSB) com destaque ao subsítio S2' onde está localizado o par de histidinas seletivas (em relação às outras proteínas do trabalho: CatK, CatL, CatS e Cz) que participam de pontes salinas feita com o carboxilato do inibidor cocristalizado.

Notas: Hidrogênios não são exibidos nas estruturas.

Fonte: autoria própria.

No grupo 23, ainda na volta ocludente da CatB há um resíduo cisteína que pode ser um alvo a ser explorado.

No grupo 24, CatB e CatS possuem serina na região e CatL, CatK e Cz possuem glutamina. Um possível lugar a ser explorado em busca de seletividade.

4.2 Campos de interação molecular

Todas estruturas 3D das enzimas selecionadas para o trabalho (catepsinas humanas B, K, L, S e cruzaina do *Trypanosoma cruzi*) provenientes do banco de dados do RCSB PDB foram obtidas por experimentos de cristalografia de raio-X, pois eram os únicos disponíveis.

Considerando o organismo de origem da proteína, foram excluídas 41 estruturas de CatB, CatK, CatL e CatS por não serem de *Homo sapiens*: 11 *Bos taurus*, 9 *Mus Musculus*, 6 *Schistosoma mansoni*, 4 *Rattus norvegicus*, 3 *Clostridioides difficile*, 2 *Tenebrio molitor*, 1 *Fasciola hepatica*, 1 *Macaca mullata*, 1 *Oryctogalus cuniculus*, 1 *Toxoplasma gondii*, 1 *Trypanosoma brucei*, 1 *Trypanosoma congolense*.

A resolução cristalográfica foi um critério considerado para a seleção das estruturas 3D de proteínas²⁰¹. Em suma, quanto menor o valor do parâmetro mais alta é a resolução. Foram considerados valores inferiores a 3,00 Å e todas estruturas atenderam ao critério, portanto nenhuma estrutura foi filtrada nesta etapa.

Proteínas na forma *apo* foram excluídas já que a forma da proteína é alterada em função das interações e a própria ocupação do ligante também modifica o sítio ativo das proteínas

em geral²²⁴, incluindo as cisteíno proteases. Este critério resultou na exclusão de 11 estruturas.

Os ligantes com alta afinidade e com características similares a fármacos para esta classe de proteínas apresentam massa molecular acima do limite inferior estabelecido de 285 g mol^{-1} para os inibidores cocrystalizados nos complexos. Proteínas cujo ligantes com massa molar inferior a este valor são considerados similares a fragmentos (conceitualmente) e interagem com subsítios específicos das cisteíno proteases implicando em diferenças conformacionais no sítio ativo. Este último filtro levou a remoção de 13 estruturas.

Ligantes que são polipeptídeos também modificam o formato do sítio catalítico de forma distinta de um ligante que é uma pequena molécula, por isso proteínas com polipeptídeos ancorados no sítio ativo foram excluídas.

Houve a exclusão de duas estruturas por não estarem em sua forma ativa, mas como pró-enzimas: PDB 1CLJ e 1CS8.

As estruturas cristalográficas em que há mutação na cisteína catalítica foram excluídas.

A escolha do monômero a ser utilizado no banco de proteínas será ilustrada através de três seleções ilustrados na Figura 37.

Figura 37 — Exemplos de critérios utilizados para escolha dos monômeros que compuseram o conjunto de proteínas do trabalho.

Notas específicas:

A) Relatório total de estrutura PDB ID 2XU3²²⁵ mostra a qualidade do modelo, excluem-se os monômeros A e D, pois eles possuem resíduos que não aparecem no modelo (mas aparecem na amostra), coloridos da cor cinza. O monômero C possui três resíduos que não se ajustam a densidade eletrônica (pontos vermelhos acima dos resíduos) enquanto o monômero B possui apenas dois, todos estes resíduos não localizam-se no sítio ativo. Assim, o monômero identificado por B foi selecionado.

B) Unidade assimétrica da proteína PDB 4QH6 ilustrando a escolha do monômero verde, o único que cristalizou com o ligante no sítio ativo (forma holo) que na imagem está circulado.

C) Modelo da unidade assimétrica PDB 1GMV em que o monômero branco foi selecionado por estar realizando menos contatos com os outros monômeros, evitando a influência do empacotamento do cristal¹⁵³.

Fonte: A) Retirado de:

https://files.rcsb.org/pub/pdb/validation_reports/xu/2xu3/2xu3_full_validation.pdf.

B e C) Fonte: autoria própria.

Ao final do filtro foram selecionadas 144 monômeros de estruturas cristalográficas de enzimas que atenderam aos critérios estabelecidos neste trabalho depositadas até março de 2019 e estão relacionadas na Tabela 19, sendo composto por 3 CatB; 23 CatL; 23 CatS; 42 CatK e 23 cruzainas.

Tabela 19 — Códigos PDB ID do conjunto de proteínas do trabalho: catepsinas B, K, L, S e cruzaína.

Enzima	"PDB ID"_"Monômero selecionado"
CatB	1CSB_A, 1GMY_A, 6AY2_A,
CatL	2XU1_B, 2XU3_A, 2XU4_A, 2XU5_A, 2YJ2_A, 2YJ8_A, 2YJ9_A, 2YJB_A, 2YJC_A, 3H89_A, 3H8B_A, 3HHA_A, 3HWN_C, 3OF8_A, 3OF9_A, 4AXM_L, 5F02_A, 5MAE_A, 5MAJ_A, 5MQY_A, 6EZP_A, 6EZX_A, 6F06_A
CatS	1MS6_A, 1NPZ_A, 1NQC_A, 2F1G_A, 2FQ9_A, 2FRA_B, 2FRQ_B, 2FT2_A, 2FUD_B, 2FYE_A, 2G6D_A, 2G7Y_A, 2H7J_A, 2HH5_B, 2HHN_B, 2HXZ_A, 2R9M_A, 2R9O_B, 3N3G_A, 3N4C_A, 3OVX_B, 4P6E_A, 4P6G_C
CatK	1ATK_A, 1AU0_A, 1AU2_A, 1AU3_A, 1AU4_A, 1AYU_A, 1AYV_A, 1AYW_A, 1BGO_A, 1MEM_A, 1NL6_A, 1NLJ_A, 1SNK_A, 1TU6_B, 1U9V_A, 1U9W_A, 1U9X_A, 1VSN_A, 1YK7_A, 1YT7_A, 2ATO_A, 2AUX_A, 2AUZ_A, 2BDL_A, 2R6N_A, 3C9E_A, 3H7D_A, 3KW9_A, 3KWB_X, 3KWZ_A, 3KX1_A, 3O0U_A, 3O1G_A, 3OVZ_A, 4DMX_A, 4DMY_A, 4X6H_A, 4X6I_A, 4X6J_A, 4YV8_A, 5TDI_A, 6QBS_B
Cruzaína	1AIM_A, 1EWL_A, 1EWM_A, 1EWO_A, 1EWP_A, 1F29_B, 1F2A_A, 1F2B_A, 1F2C_A, 1ME3_A, 1ME4_A, 1U9Q_X, 2AIM_A, 2OZ2_C, 3HD3_A, 3I06_A, 3IUT_A, 3KKU_A, 3LXS_A, 4KLB_C, 4PI3_B, 4QH6_A, 4XUI_A

Notas: A proteína de código PDB ID 2R9N apesar de atender os critérios de seleção não foi utilizada por erro, sendo impossibilitado a sua inserção posterior nos experimentos com MIF pela expiração da validade do programa GRID. Julga-se não relevante sua inclusão dado a porcentagem de similaridade com a proteína 2R9M (RMSD=0,13).

Fonte: autoria própria.

Na alinhamento estrutural o algoritmo da minimização do desvio da raiz quadrada média entre a distância dos carbonos α das proteínas é usado para a superposição das estruturas através do programa PyMOL¹⁸¹. É um método em que a estrutura 3D rígida da proteína está com liberdade para transladar e rotacionar de modo que se atinja a minimização da média aritmética da distância dos respectivos átomos carbonos α da cadeia principal entre as proteínas. A Figura 38 ilustra a superposição estrutural feita com o programa PyMOL.

Figura 38 — Ilustração de superposição estrutural com duas enzimas diferentes da mesma família.

Notas: RMSD = 0,936 Å.

Fonte: autoria própria.

Como forma de medida da superposição obtém-se o desvio da raiz quadrada da média (*Root Mean Square deviation*, RMSD)²²⁶. No alinhamento estrutural os Cα que não se corresponderam nas duas estruturas não farão parte do cálculo de RMSD. O modelo de alinhamento estrutural ilustrado obteve 148 átomos que corresponderam-se segundo critério do algoritmo — PyMOL — e assim são variáveis na equação 14²²⁷, onde é mostrada de forma genérica o cálculo de RMSD entre duas estruturas A e B com N carbonos α que se corresponderam.

$$\left(\frac{1}{N} \sum_{i=1}^N (\|a_i - b_i\|^2)\right)^{0.5} \quad 14$$

Minimizando o valor da equação 14 pela variação de a_i e b_i (Cα emparelhados) obtém-se as coordenadas da superposição estrutural.

Os resultados obtidos são mostrados na seção ANEXO ordenados por enzima na Tabela 25.

As estruturas cristalográficas da cruzaína apresentaram menor valor médio de RMSD, uma vez que uma estrutura cristalográfica da cruzaína (Código PDB ID 4QH6) foi usada como padrão de referência para a superposição. A Tabela 20 sumariza parâmetros estatísticos dos valores de RMSD das classes do conjunto de proteínas do trabalho, auxiliando na construção de gráfico diagrama de caixas (*box-plot*).

Tabela 20 — Dados estatísticos de valores do desvio da raiz quadrada média (RMDS) do conjunto de proteínas do trabalho: catepsinas B, K, L, S e cruzaína.

Classe	N	Mínimo	Q ₁	Mediana	Q ₃	Máximo	Média	D.P
CatB	3	0,936	0,936	0,973	0,996	0,996	0,968	0,030
CatK	42	0,563	0,617	0,638	0,670	0,715	0,715	0,037
CatL	23	0,592	0,601	0,622	0,649	0,694	0,694	0,031
CatS	24	0,611	0,649	0,676	0,693	0,723	0,723	0,029
Cz	22	0,198	0,294	0,311	0,329	0,366	0,366	0,037

Notas: N é o número de amostras, no caso de proteínas, Q₁ e Q₃ é o primeiro e o terceiro quartil, D.P. é desvio padrão. A Família Cz possui 22 amostras por não considerar o padrão PDB ID 4QH6. A Família CatS possui 24 amostras por considerar a proteína PDB ID 2R9N (não utilizada nos cálculos, ver notas da Tabela 19).

Fonte: autoria própria.

A Figura 39 organiza os valores de RMSD em um gráfico de diagrama caixas (*box plot*).

Figura 39 — Valor de RMSD por classe enzimática representado por gráfico diagrama de caixas.

Fonte: Gráfico feito no programa Minitab 19 (Arend, Dominic N. (1993). Choices (Version 4.0) [Computer software]. Champaign, IL: U.S. Army Corps of Engineers Research Laboratory. (CERL Report No.CH7-22510)).

As estruturas cristalográficas da CatB tiveram o maior valor médio de RMSD, mostrando a maior divergência estrutural em relação à estrutura da cruzaína usada. Podemos concluir que estes valores de médios obtidos são esperados, uma vez que a classificação da base de dados MEROPS⁴¹ distingue as cisteíno proteases da família C1 do clã CA em enzimas tipo catepsinas L (*CatL-like*) e catepsina B. A Figura 40 mostra a discrepância estrutural de modo visual entre as proteínas das outras classes do conjunto de proteínas.

Figura 40 — Diferenças estruturais na catepsina B em relação às outras enzimas estudadas (catepsinas: K, L, S e cruzaína).

Notas: A proteína CatB está à direita e a Cz à esquerda, representando: Cz, CatK, CatL, CatS. A extensão das estruturas supersecundárias folha beta (representada na cor preta) na catepsina B aparentemente é responsável pela diferença nos valores de RMSD calculados.

Fonte: autoria própria.

A escolha da exclusão dos átomos de hidrogênio quando presentes nos arquivos deve-se ao fato que a densidade eletrônica modelada nos experimentos cristalográficos são de resolução insuficiente para determinar suas coordenadas²⁰¹, sendo necessário resoluções abaixo de 1 Å²⁰. Também foi desconsiderado as moléculas de água nas estruturas cristalográficas neste estudo, baseado na dificuldade de atribuir sua localização através da densidade eletrônica²²⁸ e da dificuldade de distinguir entre moléculas de água encontradas no ambiente nativo do sítio ativo da proteína e de moléculas provindas do procedimento experimental²²⁹. Demais espécies químicas utilizadas no processo físico de cristalização (e.g. precipitantes) também foram excluídas por não serem alvo do mapeamento energético que fora realizado.

Uma correção feita nos arquivos das proteínas envolve a inserção de coordenadas para átomos ausentes pertencentes a resíduos conhecidos. Ou seja, a sequência primária das proteínas é conhecida, porém os átomos não foram identificados através de sua densidade eletrônica na técnica de cristalografia de raio-X. Inferir a conformação com auxílio de algoritmos neste caso é uma solução usualmente utilizada. Há duas possibilidades em relação a localização dos resíduos:

- a) resíduos distantes do sítio ativo: não são envolvidas diretamente nas interações intermoleculares no sítio ativo;
- b) resíduos pertencentes ao sítio ativo: o cálculo das energias de campos de interação molecular será feito atribuindo-se certa mobilidade aos átomos das cadeias laterais dos resíduos, o que diminui a importância nas incertezas das

coordenadas desses átomos além que o cálculo quimiométrico (abordado na próxima seção) atenua diferenças estruturais por considerar várias estruturas da mesma classe de proteína.

Os resíduos ausentes no arquivo de coordenadas atômicas são parte de um *loop* lateral exposto ao solvente, optou-se por modelá-lo. Apesar de distantes do sítio ativo, é um artefato computacional sugerido pelo manual do programa GRID¹¹⁵ para evitar erros de cálculo no programa. Foi usado o programa FREAD (informações na Tabela 21) para modelagem dos resíduos ausentes.

Tabela 21 — Proteínas que tiveram seu loop exposto ao solvente modelado pelo programa FREAD por haver resíduos não resolvidos nas estruturas cristalográficas.

Código PDB da proteína com <i>loop</i> reconstruído	Modelo utilizado pelo programa FREAD	Sequência Peptídica ausente no arquivo PDB	Resíduos inseridos nos arquivos PDB
2XU4	2NQD_CatB	176-178	173-181
2XU5	2NQD_CatB	174-177	172-179
2YJ2	2NQD_CatB	176	173-179
2YJ8	2NQD_CatB	176	173-179
2YJ9	3KSE_A	174-177	172-179
2YJC	2NQD_CatB	176	173-179
3H89	3KSE_A	175-179	173-181
3H8B	3KSE_A	175-179	173-181
4AXM	2NQD_CatB	175-176	173-180
5F02	2NQD_CatB	176	173-179
5MAE	3KSE_A	176-178	173-181
5MAJ	3HHA_A	176	172-180
5MQY	2NQD_CatB	176-179	174-181
6EZP	2NQD_CatB	176-178	174-180

Notas: No campo: "Modelo utilizado pelo programa FREAD para modelar *loop*", após o Código PDB ID está a letra da cadeia polipeptídica (separadas pelo caractere *underline*) da proteína selecionada pelo algoritmo do programa. Ilustrando um exemplo, a proteína 5MAE não possuía as coordenadas espaciais referentes a parte de seu *loop* lateral exposto ao solvente. Assim, indicando a sequência do *loop*, o programa selecionou a cadeia A da estrutura 3KSE como modelo. Os resíduos modelados estão indicados no campo "Resíduos inseridos nos arquivos PDB" que neste caso foram os aminoácidos 173-181, que não coincidem com os resíduos ausentes.

