



Universidade de São Paulo

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Departamento de Química

Programa de Pós-Graduação em Química

Planejamento computacional e avaliação biológica de inibidores alostéricos de glicogênio sintase cinase 3- $\beta$  com interesse na doença de Alzheimer

Guilherme Martins da Silva

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Doutor em Ciências, Área: **Química**

RIBEIRÃO PRETO – SP

2023

GUILHERME MARTINS DA SILVA

**Planejamento computacional e avaliação biológica de  
inibidores alostéricos de glicogênio sintase cinase 3- $\beta$  com  
interesse na doença de Alzheimer**

*Versão corrigida*

Tese apresentada à Faculdade de Filosofia,  
Ciências e Letras de Ribeirão Preto da Universidade de São  
Paulo, como parte das exigências para a obtenção do título de  
Doutor em Ciências

Área de Concentração: Química

Orientador: Prof. Dr. Carlos Henrique Tomich de Paula da Silva

RIBEIRÃO PRETO – SP

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### FICHA CATALOGRÁFICA

Silva, Guilherme Martins

Planejamento computacional e avaliação biológica de inibidores alostéricos de glicogênio sintase cinase 3- $\beta$  com interesse na doença de Alzheimer. Ribeirão Preto, 2023.

143 p. : il. ; 30cm

Tese de Doutorado, apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto/USP – Área de concentração: Química.

Orientador: da Silva, Carlos Henrique Tomich de Paula.

1. Doença de Alzheimer. 2. Glicogênio sintase cinase 3- $\beta$ . 3. Planejamento de fármacos. 4. Modelagem molecular. 5. QSAR.

Silva, G. M. **Planejamento computacional e avaliação biológica de inibidores alostéricos de glicogênio sintase cinase 3- $\beta$  com interesse na doença de Alzheimer**. 2023. Tese (Doutorado em Química) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

Aprovado em: **14/04/2023**

*Banca Examinadora*

Prof. Dr.: Carlos Henrique Tomich de Paula da Silva

Instituição: FCFRP – USP

Julgamento: **Aprovado**

Assinatura: \_\_\_\_\_

Profa. Dra.: Rafaela Salgado Ferreira

Instituição: ICB – UFMG

Julgamento: **Aprovado**

Assinatura: \_\_\_\_\_

Prof. Dr.: Luciano Tavares da Costa

Instituição: IQ – UFF

Julgamento: **Aprovado**

Assinatura: \_\_\_\_\_

Prof. Dr.: Giuliano César Clososki

Instituição: FFCLRP – USP

Julgamento: **Aprovado**

Assinatura: \_\_\_\_\_

*À Mariquinha e  
ao Doutor