Fonte: autoria própria.

Conforme visualizado na Figura 41, a volta (*loop*) modelada pelo programa FREAD está localizada em uma região periférica na cisteína protease.

Figura 41 — Ilustração das estruturas antes e após modelagem do loop lateral exposto ao solvente.

Notas: A estrutura de Código PDB ID 2XU5 em ciano contém o loop com resíduos ausentes na estrutura 3D, posteriormente modelado pelo FREAD resultando na estrutura de cor verde com o loop modelado.

Fonte: autoria própria.

Demais edições foram feitas, presentes em rotinas de preparação de proteínas para posterior utilização em SBDD: ajuste de nomenclatura para o formato do programa utilizado (GRID neste estudo), e.g. grupos terminais da estrutura (C-terminal e N-terminal). Escolha da conformação do resíduo quando as coordenadas atômicas possuem mais de uma possibilidade,

O estado de protonação dos resíduos ácido/básicos do sítios ativos de amostras das classes de proteínas foram verificadas quanto ao seu estado de protonação através do programa PROPKA. Concluiu-se da análise que o pKa dos resíduos utilizados como padrão no programa GRID eram similares, não sendo editado, portanto as cargas formais dos átomos dos resíduos.

À neutralização das proteínas adicionou-se íons móveis (em relação a sonda química) ao redor da proteína. A neutralização da proteína é um artefato requerido pelo cálculo computacional quando este é seguido do modelo quimiométrico, a opção deles serem móveis evita que as energias de interações intermoleculares entre sonda-proteína sofram interferência pela sua presença durante o mapeamento energético do sítio ativo¹¹⁵.

Os detalhes do procedimento computacional MIF já foram descritos com detalhes na seção Introdução. O método mapeia regiões de (macro)moléculas através do cálculo das energias de ligações não covalentes realizadas entre a molécula de interesse — o sítio ativo da enzima neste trabalho — e sondas químicas virtuais: conjunto de parâmetros que mimetizam espécies moleculares. O programa GRID calcula as energias baseando-se em funções de base de mecânica molecular (campo de força GRID) e utiliza parâmetros empíricos em suas

funções.

A superposição do conjunto de ligantes do estudo, mostrada na Figura 42, serviu como base para estabelecer qual o limite no sítio ativo a serem realizados os cálculos energéticos com as sondas químicas virtuais.

Figura 42 — Caixa com malha delimitando as coordenadas do cálculo com intersecções entre as coordenadas x, y e z igualmente espaçadas em 1 Å.

Notas: A caixa engloba todos os ligantes cocrystalizados do conjunto de proteínas do trabalho com distância mínima 5 Å de qualquer átomo. Os pontos cinzas representam a grade de coordenadas cartesianas tridimensionais construída de forma a englobar todo conjunto de dipeptidil nitrilas cocrystalizadas mais as proteínas filtradas por conter o resíduo catalítico com mutação. Cada ponto da figura na cor cinza é uma coordenada cartesiana em que o cálculo foi realizado.

Fonte: autoria própria.

Estabelecidas as coordenadas tridimensionais onde os cálculos foram feitos, escolheu-se aproximadamente o maior número de sondas químicas que o programa oferecia, hipotetizando-se que seria a forma mais detalhada de análise possível. Os parâmetros que definem cada uma das sondas são mostrados na seção ANEXO na Tabela 26.

O resultado obtido através dos cálculos do campo de interação molecular nesta seção pode ser utilizado pelo modo principal de uso do GRID, que é a visualização dos campos de interação molecular no sítio ativo de enzimas. Porém o objetivo desta parte do trabalho é explorar a seletividade enzimática, aproveitando-se da grande quantidade de estruturas de raio-X disponíveis para as classes enzimáticas estudadas (exceto pela CatB com apenas três estruturas que atenderam os critérios de filtro estabelecidos). Por isso os valores calculados para cada uma das sondas e das enzimas nas coordenadas pertencentes a caixa 3D foram organizados em uma matriz — de acordo com as características descritas por Cruciani¹¹⁶ e descrito na seção respectiva em métodos — para serem dados de entrada no programa R, servindo para alimentar um modelo estatísticos de PCA com o objetivo de validar regiões do sítio ativo propícias de serem feitas interações intermoleculares seletivas específicas.

4.3 Análise de componentes principais

Os valores das energias calculados com o programa GRID na seção anterior foram utilizados para a construção de um modelo de PCA com 6 sondas químicas selecionadas por representar a maioria dos tipos de interações intermoleculares:

- Sonda química metil, CH_3^- (símbolo no GRID C3) representa o caráter apolar, investigando regiões hidrofóbicas;
- Sonda química nitrogênio primário amídico, H_2NCO (símbolo N2): realiza duas doações de ligação de hidrogênio (parâmetros definidos pelo programa GRID);
- Sonda química água, H_2O (OH2), que realiza 2 doações de ligação de hidrogênio e 2 interações como aceitador de ligação de hidrogênio;
- Sonda química oxigênio de carbonila, CO (símbolo O), por ser parametrizada como aceitador de duas ligações de hidrogênio;
- Sondas que representam cargas: positiva e negativa: cátion sódio (símbolo GRID Na^+) e ânion cloreto (Cl^-), buscando regiões eletropositivas e eletronegativas respectivamente.

Na Figura 43 são observadas as dez componentes principais de maior variância do modelo.

Figura 43 — Gráfico de barras mostrando porcentagens de variância explicada pelas componentes principais de 1 à 10.

Fonte: autoria própria.

Através de dois gráficos utilizando-se variáveis da matriz de pontuação, Figura 44 e

Figura 45, apesar de difícil visualização por 3 dimensões estarem plotados no plano do papel, foi possível constatar através da utilização do software Gnuplot, que utilizando-se as 4 primeiras componentes principais (que explicam 29% de variância do modelo), é possível obter a separação de todas as classes de proteínas, assim como selecionar objetos relevantes na matriz de pesos, ou seja, interações intermoleculares seletivas. Ao se utilizar o método de agrupamento em 5 classes ($k=5$) pela média (k -means) nas observações da matriz de dados energéticos utilizada na PCA, foi constatado a separação das 5 classes de proteínas, conferindo robustez à análise. O gráfico de pontuações, Figura 44, mostra a separação das classes enzimáticas CatS, CatB e Cz através das variáveis PC1, PC2 e PC3.

Figura 44 — Gráfico de pontuação da PCA com as variáveis PC1, PC2 e PC3 realizada com o conjunto de proteínas do trabalho.

Notas: Cada ponto representa uma proteína e todas foram verificadas quanto a sua identidade.
Fonte: autoria própria.

Através dos valores da matriz de pontuação é possível selecionar os valores dos objetos na matriz de pesos que mais contribuem para a separação das amostras e portanto para a seletividade entre as classes de proteínas. A CatB possui os maiores valores de PC1. Catepsina S possui os menores valores de PC2 e cruzaiña os maiores valores de PC3.

Já através do gráfico de pontuações da Figura 45, é possível verificar a importância da variável PC4 para a separação da CatL, com os menores valores para essa variável.

Essas observações implicaram nos critérios de classificação explicitadas na Tabela 10 na seção Métodos desta dissertação.

Figura 45 — Gráfico de pontuação da PCA com as variáveis PC1, PC2 e PC4 realizada com o conjunto de proteínas do trabalho.

Notas: Cada ponto representa uma proteína e todas foram verificadas quanto a sua identidade.
Fonte: autoria própria.

O destaque de pontos seletivos feito a partir da interpretação da PCA, descrita nas etapas seguintes, foram feitas para as proteínas CatK, CatS e Cz, como teste de hipótese do modelo desenvolvido. Acredita-se que a análise possa se estender a qualquer nova estrutura depositada que pertença as famílias estudadas através de um modelo quimiométrico *soft independent modeling of class analogies* (SIMCA)²³⁰, ou que o modelo possa ser ampliado através da inclusão de classes de proteínas da mesma família da papaína.

Para poder explorar as interações intermoleculares seletivas encontradas, os pontos seletivos foram plotados em conjunto com o respectivo sítio ativo, assim foi necessário codificar a natureza das interações que a sonda realiza, a forma escolhida foi a cor, e estas são mostradas na Tabela 22, que auxilia a interpretação das figuras relacionando o sítio ativo com os pontos seletivos.

Tabela 22 — Legenda de cores relacionando a natureza das interações das sondas químicas virtuais.

Símbolo utilizado	Sonda química	Propriedade da sonda química	Cor
N2	Nitrogênio primário de amida	Doador de ligação de hidrogênio (2)	azul
O	Oxigênio de carbonila	Aceitador de ligação de hidrogênio (2)	vermelho
OH2	Oxigênio de Água	Doador e aceitador de ligação de hidrogênio (2,2)	laranja
NA+	Íon sódio	Carga positiva	verde
CL-	Cloreto	Carga negativa	amarelo
C3	Grupo metil	Interações hidrofóbicas	cinza

Notas: Os números em parênteses indicam a quantidade de ligações de hidrogênio segundo parametrização da sonda química utilizada pelo software GRID que foi utilizado para os cálculos de campo de interação molecular. Informações mais detalhadas sobre a parametrização de sondas estão na Tabela 26.

Fonte: autoria própria

A Figura 46 apresenta as seleções de variáveis após critérios aplicados e descritos na seção métodos, pontos que representam interações intermoleculares seletivas, para a CatK, CatS e Cz.

Figura 46 — Pontos encontrados pela utilização sequencial MIF-PCA que representam interações intermoleculares entre as sondas químicas específicas e o sítio ativo que apresentam seletividade dentre o grupo de proteínas estudados.

Notas: Superfície colorida em amarelo localiza a cisteína catalítica. Os pontilhados indicam a distância entre o átomo do resíduo indicado pelo código de uma letra e o ponto seletivo na unidade angstrom. A legenda das esferas podem ser consultadas na Tabela 19. Os ligantes nativos das estruturas proteicas estão desenhadas na forma de linhas, com carbono em preto.

Fonte: autoria própria.

Interações intermoleculares que representam seletividade em coordenadas específicas do sítio ativo formaram um modelo seletivo — lembrando que os eventos do reconhecimento molecular são fenômenos complexos, possibilitando que as definições em seu escopo sejam extremamente plásticas²³¹. A inspeção de resíduos com características complementares quimicamente ao modelo é apenas uma tentativa de racionalização dos resultados, já que o ponto seletivo é resultado de interação com todo o ambiente próximo.

A CatK apresentou cargas positivas como sendo interações favoráveis no subsítio S3. Há evidências que grupos básicos em P3 favorecem a potência e a seletividade em enzimas catepsina K²³², tanto que o composto Balicatib²³³ chegou a fase II como inibidor de CatK²¹⁶. A substituição de 3-(tert-butil)-1-metil-1H-pirazol por 1H-pirrol[2,3-b]piridina, grupo que aumenta o alcance do nitrogênio básico do grupo P3, publicada em SAR recente que, além de outras, também explora a interação P3-S3, resultando em acréscimo de 0,4 unidades logarítmicas de pK_i para a catepsina K, enquanto não houve melhoras nas outras proteínas do estudo¹⁰⁴.

A CatS apresentou cargas negativas como sendo mais favoráveis tanto no subsítio S2 quanto no subsítio S3. Uma racionalização possível no subsítio S2 seria o resíduo não conservado de lisina que está a 4 Å de distância de um dos pontos mais seletivo encontrado pelo modelo, mas o fato do resíduo não interagir com o ligante em nenhuma das 52 estruturas cristalográficas da catepsina S depositadas não colabora com a hipótese.

A cruzaína apresentou cargas positivas próximas ao resíduo de glutamato, bastante explorado no planejamento de inibidores da Cz, onde é possível verificar sua interação em 5 estruturas cristalográficas^{234,235,236}. Esta informação é consistente com o fato que o subsítio S2 acomoda bem grupos carregados positivamente²³⁷.

Os pontos selecionados foram utilizados para a escolha de grupos químicos das moléculas docadas ao final do filtro virtual para o planejamento de híbridos moleculares. Estes resultados serão mostrados nas respectivas seções de docagem e hibridização molecular. Na próxima seção é feita a escolha do software de docagem molecular a ser utilizado.

4.4 Redocagem do conjunto de dipeptidil nitrilas cocrystalizadas

A Tabela 23 contém os ligantes que serão utilizados como base de dados para os testes de performance.

Tabela 23 — Conjunto de dipeptidil nitrilas cocrystalizadas nas proteínas do estudo: catepsinas B, K, L, S e cruzaina.

Classe	PDB – ID
Cz	4QH6
CatB	1GMY
CatK	1VSN, 4X6H_3xt, 4X6J, 5TDI
CatL	2XU1 ^a , 2XU3 ^a , 2XU4 ^a , 2XU5 ^a , 2YJ2 ^a , 2YJ8 ^a , 2YJ9 ^{a,c} , 2YJB ^{a,c} , 2YJC ^a , 3HHA, 3HWN, 5F02
CatS	1MS6, 2FQ9, 2FRA, 2FRQ, 2FT2, 2FUD, 2R9M ^b , 2R9N ^b , 2R9O ^b

Notas: O ligante da estrutura 6AY2 não foi utilizado por considerar a arcação com diferenças importantes de um mimético de dipeptidil nitrila. Os ligantes das proteínas 2G7Y, 4X6I e 6F06 não foram utilizadas na docagem pois não se ligaram covalentemente à proteína. O ligante da proteína 4X6H possui mais de uma conformação em seu arquivo, por isso está identificado com o nome do ligante após *underline* para identificar a conformação utilizada. O ligante 2XU3 possui estrutura com mais de uma conformação, a utilizada é a que está com o campo de localização alternativa (*alternate location*, coluna 17 do arquivo em formato PDB) não preenchida.

Notas específicas: a) e b) identificam compostos que fazem parte de uma mesma série respectivamente.

c) identifica estruturas cristalográficas que possuem o mesmo ligante cocrystalizado.

Fonte: autoria própria.

A Figura 47 possui as estruturas 2D dos compostos mostrados na Tabela 23.

Figura 47 — Estruturas dos compostos dipeptidil nitrilas cocrystalizados com as enzimas estudadas: catepsinas B, K, L, S e cruzaina.

Notas: Compostos da Tabela 23. Ligante 4X6J é o composto CHEMBL3605409, ligante 4X6H é o composto CHEMBL3605413, ligante 4QH6 é o composto CHEMBL371466. Ligante 5TDI é o Odanacatib.
Fonte: autoria própria

Através do programa Protein Preparation Wizard¹⁹⁴ a cisteína catalítica dos receptores

(proteínas com código PDB da Tabela 23) foi modificada para glicina para evitar o impedimento estérico e permitir a aproximação do grupo eletrofílico imina (a nitrila foi modificada para imina para simular a estrutura da molécula após se ligar com a enzima). Pelo módulo PROPKA^{187,188} foi atribuído os estados de protonação das cadeias laterais dos aminoácidos no pH de 5,5, que é próximo do valor da organela lisossomo. A utilização de apenas um estado de protonação para os ligantes a serem docados é uma aproximação²³⁸, mas totalmente válida para o objetivo da docagem molecular no escopo do trabalho.

Foram realizadas duas redocagens moleculares com os programas da OEDocking (FRED e HYBRID). Os ligantes das proteínas alvo são convertidos em 2D e redocados na proteína em que está cristalizada. Experimento utilizado como validação da docagem diante de uma biblioteca de complexos inibidor-enzima, reunidos na Tabela 23. Para a métrica do método é utilizada as medidas de RMSD da molécula docada com a estrutura nativa. Os dados referentes a redocagem, extraídos dos 2 experimentos que foram agora descritos estão apresentados na Figura 48.

Figura 48 — Gráfico de barras relacionando-se o RMSD entre as estruturas redocadas com os programas HYBRID e FRED e os respectivos ligantes cognatos.

Notas: Curva azul refere-se ao programa FRED, curva vermelha refere-se ao programa HYBRID.
Fonte: autoria própria.

Os valores em sua maioria estão abaixo de 2 Å (valor usualmente aceito como bom²³⁹) e os dois algoritmos obtiveram resultado similar. Como o programa HYBRID utiliza-se de informações do ligante cristalizado, decidiu-se por utilizar o programa FRED na docagem realizada no conjunto de inibidores do filtro virtual, pois este apresenta um espaço químico mais amplo nos substituintes P1, P2 e P3 que o conjunto de ligantes dipeptidil nitrilas cocrystalizadas com o grupo de proteínas do trabalho, utilizado como base de dados no

experimento de redocagem.

4.5 Docagem do conjunto de inibidores do filtro virtual

O valor de corte para K_i para a seleção dos compostos foi considerado como 100 nM (ou $pK_i > 7$) com relação às enzimas estudadas (CatB, CatK, CatL, CatS e Cz) pois assim seria considerado somente um conjunto de inibidor com elevada potência para a enzima alvo.

Após considerar a potência, selecionou-se as estruturas que continham a nitrila como grupo reativo e que tivessem estrutura do tipo dipeptídica, ou seja, compostos miméticos de dipeptidil nitrilas. As estruturas selecionadas estão representadas na Figura 49 e compõe a base de dados intitulado conjunto de inibidores do filtro virtual que é a base de dados de compostos miméticos de dipeptidil nitrilas inibidores de ao menos uma proteína entre as classes CatB, CatK, CatL, CatS e Cz disponíveis na base de dados ChEMBL, banco de dados de moléculas virtuais com informações associadas curadas²⁴⁰.

Figura 49 — Conjunto de inibidores da base de dados ChEMBL obtido após seleção virtual.

Notas: A carga formal molecular foi estimada com o pH 5,5 (por simular o ambiente da organela lisossomo) com o programa Open Babel e conferidos com o programa Quacpac, descrito na seção métodos. O composto CHEMBL371064 é o Balicatib, CHEMBL3605409 é ligante no complexo PDB 4X6J, CHEMBL3605413 é ligante no complexo PDB 4X6H, CHEMBL371466 é ligante no complexo 4QH6. O números que rotulam os compostos é seu código ChEMBL.

Fonte: autoria própria.

Foram realizados experimentos com o programa de docagem FRED onde a pontuação atribuída pelo algoritmo de pontuação do programa está na seção ANEXO na Tabela 27. Resultados de moléculas de interesse serão apresentados e discutidos ao decorrer da próxima seção, onde as moléculas docadas serão analisadas em conjunto com o modelo seletivo encontrado pelo método integrado MIF-PCA considerando-se a potência das moléculas, Tabela 24, lembrando do objetivo de planejar moléculas que além de seletivas tenham potência frente ao alvo.

Tabela 24 — Valores de pK_i dos inibidores resgatados no filtro realizado na base de dados ChEMBL .

CÓDIGO ChEMBL	CatB	CatK	CatL	CatS
CHEMBL1651358	4,66 ^a	7,47	5,33	6,74
CHEMBL1651359	4,66 ^a	8,54	6,03	6,85
CHEMBL200161	3,91 4,49	7,46	5,55	6,84
CHEMBL200557	5,46	4,08		7,08
CHEMBL200954	5,52		7,27	6,74
CHEMBL206525	4,74			7,34
CHEMBL2153162	4,40	4,40 ^a	4,40 ^a	7,40
CHEMBL2153163	4,62	4,40 ^a	4,40 ^a	7,35
CHEMBL2153164	5,29	5,40 ^a	5,40 ^a	7,21
CHEMBL2153165	5,19	5,40 ^a	5,40 ^a	7,22
CHEMBL2153166	5,27	5,40 ^a	5,40 ^a	7,21
CHEMBL2153167	4,47	5,40 ^a	5,40 ^a	7,12
CHEMBL2153169	5,47	5,96	5,40 ^a	7,03
CHEMBL2153170	5,17	5,40 ^a	5,40 ^a	7,05
CHEMBL2153171	5,14	5,40 ^a	5,40 ^a	7,00
CHEMBL2153172	4,62	4,40 ^a	4,43	7,48
CHEMBL3605408		7,15		5,40
CHEMBL3605409		7,80	5,00	
CHEMBL3605410		8,00	4,96	
CHEMBL3605413		7,66	6,14	5,30
CHEMBL3605414		7,26		6,00
CHEMBL3640708	7,36	8,52		8,70
CHEMBL3640709	6,25	7,68		7,89
CHEMBL3640710	5,35	7,49		7,96
CHEMBL3640711	5,92	7,68		7,62
CHEMBL3640712	5,66	7,28		7,80
CHEMBL3640713	5,82	8,35		7,28
CHEMBL3640714	6,68	8,12		8,14
CHEMBL3640715	6,59	7,66		8,30
CHEMBL3640716	6,96	6,04		9,52
CHEMBL3640718	6,03	7,33		7,66
CHEMBL3640719	6,66	7,70		8,15
CHEMBL3640720	6,28	7,42		7,72
CHEMBL3640721	6,21	7,80		8,05
CHEMBL3640722	5,41	7,44		8,05
CHEMBL3640723	5,64	7,55		8,10
CHEMBL3640724	6,51	7,82		8,05
CHEMBL3640725	6,85	7,37		8,70
CHEMBL3640726	7,08	7,62		8,70
CHEMBL3640727	7,00	6,40		9,00

Tabela 24 — Valores de pK_i dos inibidores resgatados no filtro realizado na base de dados ChEMBL .

CÓDIGO ChEMBL	CatB	CatK	CatL	CatS
CHEMBL3640728	7,21	7,48		9,30
CHEMBL3640729	7,00	7,08		7,57
CHEMBL3640730	5,15	7,51		6,82
CHEMBL3640731	7,34	8,74		8,49
CHEMBL370188	6,48	5,28		7,36
	5,42	8,85	4,74	4,19
CHEMBL371064	5,54		5,49	
			6,30	
CHEMBL371466	6,48 ^a	5,82		7,11
CHEMBL3804928	4,48	6,43	8,90	5,65
CHEMBL3804975	5,91	7,61	6,15	5,58
CHEMBL3805093	7,72	6,47	5,20	6,45
	6,88	7,17	6,37	6,73
CHEMBL3805885	6,87			
	7,59			
CHEMBL4064430			7,03	
CHEMBL4065483			8,10	
CHEMBL4065829			8,40	
CHEMBL4074099			7,60	
CHEMBL4078079			7,19	
CHEMBL4080286			7,09	
CHEMBL4085387			7,85	
CHEMBL4093144			7,96	
CHEMBL4093444			7,82	
CHEMBL4095566			7,09	
CHEMBL4099235			7,09	
CHEMBL4100863			7,12	
CHEMBL4101430			8,15	
CHEMBL4102351			7,11	

Notas: Quando tem-se mais de um valor de pK_i reportado para uma mesma enzima foi usado o menor valor. Campos com o sobrescrito "a" significam que na base de dados estava reportado a relação $\geq$ para o valor de pK_i que está relatado.

Fonte: autoria própria.

4.6 Hibridização molecular

O resultado encontrado na análise da PCA destacou que a região em S3 possui uma seletividade da proteína CatK em relação às outras proteínas para grupos moleculares carregados positivamente. Dentre os compostos inibidores filtrados do ChEMBL apenas alguns possuem carga positiva no P3 (contidos na Figura 49), estes foram docados (seção anterior).

Entre estes compostos há uma série a qual o grupo P3 possui anel espiro-anelado²⁴¹ com derivatização de um anel tríciclo ligado através de um nitrogênio (CHEMBL3640727, 3640714 e 3640726). Um destes compostos (CHEMBL3640727) é mostrado na Figura 50, resultado da docagem molecular realizada com o grupo de inibidores do filtro virtual.

Figura 50 — Dipeptidil Nitrila CHEMBL3640727 com P3 carregado positivamente e pontos seletivos que representam sondas de cargas positivas docado em receptor catepsina K (PDB 1VSN).

Notas: Superfície em amarelo representa o local do resíduo catalítico para orientação espacial da enzima. As esferas são os pontos mais seletivos para a CatK segundo critérios discutidos na seção destinada à análise da PCA. O nitrogênio básico do anel encontra-se a uma distância de 1,4 Å do ponto mais próximo.

Fonte: autoria própria.

A pesquisa dessa série de dipeptidil nitrilas espiro-aneladas em P3 na empresa Sanofi estava focada na inibição da catepsina S²⁴². De acordo com a plataforma on-line SciFinder (SciFinder A CAS SOLUTION) não há SAR descrita em artigo científico onde o foco seja inibir a CatK ou outra protease, embora a potência de uma série de inibidores seja descrita na base de dados ChEMBL (Tabela 24) e tenha uma patente sobre a série de compostos (US008609681B2). O grupo P3 espiro-anelado contendo a amina na posição 4 do composto CHEMBL3640727 foi utilizado no primeiro híbrido molecular planejado, hm_1, juntamente com o arcabouço molecular da estrutura do composto Neq409⁵², sintetizado e cocrystalizado com a cruzaina pelo grupo NEQUIMED com potência descrita para catepsina L (pKi = 6,7, não testada para a catepsina K) é mostrado na parte superior da Figura 51.

Figura 51 — Proposta de inibidores da catepsina K com P3 selecionado a partir da aplicação do mapeamento de seletividade (CHEMBL360727 e CHEMBL371064 (balicatib)) e hibridização molecular com o composto Neq0409 previamente descrito como inibidor de catepsina L.

Fonte: autoria própria.

Além destes compostos, o Balicatib (CHEMBL371064) apresenta elevada seletividade para a catepsina K em relação às outras enzimas desta família (Tabela 24)²³³. Este composto foi estudado para o tratamento da osteoporose a partir do mecanismo de inibição da catepsina K e alcançou as fases clínicas²³³. Devido a complexidade sintética do composto hm_1, é proposta outra hibridização com o fragmento (*N*-propil piperazina em *para* à fenila) do composto Balicatib em P3 com o arcabouço de hm_1, dando origem ao composto hm_2, Figura 51, híbrido molecular selecionado para posterior estudo. A pose da docagem covalente com os dois híbridos planejados é mostrada na Figura 52.

Figura 52 — Híbridos moleculares 1 (hm_1) e 2 (hm_2) docados covalentemente com o programa GOLD visualizados com os pontos seletivos encontrados no modelo MIF-PCA para a catepsina K.

Notas: O receptor utilizado na docagem foi a catepsina K de código PDB ID 1VSN. Esferas em verde são pontos seletivos da sonda química cátion sódio. No híbrido molecular planejado hm_1 o nitrogênio protonado encontra-se a 2,5 Å do ponto seletivo mais próximo. No híbrido molecular planejado hm_2 o ponto seletivo mais próximo do nitrogênio protonado está a distância de 2,3 Å.

Fonte: autoria própria.

Além destes, um estudo foi realizado para o composto *in house* Neq0682, que possui uma piperazina na posição P3 (Figura 53).

Figura 53 — Estudo de docagem associado aos pontos de seletividade obtidos pela PCA sugere que a utilização de um grupo espaçador modifica o composto Neq0682 um candidato mais adequado para a catepsina K segundo hipótese encontrada no modelo seletivo MIF-PCA.

Notas: O receptor utilizado na docagem foi a catepsina K de código PDB ID 1VSN. Esferas em verde são pontos seletivos da sonda química cátion sódio. A superfície colorida de amarelo representa a cisteína catalítica ligada covalentemente ao ligante. O nitrogênio protonado em P3 do composto neq0682 está a uma distância de 3,8 Å do ponto seletivo mais próximo, enquanto que no derivado neq0682_v2 o ponto seletivo mais próximo encontra-se a uma distância de 2,5 Å do nitrogênio protonado, observando que a espaçador (grupo metileno) adicionado torna os dois nitrogênios do grupo piperazina protonado, aumentando a basicidade total de P3, o que torna o grupo mais adaptado a hipótese encontrada para catepsina K no modelo seletivo MIF-PCA.

Fonte: autoria própria.

Este composto é um bom inibidor da catepsina L ($pK_i = 6,6 \pm 0,05$), mas não foi testado para a catepsina K. Embora esta molécula tenha o grupo básico similar ao do Balicatib (Figura 51, parte inferior), Neq0682 não apresenta um grupo espaçador em P3, o que leva a uma grande distância entre o grupo básico (amina) e a região favorável à interação no sítio da enzima, com distância estimada em 3,8 Å usando os pontos obtidos no modelo da PCA e a conformação estimada por docagem no programa GOLD com a estrutura cristalográfica da CatK com código PDB 1VSN. Propondo-se portanto o derivado neq0682_v2, Figura 53, diminuído a distância para 2,5 Å.

O grupo de pontos mais seletivos encontrados na Catepsina S estão localizados no subsítio S3 e S2 e correspondem às sondas químicas íons cloreto. Um dos pontos em S3 fica a uma distância de 0,5 Å do átomo mais próximo do ligante cocrystalizado a proteína 1MS6, observações que podem ser vistas na Figura 46. Esse átomo está em um anel em P3. O composto de maior seletividade e também o mais potente (Tabela 24) é o de código

CHEMBL3640716, exibido com o resultado da docagem na Figura 54.

Figura 54 — Dipeptidil Nitrila CHEMBL3640716 e pontos seletivos que representam sondas de cargas negativas docado em receptor catepsina S (PDB 1MS6).

Notas: Superfície em amarelo representa o local do resíduo catalítico para orientação espacial da enzima. As esferas amarelas são os pontos mais seletivos para a CatS segundo critérios discutidos na seção destinada à análise da PCA. As distâncias rotuladas aos pontos mais próximos são de 2,6 e 4,0 Å. Fonte: autoria própria.

Propõe-se montar uma série de compostos utilizando o arcabouço do composto CHEMBL3640716, Figura 54, modificando-se o grupo P2 para atender aos critérios encontrados no modelo, ou seja, inserção de grupos com alta densidade eletrônica, e.g. cloro ou oxigênio (exemplos são: grupo funcional hidroxila, grupo funcional carbonila). Porém é estratégico validar o método através das propostas feitas dos compostos com seletividade para a catepsina K (hm_2, neq0682_v2) para prosseguir com o planejamento para compostos seletivos para catepsina S.

Os resultados encontrados em relação a cruzaina na seção que foi realizada a PCA obteve resultados que já são conhecidos e explorados, inclusive pelo grupo NEQUIMED. No entanto, considera-se os resultados relevantes, pois houve convergência entre o trabalho que já vem sendo realizado, que busca a seletividade em S2 através da interação seletiva com o resíduo de glutamato 208 (excluindo-se a catepsina B que também possui um resíduo de glutamato) e o resultado alcançado neste trabalho, que destacou pontos seletivos de carga positiva no subsítio S2 da cruzaina, próximos ao resíduo E208.

5 CONCLUSÃO E PERSPECTIVAS

Os sítios ativos de cisteíno proteases da família C1A (cruzaína e catepsinas B, K, L e S) foram mapeados energeticamente por meio de estudo quimiométrico (PCA) que utilizou como variável descritores moleculares 3D MIF, identificando regiões que possuem seletividade para tipos específicos de interações intermoleculares.

Grupos moleculares em S3 básicos favorecem a seletividade na catepsina K, enquanto na catepsina S grupos ácidos se mostraram seletivos neste subsítio da enzima. Pontos carregados positivamente no subsítio S2 da cruzaína encontradas pela PCA próximos ao resíduo seletivo (E208) da cruzaína convergem com a região que atualmente está sendo explorada no grupo de pesquisa.

Inibidores miméticos de dipeptidil nitrilas com potência elevada ($pK_i > 7$) descritos na literatura (ChEMBL) foram docados, o que possibilitou o estudo da complementaridade química entre inibidor e sítio ativo.

Foram planejados: um híbrido molecular (hm_2) e um derivado do composto in-house neq0682 (neq0682_v2) para a catepsina K. Estudo da viabilidade sintética e metabólica das moléculas planejadas precisam ser realizados para verificar a possibilidade de se explorar os resultados encontrados, validando o modelo ou voltando para a etapa de planejamento.

REFERÊNCIAS

- 1 Montanari, C.A (org.). *Química Medicinal: Métodos e Fundamentos em Planejamento de Fármacos*. São Paulo: EDUSP, 2011. p.
- 2 TOSCO, P.; MACKEY, M. Lessons and Successes in the Use of Molecular Fields. In: CHACKALAMANNIL, S.; ROTELLA, D. P.; WARD, S. E. (Eds). *Comprehensive Medicinal Chemistry III*. 3. ed. Brighton: Elsevier, 2017. p. 253–296.
- 3 WILLIAMS, D.H. et al. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angewandte Chemie International Edition*, v. 43, n. 48, p. 6596–6616, 2004.
- 4 SURABHI; SINGH, B.K. Computer Aided Drug Design: An Overview. *Journal of Drug Delivery and Therapeutics*, v. 8, n. 5, p. 504–509, 2018.
- 5 S. SARI. Molecular Modelling and Computer Aided Drug Design: The Skill Set Every Scientist in Drug Research Needs and Can Easily Get. *Hacettepe University Journal of the Faculty of Pharmacy*, v. 40, n. 1, p. 34–47, 2020.
- 6 CHAGAS, A.P. 100 anos de Nobel: Jacobus Henricus van't Hoff. *Química Nova na Escola*, v. 14, p. 25–27, 2001.
- 7 MONTANARI, C. A.; GAUDIO, A. C. Estratégias e Princípios do Planejamento Molecular de Fármacos. In: MONTANARI, C. A. (ORG). *Química Medicinal: Métodos e Fundamentos em Planejamento de Fármacos*. São Paulo: EDUSP, 2011. p. 210.
- 8 ZHAO, L. et al. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug Discovery Today*, v. 25, n. 9, p. 1624–1638, 2020.
- 9 BERMAN, H.M.; et al. The Protein Data Bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242, 2000.
- 10 LANGE, J. et al. Facilities that make the PDB data collection more powerful. *Protein Science*, v. 29, n. 1, p. 330–344, 2020.
- 11 TAULER, R.; PARASTAR, H. Big (Bio)Chemical Data Mining Using Chemometric Methods: A Need for Chemists. *Angewandte Chemie International Edition*, 2018.

12 VARNEK, A.; BASKIN, I.I. Chemoinformatics as a Theoretical Chemistry Discipline. *Molecular Informatics*, v. 30, n. 1, p. 20–32, 2011.

13 FOKOUE, H.H. et al. Há algo novo no reconhecimento molecular aplicado à química medicinal? *Química Nova*, v. 43, p. 78–89, 2020.

14 IUPAC. *Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*. Compilado por McNaught, A.D. e A. Wilkinson, A. Blackwell Scientific Publications, Oxford (1997). Online version (2019-) created by S. J. Chalk. ISBN 0-9678550-9-8. Disponível em <<https://doi.org/10.1351/goldbook>>. Acesso em 15 dez. 2020.

15 RODRIGUES, C.R. Processos Modernos no desenvolvimento de Fármacos: Modelagem molecular. *Cadernos Temáticos de Química Nova na Escola*, v. 3, p. 43–49, 2001.

16 BATSANOV, S.S. Van der Waals Radii of Elements. *Inorganic Materials*, v. 37, n. 9, p. 871–885, 2001.

17 Schrödinger Documentation—2020-1 Release. Disponível em: <https://www.schrodinger.com//sites/default/files/s3/mkt/Documentation/2020-1/docs/Documentation.htm#maestro_help/preferences.html#measurements>. Acesso em: 1 out. 2020.

18 BOND, A. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, v. 68, n. 3, p. 441–451, 1964.

19 WANG, X. et al. The Lennard-Jones potential: when (not) to use it. *Physical Chemistry Chemical Physics*, v. 22, n. 19, p. 10624–10633, 2020.

20 WILLIAMS, M. A.; LADBURY, J. E. Hydrogen Bonds in Protein-Ligand Complexes. In: BÖHM, H. J.; SCHNEIDE, G. (Eds.). *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*. Methods and Principles in Medicinal Chemistry. 1. ed. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2003. p. 137–161.

21 HONIG, B.; NICHOLLS, A. *Classical electrostatics in biology and chemistry*. *Science*, v. 268, n. 5214, p. 1144 LP – 1149, 1995.

22 WANG, X. et al. A quantum mechanical computational method for modeling electrostatic and solvation effects of protein. *Scientific Reports*, v. 8, n. 1, p. 5475, 2018.

23 LEMKE, T. L. et al. *Foye's principles of medicinal chemistry*. 7^a ed. Baltimore: Lippincott Williams & Wilkins, 2012. p. 36.

24 MA, S. et al. Incorporation of Hydrogen Bond Angle Dependency into the Generalized Solvation Free Energy Density Model. *Journal of Chemical Information and Modeling*, v. 58, n. 4, p. 761–772, 2018.

25 SHAHI, A.; ARUNAN, E. Why are Hydrogen Bonds Directional? *Journal of Chemical Sciences*, v. 128, n. 10, p. 1571–1577, 2016.

26 AHMED, M.; JELSCH, C. et al. Relationship between Stereochemistry and Charge Density in Hydrogen Bonds with Oxygen Acceptors. *Crystal Growth & Design*, v. 13, n. 1, p. 315–325, 2013.

27 KORTEMME, T.; MOROZOV, A. V; BAKER, D. An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein–Protein Complexes. *Journal of Molecular Biology*, v. 326, n. 4, p. 1239–1259, 2003.

28 ARUNAN, E et al. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure and Applied Chemistry*, v. 83, n. 8, p. 1637–1641, 2011.

29 KENNY, P.W. Hydrogen bonding, electrostatic potential, and molecular design. *Journal of chemical information and modeling*, v. 49, n. 5, p. 1234–44, 2009.

30 ZHOU, P. et al. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins: Structure, Function, and Bioinformatics*, v. 76, n. 1, p. 151–163, 2009.

31 KURCZAB, R. et al. Salt Bridge in Ligand–Protein Complexes—Systematic Theoretical and Statistical Investigations. *Journal of Chemical Information and Modeling*, v. 58, n. 11, p. 2224–2238, 2018.

32 HARDEGGER, L.A. et al. Systematic investigation of halogen bonding in protein-ligand interactions. *Angewandte Chemie (International ed. in English)*, v. 50, n. 1, p. 314–318, 2011.

33 HARDEGGER, L.A et al. Halogen Bonding at the Active Sites of Human Cathepsin L and MEK1 Kinase: Efficient Interactions in Different Environments. *ChemMedChem*, v. 6, n. 11, p. 2048–2054, 2011.

- 34 AYKUL, S.; MARTINEZ-HACKERT, E. Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. *Analytical Biochemistry*, v. 508, p. 97–103, 2016.
- 35 LAMEIRA, J.; BONATTO, V.; CIANNI, L.; ROCHO, F.R.; LEITÃO, A.; MONTANARI C.A. Predicting the affinity of halogenated reversible covalent inhibitors through relative binding free energy. *Physical Chemistry Chemical Physics*, v. 21, n. 44, p. 24723–24730, 2019.
- 36 HONÓRIO, K.M.; DA SILVA, A.B.F. An AM1 study on the electron-donating and electron-accepting character of biomolecules. *International Journal of Quantum Chemistry*, v. 95, n. 2, p. 126–132, 2003.
- 37 DE FREITAS, R.F.; SCHAPIRA, M. A systematic analysis of atomic protein–ligand interactions in the PDB. *MedChemComm*, v. 8, n. 10, p. 1970–1981, 2017.
- 38 KLEBE, G. Applying thermodynamic profiling in lead finding and optimization. *Nature Reviews Drug Discovery*, v. 14, n. 2, p. 95–110, 2015.
- 39 HAYAKAWA, D. et al. A molecular interaction field describing nonconventional intermolecular interactions and its application to protein–ligand interaction prediction. *Journal of Molecular Graphics and Modelling*, v. 96, p. 107515, 2020
- 40 RAWLINGS, N.D.; BARRETT, A.J.; BATEMAN, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, v. 40, n. D1, p. D343–D350, 2011.
- 41 RAWLINGS, N.D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, v. 46, n. D1, p. D624–D632, 2017.
- 42 MEROPS the Peptidase Database. Disponível em: <<https://www.ebi.ac.uk/merops/>>. Acesso em: 9 ago. 2020.
- 43 FASSLER J, COOPER P. BLAST Glossary. 2011 Jul 14. In: BLAST® Help [Internet]. Bethesda (MD): *National Center for Biotechnology Information (US)*; 2008-. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK62051/>>. Acesso em: 15 dez. 2020.
- 44 MEROPS the Peptidase Database. Disponível em <https://www.ebi.ac.uk/merops/cgi-bin/family_index?type=P>. Acesso em 26 fev. 2019.

45 MEROPS the Peptidase Database. Disponível em <https://www.ebi.ac.uk/merops/cgi-bin/clan_index?type=P>. Acesso em 04 dez. 2020.

46 LÖSER, R. et al. Interaction of papain-like cysteine proteases with dipeptide-derived nitriles. *Journal of Medicinal Chemistry*, v. 48, n. 24, p. 7688, 2005.

47 TURK, B. Targeting proteases: successes, failures and future prospects. *Nature Reviews Drug Discovery*, v. 5, p. 785, 2006.

48 SIKLOS, M.; BENAÏSSA M.; THATCHER, R.J. Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. *Acta pharmaceutica Sinica. B*, v. 5, n. 6, p. 506–519, 2015.

49 SCHECHTER, I.; BERGER, A. On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, v. 27, n. 2, p. 157–162, 1967.

50 POREBA, M.; DRAG, M. Current Strategies for Probing Substrate Specificity of Proteases. *Current Medicinal Chemistry*, v. 17, n. 33, p. 3968–3995, 2010.

51 GRZONKA, Z. et al. Structural studies of cysteine proteases and their inhibitors. *Acta Biochimica Polonica*, v. 48, n. 1, p. 1–20, 2001.

52 AVELAR, L.A.A. CAMILO, C.D.; ALBUQUERQUE, S.D.; FERNANDES, W.B.; GONÇALEZ, C.; KENNY, P.W.; LEITÃO, A.; MCKERROW, J.H.; MONTANARI, C.A.; OROZCO, E.V.M.; RIBEIRO, J.F.R.; ROCHA, J.R.; ROSINI, F.; SAIDEL, M.E. Molecular Design, Synthesis and Trypanocidal Activity of Dipeptidyl Nitriles as Cruzain Inhibitors. *PLOS Neglected Tropical Diseases*, v. 9, n. 7, p. e0003916, 2015.

53 GREENSPAN, P.D. et al. Identification of Dipeptidyl Nitriles as Potent and Selective Inhibitors of Cathepsin B through Structure-Based Drug Design. *Journal of Medicinal Chemistry*, v. 44, n. 26, p. 4524–4534, 2001.

54 ILLY, C. et al. Role of the Occluding Loop in Cathepsin B Activity. *Journal of Biological Chemistry*, v. 272, n. 2, p. 1197–1202, 1997.

55 RENKO, M. et al. Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft. *The FEBS Journal*, v. 277, n. 20, p. 4338–4345, 2010.

- 56 SCHMITZ, J. et al. Cathepsin B Inhibitors: Combining Dipeptide Nitriles with an Occluding Loop Recognition Element by Click Chemistry. *ACS Medicinal Chemistry Letters*, v. 7, n. 3, p. 211–216, 2016.
- 57 QURAIISHI, O. et al. The Occluding Loop in Cathepsin B Defines the pH Dependence of Inhibition by Its Propeptide. *Biochemistry*, v. 38, n. 16, p. 5017–5023, 1999.
- 58 REDZYNIA, I. et al. Displacement of the Occluding Loop by the Parasite Protein, Chagasin, Results in Efficient Inhibition of Human Cathepsin B. *The Journal of biological chemistry*, v. 283, n. 33, p. 22815–22825, 2008.
- 59 MUSIL, D. et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *The EMBO journal*, v. 10, n. 9, p. 2321–2330, 1991.
- 60 LI, Y.-Y.; FANG, J.; AO, G.-Z. Cathepsin B and L inhibitors: a patent review (2010 - present). *Expert Opinion on Therapeutic Patents*, v. 27, n. 6, p. 643–656, 2017.
- 61 MOHAMED, M.M.; SLOANE, B.F. Cysteine cathepsins: multifunctional enzymes in cancer. *Nature reviews. Cancer*, v. 6, n. 10, p. 764–775, 2006.
- 62 DUSCHAK, V.G.; COUTO, A.S. Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Current medicinal chemistry*, v. 16, n. 24, p. 3174–3202, 2009.
- 63 SILVA-LÓPEZ, R.E. Proteases de *Leishmania*: novos alvos para o desenvolvimento racional de fármacos. *Química Nova*, v. 33, n. 7, p. 1541–1548, 2010.
- 64 SCHMITZ, J. et al. Cathepsin B: Active site mapping with peptidic substrates and inhibitors. *Bioorganic & Medicinal Chemistry*, v. 27, n. 1, p. 1–15, 2019.
- 65 THANEI, S. et al. Cathepsin S inhibition suppresses autoimmune-triggered inflammatory responses in macrophages. *Biochemical Pharmacology*, v. 146, p. 151–164, 2017.
- 66 DAI, R. et al. Cathepsin K: The Action in and Beyond Bone. *Frontiers in Cell and Developmental Biology*, v. 8, p. 433., 2020.
- 67 AGGARWAL, N.; SLOANE, B.F. Cathepsin B: multiple roles in cancer. *Proteomics Clinical applications*, v. 8, n. 5–6, p. 427–437, 2014.

68 LANKELMA, J.M. et al. Cathepsin L, target in cancer treatment? *Life sciences*, v. 86, n. 7–8, p. 225–233, 2010.

69 OLSON, O.C.; JOYCE, J.A. Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. *Nature Reviews Cancer*, v. 15, n. 12, p. 712–729, 2015.

70 QUILLES JR, J.C.; BERNARDI, M.D.L.; BATISTA, P.H.J.; SILVA, S.C.M.; ROCHA, C.M.R.; MONTANARI, C.A.; LEITÃO, A. Biological Activity and Physicochemical Properties of Dipeptidyl Nitrile Derivatives Against Pancreatic Ductal Adenocarcinoma Cells. *Anti-Cancer Agents in Medicinal Chemistry*, v. 19, n. 1, p. 112–120, 2019.

71 DOYLE, P.S. et al. A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrobial agents and chemotherapy*, v. 51, n. 11, p. 3932–3939, 2007.

72 Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. *Doença de Chagas: 14 de abril – Dia Mundial*. Bol Epidemiol [Internet]. 2020 abr [data da citação]; 51(n.esp.):1-43. Disponível em: <<http://www.saude.gov.br/boletins-epidemiologicos>>. Acesso em 15 dez. 2020.

73 COURA, J.R.; BORGES-PEREIRA, J. Chagas disease: What is known and what should be improved: a systemic review. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 45, n. 3, p. 286–296, 2012.

74 Grennan, D. What Is a Pandemic? *JAMA*, v. 321, n. 9, p. 910, 2019.

75 CASTRO, M.F. et al. Epidemiologia da Chagas aguda no Brasil de 2007 a 2018. *Brazilian Journal of Health Review*, v. 3, n. 5, p. 11448–11460, 2020.

76 VIEIRA, M.L.F.P (organizadora). *Pesquisa de orçamentos familiares : 2017-2018 : perfil das despesas no Brasil : indicadores selecionados / IBGE, Coordenação de Trabalho e Rendimento*. Rio de Janeiro: IBGE, 2020, 119 p. Disponível em <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101761.pdf>>

77 Lima, N. de J. de F.; Farias, M.P.O. Estudo retrospectivo e transversal dos casos de doença de Chagas aguda no Brasil de 2007 a 2018. v. 14, n. 10, p. 148.

78 JUDICE, W.A.S. et al. Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzipain and its

recombinant form, cruzain, from *Trypanosoma cruzi*. *European Journal of Biochemistry*, v. 268, n. 24, p. 6578–6586, 2001.

79 JOYCE, J.A.; et al. Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell*, v. 5, n. 5, p. 443–453, 2004.

80 HOOK, V. et al. Cathepsin B in neurodegeneration of Alzheimer's disease, traumatic brain injury, and related brain disorders. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, v. 1868, n. 8, p. 140428, 2020.

81 LV, B.-J. et al. Plasma levels of cathepsins L, K, and V and risks of abdominal aortic aneurysms: a randomized population-based study. *Atherosclerosis*, v. 230, n. 1, p. 100–105, 2013.

82 ZHANG, L. et al. Cathepsin L is involved in proliferation and invasion of ovarian cancer cells. *Molecular medicine reports*, v. 11, n. 1, p. 468–474, 2015.

83 SUDHAN, D.R. et al. Cathepsin L in tumor angiogenesis and its therapeutic intervention by the small molecule inhibitor KGP94. *Clinical & experimental metastasis*, v. 33, n. 5, p. 461–473, 2016.

84 WANG, H. et al. Cathepsin L mediates the degradation of novel APP C-terminal fragments. *Biochemistry*, v. 54, n. 18, p. 2806–2816, 2015.

85 CHOWDHURY, S.R. et al. Discovery of covalent enzyme inhibitors using virtual docking of covalent fragments. *Bioorganic & Medicinal Chemistry Letters*, v. 29, n. 1, p. 36–39, 2019.

86 AHMAD, S. et al. Molecular modeling assisted identification and biological evaluation of potent cathepsin S inhibitors. *Journal of Molecular Graphics and Modelling*, v. 96, p. 107512, 2020.

87 MCGRATH, M.E. The lysosomal cysteine proteases. *Annual review of biophysics and biomolecular structure*, v. 28, p. 181–204, 1999.

88 ALMEIDA, P.C. et al. Cathepsin B activity regulation heparin-like glycosaminoglycans protect human cathepsin B from alkaline pH-induced inactivation. *Journal of Biological Chemistry*, v. 276, n. 2, p. 944–951, 2001.

- 89 VERMELHO, A.B.; RODRIGUES, G.C.; SUPURAN, C.T. Why hasn't there been more progress in new Chagas disease drug discovery? *Expert Opinion on Drug Discovery*, v. 15, n. 2, p. 145–158, 2020.
- 90 WEN, C. et al. Systematic Studies on the Protocol and Criteria for Selecting a Covalent Docking Tool. *Molecules*, v. 24, n. 11, p. 2183, 2019.
- 91 GAUTHIER, J, Y. et al. The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorganic & Medicinal Chemistry Letters*, v. 18, n. 3, p. 923–928, 2008.
- 92 MATOS, T.K.B.; BATISTA, P.H.J.; ROCHO, F.R.; DE VITA, D.; PEARCE, N.; KELLAM, B.; MONTANARI, C.A.; LEITÃO, A. Synthesis and matched molecular pair analysis of covalent reversible inhibitors of the cysteine protease CPB. *Bioorganic & Medicinal Chemistry Letters*, v. 30, n. 18, p. 127439, 2020.
- 93 STONE, J.A. et al. Clinical and translational pharmacology of the cathepsin K inhibitor odanacatib studied for osteoporosis. *British Journal of Clinical Pharmacology*, v. 85, n. 6, p. 1072–1083, 2019.
- 94 MULLARD, A. Merck & Co. drops osteoporosis drug odanacatib. *Nature Reviews Drug Discovery*, v. 15, n. 10, p. 669, 2016.
- 95 LIANG, T.-C.; ABELES, R.H. Inhibition of papain by nitriles: mechanistic studies using NMR and kinetic measurements. *Archives of biochemistry and biophysics*, v. 252, n. 2, p. 626–634, 1987.
- 96 Clayden, J. et al. *Organic Chemistry*. Nova Iorque: Oxford University Press, 2001. p. 35.
- 97 LAW, S. et al. Identification of mouse cathepsin K structural elements that regulate the potency of odanacatib. *Biochemical Journal*, v. 474, n. 5, p. 851–864, 2017.
- 98 COSTA, C.H.S.; BONATTO, V.; SANTOS, A.M.; LAMEIRA, J.; LEITÃO, A.; MONTANARI, C.A. Evaluating QM/MM Free Energy Surfaces for Ranking Cysteine Protease Covalent Inhibitors. *Journal of Chemical Information and Modeling*, v. 60, n. 2, p. 880–889, 2020.
- 99 MÊNARD, R. et al. Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. *Biochemistry*, v. 30, n. 37, p. 8924–8928, 2002.
- 100 SILVA, J.R.A.; CIANNI, L.; ARAUJO, D.; BATISTA, P.H.J.; DE VITA, D.; ROSINI, F.; LEITÃO, A.; LAMEIRA, J.; MONTANARI, C.A. Assessment of the Cruzain Cysteine Protease Reversible and

Irreversible Covalent Inhibition Mechanism. *Journal of Chemical Information and Modeling*, v. 60, n. 3, p. 1666–1677, 2020.

101 CIANNI, L.; SARTORI, G.; ROSINI, F.; DE VITA, D.; PIRES, G.; LOPES, B.R.; LEITÃO, A.; BURTULOSO, A.C.B.; MONTANARI, C.A. Leveraging the cruzain S3 subsite to increase affinity for reversible covalent inhibitors. *Bioorganic Chemistry*, v. 79, p. 285–292, 2018.

102 GOMES, J.C.; CIANNI, L.; RIBEIRO, J.F.R.; ROCHO, F.R.; SILVA, S.C.M.; BATISTA, P.H.J.; MORAES, C.B.; FRANCO, C.H.; FREITAS-JUNIOR, L.H.G.; KENNY, P.W.; LEITÃO, A.; BURTULOSO, A.C.B.; DE VITA, D.; MONTANARI, C.A. Synthesis and structure-activity relationship of nitrile-based cruzain inhibitors incorporating a trifluoroethylamine-based P2 amide replacement. *Bioorganic & Medicinal Chemistry*, v. 27, n. 22, p. 115083, 2019.

103 CIANNI, L.; ROCHO, F.R.; ROSINI, F.; BONATTO, V.; RIBEIRO, J.F.R.; LAMEIRA, J.; LEITÃO, A.; SHAMIM, A.; MONTANARI, C.A. et al. Optimization strategy of single-digit nanomolar cross-class inhibitors of mammalian and protozoa cysteine proteases. *Bioorganic Chemistry*, v. 101, p. 104039, 2020.

104 CIANNI, L.; LEMKE, C.; GILBERG, E.; FELDMANN, C.W.; ROSINI, F.; ROCHO, F.R.; RIBEIRO, J.F.R.; TEZUKA, D.Y.; LOPES, C.D.; ALBUQUERQUE, S.; BAJORATH, J.; LAUFER, S.; LEITÃO, A.; GÜTSCHOW, M.; MONTANARI, C.A. Mapping the S1 and S1' subsites of cysteine proteases with new dipeptidyl nitrile inhibitors as trypanocidal agents. *PLOS Neglected Tropical Diseases*, v. 14, n. 3, p. e0007755, 2020.

105 MARTIN, Y.C. et al. Glossary of terms used in computational drug design, part II (IUPAC Recommendations 2015). *Pure and Applied Chemistry*, v. 88, n. 3, p. 239–264, 2016.

106 LEITÃO, A.; MONTANARI, C.A.; DONNICI, C.L. Sobre o uso de métodos quimiométricos em química combinatória. *Química Nova*, v. 23, n. 2, p. 178–184, 2000.

107 NEVES, P.J.C. et al. TOP-um programa de cálculo de descritores topológicos para uso em correlações entre estrutura e atividade. *Química Nova*, v. 21, n. 6, p. 709–713, 1998.

108 TOSCO, P.; BALLE, T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *Journal of Molecular Modeling*, v. 17, n. 1, p. 201–208, 2011.

109 GOODFORD, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, v. 28, n. 7, p. 849–857, 1985.

- 110 HINCHLIFFE, A. *Molecular modelling for beginners*. 2 ed. Chichester: John Wiley & Sons, 2008. p. 57.
- 111 MACKERELL JR, A.D. BANAVALI, N.; FOLOPPE, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, v. 56, n. 4, p. 257–265, 2001.
- 112 WADE, R.C.; CLARK, K.J.; GOODFORD, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups with the Ability To Form More Than Two Hydrogen Bonds. *Journal of Medicinal Chemistry*, v. 36, n. 1, p. 140–147, 1993.
- 113 Wade, R.C.; Goodford, P.J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability To Form More Than Two Hydrogen Bonds. *Journal of Medicinal Chemistry*, v. 36, n. 1, p. 148–156, 1993.
- 114 BOOBBYER, D.N.A. et al. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *Journal of Medicinal Chemistry*, v. 32, n. 5, p. 1083–1094, 1989.
- 115 Molecular Discovery Ltd., *GRID manual*, 2004, p. 1–580.
- 116 WADE, R.C. Calculation and Application of Molecular Interaction Fields. In: CRUCIANI, G. (Ed.); *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*. Weinheim: Wiley-VCH Verlag, 2006, p. 27–38.
- 117 RODRIGO, J. et al. Comparison of biomolecules on the basis of Molecular Interaction Potentials. *Journal of the Brazilian Chemical Society*, v. 13, p. 795–799, 2002.
- 118 VON ITZSTEIN, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, v. 363, n. 6428, p. 418–423, 1993.
- 119 WIWANITKIT, V. Finding a new drug and vaccine for emerging swine flu: what is the concept? *Biologics : targets & therapy*, v. 3, p. 377–383, 2009.
- 120 OXFORD, J.S. Zanamivir (Glaxo Wellcome). *IDrugs : the investigational drugs journal*, v. 3, n. 4, p. 447–459, 2000.

- 121 VARGHESE, J.N.; EPA, V.C.; COLMAN, P.M. Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase. *Protein Science*, v. 4, n. 6, p. 1081–1087, 1995.
- 122 RICHTER, L.; ECKER, G.F. Medicinal chemistry in the era of big data. *Drug Discovery Today: Technologies*, v. 14, p. 37–41, 2015.
- 123 WALTERS, W.P. Virtual Chemical Libraries. *Journal of Medicinal Chemistry*, v. 62, n. 3, p. 1116–1124, 2019.
- 124 GAULTON, A. et al. The ChEMBL database in 2017. *Nucleic Acids Research*, v. 45, n. D1, p. D945–D954, 2016.
- 125 IRWIN, J.J. et al. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*, v. 52, n. 7, p. 1757–1768, 2012.
- 126 HANSCH, C.; FUJITA, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, v. 86, n. 8, p. 1616–1626, 1964.
- 127 PICCIRILLO, E.; AMARAL, A.T.D. Busca Virtual De Compostos Bioativos: Conceitos E Aplicações. *Química Nova*, v. 41, n. 6, p. 662–677, 2018.
- 128 CRAMER III, R.D.; PATTERSON, D.E; BUNCE, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, v. 110, n. 18, p. 5959–5967, 1988.
- 129 RAJALAHTI, T.; KVALHEIM, O.M. Multivariate data analysis in pharmaceuticals: A tutorial review. *International Journal of Pharmaceutics*, v. 417, n. 1, p. 280–290, 2011.
- 130 BRO, R.; SMILDE, A.K. Principal component analysis. *Analytical Methods*, v. 6, n. 9, p. 2812–2831, 2014.
- 131 Ferreira, M.M.C. *Quimiometria: conceitos, métodos e aplicações*. Campinas: Editora da UNICAMP, 2015, p. 7–38.
- 132 JOLLIFFE, I.T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 374, n. 2065, p. 20150202, 2016.

- 133 R-Package: Singular Value Decomposition Of A Matrix. *Singular value decomposition*. Disponível em: <<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/svd>>. Acesso em março de 2020.
- 134 REY, J.; et al. PatchSearch: a web server for off-target protein identification. *Nucleic Acids Research*, v. 47, n. W1, p. W365–W372, 2019.
- 135 PASTOR, M.; CRUCIANI, G. A Novel Strategy for Improving Ligand Selectivity in Receptor-Based Drug Design. *Journal of Medicinal Chemistry*, v. 38, n. 23, p. 4637–4647, 1995.
- 136 KASTENHOLZ, M.A.; et al. GRID/CPCA: A New Computational Tool To Design Selective Ligands. *Journal of Medicinal Chemistry*, v. 43, n. 16, p. 3033–3044, 2000.
- 137 ROCHA, J.R.; FREITAS, R.F.; MONTANARI, C.A. The GRID/CPCA approach in drug discovery. *Expert Opinion on Drug Discovery*, v. 5, n. 4, p. 333–346, 2010.
- 138 FONSECA, E.M.B.; ROCHA, J.R.; MONTANARI, C.A. On the use of GRID / CPCA approach to recognize hot spots in S2 pocket of cruzain enzyme. Brazilian Chemical Society (SBQ). Division of Medicinal Chemistry. *5th Brazilian Symposium on Medicinal Chemistry*, p. 4622, 2009.
- 139 MENEZES, I.R.A.; LEITÃO, A.; MONTANARI, C.A.. Three-dimensional models of non-steroidal ligands: A comparative molecular field analysis. *STERIODS*, v. 71, n. 6, p. 417–428, 2006.
- 140 MENEZES, I.R.A.; LEITÃO, A.; MONTANARI, C.A. Estrogen Receptor: Structural Differences and Potential Implications on Selectivity Examined by the GRID/CPCA Approach. *Letters in Drug Design & Discovery*, v. 5, n. 3, p. 182–192, 2008.
- 141 PINZI, L.; RASTELLI, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *International journal of molecular sciences*, v. 20, n. 18, p. 4331, 2019.
- 142 GIOIA, D. et al. Dynamic docking: A paradigm shift in computational drug discovery. *Molecules*, v. 22, n. 11, 2017.
- 143 KELLEY, B.P. et al. POSIT: Flexible Shape-Guided Docking For Pose Prediction. *Journal of Chemical Information and Modeling*, v. 55, n. 8, p. 1771–1780, 2015.
- 144 LIPPERT, T.; RAREY, M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *Journal of Cheminformatics*, v. 1, n. 1, p. 13, 2009.

145 SCHRÖDINGER: LLC (2009). Glide User Manual Glide v 5.5. Schrodinger Press, 2009.

146 FRIESNER, R.A. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, v. 47, n. 7, p. 1739–1749, 2004.

147 LIU, Z. et al. Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere*, v. 203, p. 139–150, 2018.

148 JORGENSEN, W.L. Rusting of the lock and key model for protein-ligand binding. *Science*, v. 254, n. 5034, p. 954LP – 955, 1991.

149 KOSHLAND JR, D.E. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English*, v. 33, n. 23-24, p. 2375–2378, 1995.

150 SAIKIA, S.; BORDOLOI, M. Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. *Current Drug Targets*, v. 20, n. 5, p. 501–521, 2019.

151 PAGADALA, N.S.; KHAJAMOHIDDIN, S.; TUSZYNSKI, J. Software for molecular docking: a review. *Biophysical Reviews*, v. 9, n. 2, p. 91–102, 2017.

152 TORRES, P.H.M. et al. Key topics in molecular docking for drug design. *International Journal of Molecular Sciences*, v. 20, n. 18, p. 4574, 2019.

153 SALMASO, V.; MORO, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Frontiers in Pharmacology*, v. 9, p. 923, 2018.

154 LIU, J.; WANG, R. Classification of Current Scoring Functions. *Journal of Chemical Information and Modeling*, v. 55, n. 3, p. 475–482, 2015.

155 LI, L., WANG, B.; MEROUEH, S.O. Support Vector Regression Scoring of Receptor–Ligand Complexes for Rank-Ordering and Virtual Screening of Chemical Libraries. *Journal of Chemical Information and Modeling*, v. 51, n. 9, p. 2132–2138, 2011.

156 OEDocking 3.1: Pose Prediction in One Awesome Package, Webinar. Disponível em: <<https://www.eyesopen.com/news/webinars/2014/07/oedocking-pose-prediction>>. Acesso em: 1 out. 2019.

157 GUEDES, I.A.; PEREIRA, F.S.S.; DARDENNE, L.E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Frontiers in Pharmacology*, v. 9, p. 1089, 2018.

158 PALACIO-RODRÍGUEZ, K. et al. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Scientific Reports*, v. 9, n. 1, p. 5142, 2019.

159 FU, D.Y.; MEILER, J. Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review. *Journal of Chemical Information and Modeling*, v. 58, n. 2, p. 225–233, 2018.

160 ERICKSEN, S.S. et al. Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. *Journal of Chemical Information and Modeling*, v. 57, n. 7, p. 1579–1590, 2017.

161 MCGANN, M. FRED pose prediction and virtual screening accuracy. *Journal of Chemical Information and Modeling*, v. 51, n. 3, p. 578–596, 2011.

162 MCGANN, M. FRED and HYBRID docking performance on standardized datasets. *Journal of Computer-Aided Molecular Design*, v. 26, n. 8, p. 897–906, 2012.

163 FRED Theory: Disponível em: <http://www.molcalx.com.cn/oe-docs/oedocking/theory/fred_theory.html>. Acesso em: 7 ago. 2020.

164 JONES, G.; WILLETT, P.; GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, v. 245, n. 1, p. 43–53, 1995.

165 JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, v. 267, n. 3, p. 727–748, 1997.

166 IVASIV, V. et al. Molecular Hybridization as a Tool for Designing Multitarget Drug Candidates for Complex Diseases. *Current Topics in Medicinal Chemistry*, v. 19, n. 19, p. 1694–1711, 2019.

167 ARAUJO, C.R.M. et al. Drug Development by Molecular Hybridization: a Medicinal Chemistry Practice Class Using Paracetamol and Sulfadiazine Tablets and the Virtual Tool. *Química Nova*, v. 38, n. 6, p. 868–873, 2015.

- 168 ZAREI, O. et al. Design, synthesis, and biological evaluation of novel benzo[b]thiophene-diaryl urea derivatives as potential anticancer agents. *Medicinal Chemistry Research*, v. 29, n. 8, p. 1438–1448, 2020.
- 169 LEITÃO, A.; ANDRICOPULO A.D.; MONTANARI, C.A.. In silico screening of HIV-1 non-nucleoside reverse transcriptase and protease inhibitors. *European Journal of Medicinal Chemistry*, v. 43, n. 7, p. 1412–1422, 2008.
- 170 WILHELM, S. et al. Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. *Nature reviews Drug discovery*, v. 5, n. 10, p. 835–844, 2006.
- 171 GARUTI, L. et al. Diaryl Urea: A Privileged Structure in Anticancer Agents. *Current Medicinal Chemistry*, v. 23, n. 15, p. 1528–1548, 2016.
- 172 PROVINCIALI, N. et al. Raloxifene hydrochloride for breast cancer risk reduction in postmenopausal women. *Expert Review of Clinical Pharmacology*, v. 9, n. 10, p. 1263–1272, 2016.
- 173 KERI, R.S. et al. An overview of benzo[b]thiophene-based medicinal chemistry. *European Journal of Medicinal Chemistry*, v. 138, p. 1002–1033, 2017.
- 174 PEREIRA, O. Á. *Dicionário de substâncias farmacêuticas comerciais*. 5ª ed. Rio de Janeiro: abiquif, 2013. p. 93.
- 175 BEALES, D.L.; BURRY, H.C.; GRAHAME, R. Comparison of aspirin and benorylate in the treatment of rheumatoid arthritis. *British medical journal*, n. 5812, p. 483–485, 1972.
- 176 HARRISON, J.R. et al. A Molecular Hybridization Approach for the Design of Potent, Highly Selective, and Brain-Penetrant N-Myristoyltransferase Inhibitors. *Journal of Medicinal Chemistry*, v. 61, n. 18, p. 8374–8389, 2018.
- 177 OCHOA, R. et al. myChEMBL: a virtual machine implementation of open data and cheminformatics tools. *Bioinformatics (Oxford, England)*, v. 30, n. 2, p. 298–300, 2014.
- 178 BENTO, A.P. et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, v. 42, n. D1, p. D1083–D1090, 2014.
- 179 GAULTON, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, v. 40, n. Database issue, p. D1100–D1107, 2012.

180 Gore, S. e colaboradores. Validation of Structures in the Protein Data Bank. *Structure*, v. 25, n. 12, p. 1916–1927, 2017.

181 SCHRÖDINGER, L. The PyMOL Molecular Graphics System, v. 2.3.2, nov. 2015. Disponível em: <<https://sourceforge.net/projects/pymol/>>.

182 Align – PyMOLWiki. Disponível em <<https://pymolwiki.org/index.php/Align>>. Acesso em dez. 2020.

183 JACQUEMARD, C. et al. Binding mode information improves fragment docking. *Journal of Cheminformatics*, v. 11, n. 1, p. 24, 2019.

184 GUEX, N.; PEITSCH, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, v. 18, n. 15, p. 2714–2723, 1997.

185 wwPDB Format version 3.3: Coordinate Section. Disponível em: <<https://www.wwpdb.org/documentation/file-format-content/format33/sect9.html>>. Acesso em: 20 fev. 2019.

186 CHOI, Y.; DEANE, C.M. FREAD revisited: Accurate loop structure prediction using a database search algorithm. *Proteins: Structure, Function, and Bioinformatics*, v. 78, n. 6, p. 1431–1440, 2010.

187 OLSSON, M.H.M. et al. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation*, v. 7, n. 2, p. 525–537, 2011.

188 SØNDERGAARD, C.R. et al. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation*, v. 7, n. 7, p. 2284–2295, 2011.

189 CAROSATI, E.; SCIABOLA, S.; CRUCIANI, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *Journal of Medicinal Chemistry*, v. 47, n. 21, p. 5114–5125, 2004.

190 Crawley, M.J. *The R book*. John Wiley & Sons, 2012.

191 WILLIAMS, T.; KELLEY, C.; AL, E. Gnuplot 5.0: an interactive plotting program, 2017. Disponível em: <<http://gnuplot.sourceforge.net/>>.

192 O'BOYLE, N.M. et al. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, v. 3, n. 1, p. 33, 2011.

193 HAWKINS, P.C.D. et al. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, v. 50, n. 4, p. 572, 2010.

194 SASTRY, G. M. et al. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, v. 27, n. 3, p. 221–234, 2013.

195 obrms - Open Babel. Disponível em: <<https://manpages.debian.org/testing/openbabel/obrms.1.en.html>>. Acesso em: 25 set. 2019.

196 HANWELL, M.D. et al. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, v. 4, n. 8, 2012.

197 HALGREN, T.A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, v. 17, 1996.

198 WARREN, G.L. et al. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry*, v. 49, n. 20, p. 5912–5931, 2006.

199 RCSB PDB PROTEIN DATABANK . Disponível em < <https://www.rcsb.org/>>. Acesso em 08 dez. 2020.

200 RCSB PDB PROTEIN DATABANK. Disponível em <https://www.rcsb.org/stats/growth/growth-xray>. Acesso em 08 dez. 2020.

201 DAVIS, A.M.; TEAGUE, S.J.; KLEYWEGT, G.J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angewandte Chemie International Edition*, v. 42, n. 24, p. 2718–2736, 2003.

202 LI, C.S. et al. Identification of a potent and selective non-basic cathepsin K inhibitor. *Bioorganic & Medicinal Chemistry Letters*, v. 16, n. 7, p. 1985–1989, 2006.

- 203 BORIŠEK, J. et al. Development of N-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles as Potent Cathepsin K Inhibitors. *Journal of Medicinal Chemistry*, v. 58, n. 17, p. 6928–6937, 2015.
- 204 ASAAD, N. et al. Dipeptidyl nitrile inhibitors of Cathepsin L. *Bioorganic & Medicinal Chemistry Letters*, v. 19, n. 15, p. 4280–4283, 2009.
- 205 BETHEL, P.A. et al. Design of selective Cathepsin inhibitors. *Bioorganic and Medicinal Chemistry Letters*, v. 19, n. 16, p. 4622–4625, 2009.
- 206 KUHN, B. et al. A Real-World Perspective on Molecular Design. *Journal of Medicinal Chemistry*, v. 59, n. 9, p. 4087–4102, 2016.
- 207 GIROUD, M.; et al. Repurposing a Library of Human Cathepsin L Ligands: Identification of Macrocyclic Lactams as Potent Rhodesain and Trypanosoma brucei Inhibitors. *Journal of Medicinal Chemistry*, v. 61, n. 8, p. 3350–3369, 2018.
- 208 WARD, Y.D. et al. Design and Synthesis of Dipeptide Nitriles as Reversible and Potent Cathepsin S Inhibitors. *Journal of Medicinal Chemistry*, v. 45, n. 25, p. 5471–5482, 2002.
- 209 SCARPINO, A.; FERENCZY, G.G.; KESERŰ, G.M. Comparative Evaluation of Covalent Docking Tools. *Journal of Chemical Information and Modeling*, v. 58, n. 7, p. 1441–1458, 2018.
- 210 SOMOZA, J.R. Human Cathepsin S with CRA-27934, a Nitrile Inhibitor. A ser publicado. Disponível em <https://www.rcsb.org/structure/2FRA>. Acesso em outubro 2020.
- 211 SOMOZA, J.R. Human Cathepsin S with Inhibitor CRA-26871. A ser publicado. Disponível em <https://www.rcsb.org/structure/2FRQ>. Acesso em outubro 2020.
- 212 SOMOZA, J.R. Human Cathepsin S with Inhibitor CRA-29728. A ser publicado. Disponível em <https://www.rcsb.org/structure/2FT2>. Acesso em outubro 2020.
- 213 SOMOZA, J.R. Human Cathepsin S with Inhibitor CRA-27566. A ser publicado. Disponível em <https://www.rcsb.org/structure/2fud>. Acesso em outubro 2020.
- 214 SOMOZA, J.R. Human Cathepsin S with inhibitor CRA-16981. A ser publicado. Disponível em <https://www.rcsb.org/structure/2g7y>. Acesso em outubro 2020.

215 Ward, Y.D. et al. Design and Synthesis of Reversible Inhibitors of Cathepsin S: alpha,alpha-Disubstitution at the P1 Residue Provides Potent Inhibitors in Cellular Assays and In Vivo Models of Antigen Presentation. A ser publicado. Disponível em <https://www.rcsb.org/structure/2R9M>. Acesso em outubro de 2020.

216 CIANNI, L.; FELDMANN, C W.; GILBERG, E.; GÜTSCHOW, M.; JULIANO, L.; LEITÃO, A.; BAJORATH, J.; MONTANARI C.A. Can Cysteine Protease Cross-class Inhibitors Achieve Selectivity? *Journal of Medicinal Chemistry*, v. 62, n. 23, p. 10497–10525, 2019.

217 ZHANG, J. Rates of Conservative and Radical Nonsynonymous Nucleotide Substitutions in Mammalian Nuclear Genes. *Journal of Molecular Evolution*, v. 50, n. 1, p. 56–68, 2000.

218 EXARCHOS, K.P. et al. Detection of discriminative sequence patterns in the neighborhood of proline cis peptide bonds and their functional annotation. *BMC Bioinformatics*, v. 10, n. 113, 2009.

219 MARQUIS, R.W. et al. Conformationally Constrained 1,3-Diamino Ketones: A Series of Potent Inhibitors of the Cysteine Protease Cathepsin K. *Journal of Medicinal Chemistry*, v. 41, n. 19, p. 3563–3567, 1998.

220 TURK, D. et al. Revised Definition of Substrate Binding Sites of Papain-Like Cysteine Proteases. *Biological chemistry*, v. 379, n. 2, p. 137–147, 1998.

221 BRAK, K. et al. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for chagas disease chemotherapy. *Journal of Medicinal Chemistry*, v. 53, n. 4, p. 1763–1773, 2010.

222 MIRKOVIĆ, B. et al. Novel Mechanism of Cathepsin B Inhibition by Antibiotic Nitroxoline and Related Compounds. *ChemMedChem*, v. 6, n. 8, p. 1351–1356, 2011.

223 TURK, D. et al. Crystal Structure of Cathepsin B Inhibited with CA030 at 2.0 Å Resolution: A Basis for the Design of Specific Epoxysuccinyl Inhibitors. *Biochemistry*, v. 34, n. 14, p. 4791–4797, 1995.

224 SEELIGER, D.; GROOT, B. de Conformational Transitions Upon Ligand Binding: Holo Structure Prediction from Apo Conformations. *PLoS computational biology*, v. 6, n. 1, p. e1000634, 2010.

225 BANNER, D. W. et al. Full wwPDB X-ray Structure Validation Report : Cathepsin L with a nitrile inhibitor. Disponível em:

<<https://www wwptdb.org/validation/2017/XrayValidationReportHelp>>. Acesso em: 3 mar. 2020.

226 WU, D.; WU, Z. Superimposition of protein structures with dynamically weighted RMSD. *Journal of Molecular Modeling*, v. 16, n. 2, p. 211–222, 2010.

227 KABSCH, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, v. 34, n. 5, p. 827–828, 1978.

228 LEVITT, M.; PARK, B.H. Water: now you see it, now you don't. *Structure*, v. 1, n. 4, p. 223–226, 1993.

229 MITUSIŃSKA, K. et al. Applications of water molecules for analysis of macromolecule properties. *Computational and Structural Biotechnology Journal*, v. 18, p. 355–365, 2020.

230 MENEZES, F.A.S.; MONTANARI, C.A.; BRUNS, R. E. 3D-WHIM pattern recognition study for bisamidines. A structure-property relationship study. *Journal of the Brazilian Chemical Society*, v. 11, n. 4, p. 393–397, 2000.

231 DESIRAJU, G.R. A Bond by Any Other Name. *Angewandte Chemie International Edition*, v. 50, n. 1, p. 52–59, 2011.

232 BLACK, W.C.; PERCIVAL, M.D. The consequences of lysosomotropism on the design of selective cathepsin K inhibitors. *Chembiochem : a European journal of chemical biology*, v. 7, n. 10, p. 1525–1535, 2006.

233 FALGUEYRET, J.-P. et al. Lysosomotropism of Basic Cathepsin K Inhibitors Contributes to Increased Cellular Potencies against Off-Target Cathepsins and Reduced Functional Selectivity. *Journal of Medicinal Chemistry*, v. 48, n. 24, p. 7535–7543, 2005.

234 GILLMOR, S.A.; CRAIK, C.S.; FLETTERICK, R.J. Structural determinants of specificity in the cysteine protease cruzain. *Protein Science*, v. 6, n. 8, p. 1603–1611, 1997.

235 CHEN, Y.T. et al. In vitro and in vivo studies of the trypanocidal properties of WRR-483 against *Trypanosoma cruzi*. *PLoS neglected tropical diseases*, v. 4, n. 9, p. e825, 2010.

236 JONES, B.D. et al. Synthesis and Evaluation of Oxyguanidine Analogues of the Cysteine Protease Inhibitor WRR-483 against Cruzain. *ACS Medicinal Chemistry Letters*, v. 7, n. 1, p. 77–82, 2016.

- 237 SOUZA, M.L. et al. Discovery of Potent, Reversible, and Competitive Cruzain Inhibitors with Trypanocidal Activity: A Structure-Based Drug Design Approach. *Journal of Chemical Information and Modeling*, v. 60, n. 2, p. 1028–1041, 2020.
- 238 SAYLE, R.A. So you think you understand tautomerism? *Journal of Computer-Aided Molecular Design*, v. 24, n. 6, p. 485–496, 2010.
- 239 SILVA, B.N.M. et al. Synthesis of New Thiosemicarbazones and Semicarbazones Containing the 1,2,3-1H-triazole-isatin Scaffold: Trypanocidal, Cytotoxicity, Electrochemical Assays, and Molecular Docking. *Medicinal Chemistry*, v. 15, n. 3, p. 240–256, 2019.
- 240 MENDEZ, D. et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Research*, v. 47, n. D1, p. D930–D940, 2019.
- 241 Sanofi (Paris). SCHUDOK, M et al. Spirocyclic nitriles as protease inhibitors. n PI US8609681, 2013. Disponível em: <<http://www.nlm.nih.gov/cgi/mesh/2009/>>. Acesso em: 22 jan. 2020.
- 242 LEE-DUTRA, A.; WIENER, D.K.; SUN, S. Cathepsin S inhibitors: 2004 – 2010. *Expert Opinion on Therapeutic Patents*, v. 21, n. 3, p. 311–337, 2011.

ANEXO

Tabela 25 — Dados do alinhamento tridimensional realizado para as cisteíno proteases da base de dados baseado na estrutura cristalográfica da cruzaina 4QH6 pelo programa PyMol.

PDB ID	RMSD/Å	Nº AMINOÁCIDOS	Cα alinhados	%Cα	Classe
1GMY	0,996	253	148	69	CatB
1CSB	0,973	254	151	70	CatB
6AY2	0,936	254	148	69	CatB
4P6G	0,723	218	153	71	CatS
2HHN	0,716	217	151	70	CatS
2AUX	0,715	215	168	78	CatK
1SNK	0,700	215	169	79	CatK
3N3G	0,700	217	151	70	CatS
2FRA	0,697	218	151	70	CatS
2HH5	0,697	217	151	70	CatS
3HWN	0,694	220	170	79	CatL
1MS6	0,693	217	151	70	CatS
3N4C	0,693	217	150	70	CatS
2R9O	0,692	217	150	70	CatS
2H7J	0,690	217	150	70	CatS
2HXZ	0,689	217	151	70	CatS
3OVZ	0,684	215	166	77	CatK
2FYE	0,684	211	150	70	CatS
3H89	0,682	220	172	80	CatL
3KWB	0,681	215	165	77	CatK
1AU0	0,680	215	163	76	CatK
1TU6	0,679	329	166	77	CatK
1YK7	0,678	215	168	78	CatK
1AU2	0,677	215	165	77	CatK
1NLJ	0,677	215	165	77	CatK
2F1G	0,676	217	149	69	CatS
2R9M	0,676	217	149	69	CatS
1AYW	0,672	215	163	76	CatK
2AUZ	0,669	215	162	75	CatK
2XU3	0,667	220	168	78	CatL
3KWZ	0,663	215	164	76	CatK
3HHA	0,663	220	170	79	CatL
1U9W	0,660	215	174	81	CatK
2FT2	0,660	218	148	69	CatS
3OF9	0,659	221	172	80	CatL
2FRQ	0,659	218	148	69	CatS
2G6D	0,657	211	148	69	CatS
2FUD	0,654	218	147	68	CatS
1NL6	0,653	215	164	76	CatK
3O1G	0,652	215	167	78	CatK

Tabela 25 — Dados do alinhamento tridimensional realizado para as cisteíno proteases da base de dados baseado na estrutura cristalográfica da cruzaina 4QH6 pelo programa PyMol.

PDB ID	RMSD/Å	Nº AMINOÁCIDOS	Cα alinhados	%Cα	Classe
1U9V	0,650	215	167	78	CatK
4P6E	0,650	218	147	68	CatS
2XU1	0,649	220	168	78	CatL
2YJB	0,649	220	171	80	CatL
1U9X	0,648	215	173	80	CatK
2R9N	0,648	217	147	68	CatS
2FQ9	0,647	218	146	68	CatS
1BGO	0,644	215	164	76	CatK
2ATO	0,644	215	164	76	CatK
3H8B	0,644	220	170	79	CatL
1AYV	0,641	215	165	77	CatK
2R6N	0,638	215	165	77	CatK
3OF8	0,638	221	168	78	CatL
4YV8	0,637	215	159	74	CatK
2G7Y	0,635	218	145	67	CatS
3KW9	0,634	215	160	74	CatK
4X6I	0,633	215	157	73	CatK
3KX1	0,632	215	158	73	CatK
3OVX	0,632	217	146	68	CatS
2BDL	0,631	215	166	77	CatK
1NQC	0,631	217	147	68	CatS
3C9E	0,629	215	163	76	CatK
4X6H	0,629	215	157	73	CatK
1AU4	0,628	215	163	76	CatK
2YJ8	0,625	220	170	79	CatL
3H7D	0,623	215	168	78	CatK
5MQY	0,623	220	168	78	CatL
2XU5	0,622	220	167	78	CatL
6F06	0,620	220	168	78	CatL
3O0U	0,619	215	160	74	CatK
4X6J	0,619	215	165	77	CatK
2XU4	0,619	220	170	79	CatL
1AU3	0,612	215	160	74	CatK
1YT7	0,611	215	157	73	CatK
1NPZ	0,611	217	149	69	CatS
5MAE	0,607	220	165	77	CatL
2YJC	0,606	220	169	79	CatL
5TDI	0,604	215	159	74	CatK
6EZP	0,603	220	168	78	CatL
5MAJ	0,601	220	166	77	CatL
4DMX	0,596	215	158	73	CatK
2YJ9	0,596	220	168	78	CatL
6EZX	0,595	220	167	78	CatL

Tabela 25 — Dados do alinhamento tridimensional realizado para as cisteíno proteases da base de dados baseado na estrutura cristalográfica da cruzaina 4QH6 pelo programa PyMol.

PDB ID	RMSD/Å	Nº AMINOÁCIDOS	Cα alinhados	%Cα	Classe
4AXM	0,594	220	168	78	CatL
5F02	0,593	220	166	77	CatL
2YJ2	0,592	220	167	78	CatL
4DMY	0,587	215	162	75	CatK
1VSN	0,582	211	160	74	CatK
1ATK	0,574	215	162	75	CatK
1AYU	0,569	215	157	73	CatK
1MEM	0,566	211	158	73	CatK
6QBS	0,563	215	166	77	CatK
1AIM	0,366	212	196	91	Cz
1EWM	0,362	212	196	91	Cz
1ME3	0,342	212	194	90	Cz
1EWL	0,339	212	199	93	Cz
1EWP	0,336	212	196	91	Cz
1F2B	0,327	212	204	95	Cz
2AIM	0,325	212	193	90	Cz
1F2A	0,320	212	201	93	Cz
1ME4	0,317	212	189	88	Cz
1U9Q	0,317	212	198	92	Cz
4XUI	0,313	215	195	91	Cz
3IUT	0,309	215	183	85	Cz
1EWO	0,302	212	202	94	Cz
3I06	0,302	215	199	93	Cz
3LXS	0,301	215	206	96	Cz
4PI3	0,297	215	205	95	Cz
3HD3	0,294	215	193	90	Cz
1F29	0,292	212	199	93	Cz
2OZ2	0,287	215	204	95	Cz
1F2C	0,257	212	197	92	Cz
3KKU	0,256	215	196	91	Cz
4KLB	0,198	215	192	89	Cz

Notas: Porcentagem é calculada com referência a estrutura de comparação, PDB ID 4QH6, com 215 aminoácidos.

Fonte: autoria própria.

Tabela 26 — Informações relativas aos parâmetros das sondas químicas virtuais do programa GRID utilizadas no cálculo dos campos de interação molecular.

Símbolo	VDWR (Å)	NEFF	ALPH (Å ³)	Q/ e	EMIM/kcal. mol ⁻¹	RMIN (Å)	JD	JA
C3	1,95	8	2,17	0,00	0,00	0,00	0	0
C1=	1,90	6	2,07	0,00	0,00	0,00	0	0
N:	1,65	6	1,10	0,00	-6,50	1,50	0	1
N1=	1,65	7	1,80	0,66	-3,00	1,60	1	0
N2:	1,70	8	1,90	0,00	-4,00	1,60	2	1
NM3	2,55	33	7,60	0,84	0,00	0,00	0	0
N1+	1,65	7	1,40	0,66	-3,00	1,60	1	0
N3+	1,75	9	2,13	0,66	-2,50	1,60	3	0
N1	1,65	7	1,40	-0,08	-4,00	1,60	1	0
N2	1,70	8	1,70	0,00	-5,30	1,50	2	0
N:=	1,65	6	1,80	0,00	-5,50	1,55	0	1
NH=	1,65	7	1,80	0,00	-3,00	1,60	1	1
N-:	1,65	6	1,80	-0,33	-3,00	1,60	0	1
N2=	1,70	8	2,10	0,66	-3,00	1,60	2	0
N:#	1,65	6	2,90	0,00	-5,00	1,55	0	1
N1#	1,65	7	2,90	0,40	-3,00	1,60	1	0
OH2	1,70	7	1,20	0,00	-4,50	1,40	2	2
OES	1,60	6	0,64	-0,20	-0,50	1,55	0	2
OC2	1,60	6	0,64	-0,20	-2,00	1,50	0	2
O1	1,65	7	1,20	-0,10	-4,50	1,40	1	2
OH	1,65	7	1,50	-0,15	-4,50	1,40	1	1
O	1,60	6	0,84	-0,25	-5,00	1,40	0	2
O::	1,60	6	2,14	-0,45	-5,00	1,40	0	2
OS	1,60	6	0,30	0,00	-5,50	1,40	0	2
ON	1,60	6	2,14	-0,10	-1,30	1,50	0	2
O=	1,60	6	2,14	-0,40	-1,50	1,50	0	2
O-	1,60	6	2,14	-0,50	-4,50	1,40	0	2
F	1,45	8	0,53	0,00	-1,20	1,35	0	2
DRY	1,65	7	1,20	0,00	-4,50	1,40	2	2
BOTH	1,65	7	1,20	0,00	-4,50	1,40	2	2
CL-	1,81	13	2,90	-1,00	0,00	0,00	0	0
NA+	0,95	8	0,24	1,00	0,00	0,00	0	0

Notas: A descrição da sonda pode ser relacionada através de seu símbolo (coluna 1). Descrição das colunas: VDWR é o raio de van Der Waals, NEFF é descrito como número efetivo de elétrons em volta da sonda, ALPH é a polarizabilidade da sonda, Q é a carga eletrostática da sonda, EMIM é o valor mínimo que a função que calcula a energia da ligação de hidrogênio resulta através da distância ótima entre os átomos. RMIN é o raio da ligação de hidrogênio da sonda sem considerar os átomos de hidrogênio. JD é o número de hidrogênios que a sonda está parametrizada para doar e JA para aceitar. Dados da tabela foram retirados do manual e dos arquivos gridlont referente a cada sonda (gerados durante o cálculo com o programa).

Fonte: retirada do arquivo gridlont, gerado ao executar a rotina grid.in no programa GRID.

Tabela 27 — Pontuação da docagem com a série de inibidores virtuais

Composto	Pose	Proteína				
		1GMV	1VSN	2YJ2	1MS6	4QH6
ChEMBL1651358	0	3,12	-3,72	5,69	3,42	x
	1	3,43	-3,43	6,03	3,46	x
	2	4,24	-3,42	9,29	4,83	x
ChEMBL1651359	0	-4,30	-7,75	-1,22	-4,39	0,04
	1	-4,03	-7,44	-1,02	-4,30	0,96
	2	-3,79	-7,34	-0,50	-4,14	1,47
ChEMBL200161	0	-8,13	-9,43	-7,34	-8,81	-8,63
	1	-7,84	-9,15	-7,33	-8,56	-7,54
	2	-7,73	-9,08	-7,09	-8,42	-7,37
ChEMBL200557	0	-10,08	-8,31	-8,99	-11,37	-9,75
	1	-9,41	-7,97	-8,63	-11,27	-9,56
	2	-9,24	-7,97	-8,13	-11,13	-9,10
ChEMBL200954	0	-9,76	-8,22	-8,87	-9,65	-6,81
	1	-9,28	-7,61	-8,43	-9,62	-6,60
	2	-9,16	-6,79	-8,14	-9,59	-6,53
ChEMBL206525	0	-6,01	-4,60	-4,13	-6,99	-4,75
	1	-5,19	-4,41	-4,12	-6,83	-4,37
	2	-4,71	-4,01	-4,09	-6,27	-4,27
ChEMBL2153162	0	-6,63	-5,98	-4,89	-8,18	-6,50
	1	-6,43	-5,96	-4,79	-7,76	-6,16
	2	-6,12	-5,71	-4,67	-7,53	-6,14
ChEMBL2153163	0	-6,75	-4,82	-4,67	-7,75	-5,62
	1	-6,20	-4,23	-4,36	-7,52	-5,12
	2	-6,15	-4,12	-4,25	-7,48	-5,06
ChEMBL2153164	0	-4,95	-4,34	-4,20	-7,50	-2,75
	1	-4,64	-4,19	-3,77	-7,43	-2,32
	2	-4,25	-4,03	-3,46	-7,23	-2,17
ChEMBL2153165	0	-4,02	-3,90	-4,76	-7,78	-2,83
	1	-4,02	-3,80	-4,37	-7,36	-2,68
	2	-3,88	-3,74	-4,32	-6,85	-2,51
ChEMBL2153166	0	-5,18	-5,26	-5,18	-7,72	-3,54
	1	-5,06	-5,19	-4,81	-7,60	-2,83
	2	-4,57	-4,68	-4,67	-7,47	-2,45
ChEMBL2153167	0	-5,96	-3,61	-6,00	-8,30	-5,55
	1	-5,96	-2,67	-5,40	-8,10	-5,24
	2	-5,76	-2,48	-5,35	-7,78	-4,84
ChEMBL2153169	0	-5,39	-4,52	-5,57	-8,19	-2,74
	1	-4,84	-4,42	-4,95	-8,18	-2,72
	2	-4,70	-4,15	-4,90	-7,92	-2,71
ChEMBL2153170	0	-5,98	-4,27	-5,65	-8,39	-5,22
	1	-5,48	-4,23	-4,91	-8,31	-4,87
	2	-5,42	-4,16	-4,89	-8,28	-4,70
ChEMBL2153171	0	-5,90	-4,86	-5,16	-8,30	-6,15
	1	-5,09	-4,83	-5,03	-8,17	-5,95
	2	-5,06	-4,00	-4,89	-8,07	-5,61
ChEMBL2153172	0	-8,59	-7,78	-7,11	-9,74	-6,42

Tabela 27 — Pontuação da docagem com a série de inibidores virtuais

Composto	Pose	Proteína				
		1GMV	1VSN	2YJ2	1MS6	4QH6
	1	-8,40	-7,49	-6,74	-9,25	-6,32
	2	-8,08	-7,08	-6,60	-9,16	-6,22
CHEMBL3605408	0	-4,82	-6,49	-4,21	-7,17	-3,66
	1	-4,60	-6,44	-3,60	-6,23	-3,62
	2	-4,38	-6,08	-3,08	-6,20	-3,61
CHEMBL3605409	0	-5,32	-6,42	-6,93	-6,33	-6,77
	1	-5,30	-6,28	-6,36	-6,01	-6,48
	2	-5,16	-6,18	-6,26	-5,85	-5,89
CHEMBL3605410	0	-5,32	-6,95	-6,90	-6,30	-6,43
	1	-5,23	-6,69	-6,68	-5,98	-6,13
	2	-4,94	-6,40	-6,47	-5,95	-6,03
CHEMBL3605413	0	-5,29	-6,27	-6,13	-7,12	-5,78
	1	-4,83	-5,94	-6,10	-6,94	-5,59
	2	-4,66	-5,10	-5,67	-6,56	-5,44
CHEMBL3605414	0	-5,37	-5,98	-4,03	-7,55	-3,80
	1	-5,19	-5,70	-3,78	-6,33	-2,78
	2	-5,05	-5,68	-3,76	-6,21	-2,72
CHEMBL3640708	0	-6,87	-9,21	-7,96	-7,39	-6,27
	1	-6,34	-8,81	-7,74	-7,18	-6,11
	2	-6,19	-8,68	-7,45	-6,67	-5,52
CHEMBL3640709	0	-5,61	-8,69	-6,44	-7,32	-5,54
	1	-5,60	-8,43	-6,31	-6,64	-5,41
	2	-5,38	-7,83	-6,25	-6,55	-5,36
CHEMBL3640710	0	-3,37	-7,83	-3,27	-6,27	-4,52
	1	-3,28	-7,28	-2,62	-6,11	-4,38
	2	-3,24	-7,10	-2,52	-5,87	-3,65
CHEMBL3640711	0	-3,86	-7,87	-5,31	-3,72	-0,80
	1	-3,85	-7,69	-3,79	-3,50	-0,60
	2	-3,77	-7,59	-3,50	-3,38	-0,17
CHEMBL3640712	0	-3,11	-7,93	-0,75	-4,49	0,48
	1	-3,04	-7,87	-0,62	-4,27	0,94
	2	-3,02	-7,30	-0,33	-4,23	1,44
CHEMBL3640713	0	-4,47	-5,92	-5,47	-6,73	-4,90
	1	-4,38	-5,86	-4,74	-6,67	-3,64
	2	-4,29	-5,79	-4,52	-6,59	-3,36
CHEMBL3640714	0	-4,36	-8,60	-1,28	-4,38	2,60
	1	-3,91	-8,55	-0,97	-3,87	3,81
	2	-3,60	-8,46	-0,86	-3,75	3,85
CHEMBL3640715	0	-6,95	-8,35	-7,82	-8,46	-6,72
	1	-6,89	-8,35	-7,65	-8,40	-6,53
	2	-6,64	-8,10	-7,31	-8,39	-6,29
CHEMBL3640716	0	-7,87	-9,12	-9,11	-9,47	-6,66
	1	-7,82	-8,37	-8,91	-9,27	-6,34
	2	-7,75	-7,80	-8,88	-9,24	-6,00
CHEMBL3640718	0	-4,76	-7,86	-5,55	-5,81	-5,23
	1	-4,39	-7,56	-5,54	-5,79	-4,28

Tabela 27 — Pontuação da docagem com a série de inibidores virtuais

Composto	Pose	Proteína				
		1GMY	1VSN	2YJ2	1MS6	4QH6
	2	-4,35	-7,52	-5,03	-5,77	-3,04
CHEMBL3640719	0	-6,21	-8,62	-6,68	-6,39	-6,92
	1	-6,01	-8,45	-6,65	-6,33	-6,37
	2	-5,94	-8,17	-6,55	-6,21	-6,14
CHEMBL3640720	0	-4,33	-8,62	-5,35	-6,06	-1,58
	1	-4,10	-7,74	-3,38	-5,89	-0,98
	2	-3,89	-7,71	-3,08	-5,82	-0,93
CHEMBL3640721	0	-5,19	-8,13	-5,52	-6,19	-5,29
	1	-5,03	-7,86	-4,98	-5,82	-4,47
	2	-4,83	-7,50	-4,72	-5,55	-4,03
CHEMBL3640722	0	-6,58	-7,28	-6,34	-6,89	-6,23
	1	-6,40	-7,05	-5,39	-6,49	-5,54
	2	-6,39	-7,04	-5,38	-6,05	-5,36
CHEMBL3640723	0	-6,47	-9,18	-6,76	-6,06	-0,90
	1	-5,09	-8,57	-5,83	-6,03	-0,88
	2	-4,61	-8,42	-5,71	-6,01	-0,86
CHEMBL3640724	0	-4,55	-7,61	-1,08	-3,93	-0,51
	1	-2,69	-7,35	-0,64	-3,66	0,39
	2	-2,39	-6,80	-0,26	-3,59	0,57
CHEMBL3640725	0	-0,81	-5,97	0,70	-2,52	x
	1	-0,25	-5,94	1,45	-2,21	x
	2	-0,01	-5,61	1,71	-1,99	x
CHEMBL3640726	0	-4,34	-8,59	-0,49	-2,41	3,94
	1	-4,33	-7,89	-0,49	-2,34	4,54
	2	-4,05	-7,82	-0,39	-2,31	9,16
CHEMBL3640727	0	-3,82	-8,57	0,26	-2,95	x
	1	-3,63	-7,83	0,84	-2,93	x
	2	-3,63	-7,57	1,06	-2,90	x
CHEMBL3640728	0	-7,74	-8,16	-7,98	-8,09	-6,89
	1	-7,32	-7,85	-7,58	-8,02	-6,51
	2	-7,20	-7,66	-7,05	-7,94	-6,34
CHEMBL3640729	0	-7,11	-8,84	-8,33	-8,59	-7,16
	1	-6,84	-8,46	-8,14	-8,46	-6,67
	2	-6,59	-7,88	-8,01	-8,14	-6,28
CHEMBL3640730	0	-6,84	-7,74	-6,42	-8,23	-5,38
	1	-6,14	-7,66	-6,40	-8,20	-4,89
	2	-5,95	-7,65	-6,37	-8,14	-4,88
CHEMBL3640731	0	-6,59	-9,06	-6,69	-6,31	-5,44
	1	-6,11	-7,96	-6,60	-6,27	-5,27
	2	-5,93	-7,90	-6,58	-6,23	-4,85
CHEMBL370188	0	-9,17	-8,96	-7,32	-10,08	-7,92
	1	-9,16	-8,73	-7,28	-9,87	-7,60
	2	-8,64	-8,19	-7,12	-9,59	-7,47
CHEMBL371064	0	-5,28	-7,43	-4,39	-7,54	-3,41
	1	-5,23	-7,10	-4,16	-7,35	-2,14
	2	-5,23	-6,91	-4,04	-7,07	-1,68

Tabela 27 — Pontuação da docagem com a série de inibidores virtuais

Composto	Pose	Proteína				
		1GMV	1VSN	2YJ2	1MS6	4QH6
CHEMBL371466	0	-9,92	-8,28	-9,27	-10,87	-9,71
	1	-9,85	-8,20	-9,12	-10,11	-8,58
	2	-8,96	-8,13	-8,78	-10,11	-8,30
CHEMBL3804928	0	-4,68	-4,12	-3,79	-4,12	-3,04
	1	-4,48	-3,98	-3,14	-3,54	-2,31
	2	-4,47	-3,85	-2,79	-3,26	-2,20
CHEMBL3804975	0	-10,30	-7,43	-6,82	-7,90	-4,08
	1	-8,83	-7,08	-6,80	-6,70	-3,99
	2	-8,70	-6,70	-6,75	-6,44	-3,70
CHEMBL3805093	0	-1,53	-3,01	2,99	1,40	x
	1	-1,11	-2,92	3,67	1,55	x
	2	-0,85	-2,85	4,10	1,62	x
CHEMBL3805885	0	-1,62	-3,13	5,50	4,62	x
	1	-1,49	-2,60	6,92	6,06	x
	2	-1,30	-2,40	8,77	7,26	x
CHEMBL4064430	0	-7,87	-7,47	-6,53	-6,82	-6,27
	1	-7,80	-7,10	-6,39	-6,18	-5,48
	2	-6,72	-6,69	-6,01	-5,98	-4,53
CHEMBL4065483	0	-7,37	-5,76	-6,64	-8,20	-4,69
	1	-7,07	-5,33	-6,17	-7,99	-4,36
	2	-7,03	-5,04	-6,15	-7,67	-4,31
CHEMBL4065829	0	-7,98	-5,83	-6,51	-8,77	-5,90
	1	-7,95	-5,79	-6,31	-8,55	-5,68
	2	-7,68	-5,22	-5,98	-8,42	-5,54
CHEMBL4074099	0	-9,08	-6,56	-5,92	-8,85	-5,39
	1	-8,98	-6,19	-5,59	-8,46	-4,15
	2	-8,43	-5,62	-5,38	-8,17	-4,09
CHEMBL4078079	0	-6,28	-6,68	-5,02	-6,03	-4,00
	1	-6,05	-6,63	-4,49	-5,70	-3,09
	2	-5,49	-6,42	-4,47	-5,43	-2,97
CHEMBL4080286	0	-6,19	-6,00	-6,35	-7,19	-5,39
	1	-6,17	-5,48	-6,11	-6,49	-5,36
	2	-6,00	-4,97	-5,89	-6,30	-5,03
CHEMBL4085387	0	-9,42	-6,25	-6,26	-9,28	-4,42
	1	-8,64	-6,02	-6,15	-7,19	-4,24
	2	-8,11	-5,83	-6,08	-6,92	-4,16
CHEMBL4093144	0	-7,70	-5,99	-5,52	-7,82	-5,15
	1	-7,63	-5,72	-5,33	-7,59	-4,81
	2	-7,47	-5,54	-5,26	-7,57	-4,14
CHEMBL4093444	0	-8,18	-5,55	-7,45	-10,08	-6,74
	1	-8,04	-5,21	-7,13	-9,97	-6,70
	2	-7,30	-5,17	-7,06	-9,78	-6,13
CHEMBL4095566	0	-6,23	-5,43	-6,33	-3,65	-5,42
	1	-5,25	-5,34	-6,25	-3,45	-5,24
	2	-4,96	-5,11	-5,66	-3,09	-4,09
CHEMBL4099235	0	-6,94	-4,19	-2,90	-8,30	-2,60

Tabela 27 — Pontuação da docagem com a série de inibidores virtuais

Composto	Pose	Proteína				
		1GMY	1VSN	2YJ2	1MS6	4QH6
	1	-6,14	-4,04	-2,76	-7,93	-1,93
	2	-4,79	-3,98	-2,62	-7,72	-1,35
CHEMBL4100863	0	-6,67	-5,09	-2,85	-5,35	-1,29
	1	-6,59	-4,67	-2,68	-4,67	-0,93
	2	-6,08	-4,54	-2,63	-4,64	0,18
CHEMBL4101430	0	-7,27	-7,56	-6,38	-8,03	-5,94
	1	-7,09	-7,17	-6,32	-7,80	-5,56
	2	-7,00	-7,04	-5,88	-7,64	-4,73
CHEMBL4102351	0	-7,62	-7,31	-6,65	-7,98	-5,62
	1	-7,06	-7,03	-6,46	-7,78	-5,52
	2	-7,00	-6,60	-6,42	-7,75	-5,16

Notas: células da tabela com valores preenchidos com “X” foram docagens sem sucesso no respectivo alvo enzimático.

Fonte: autoria própria.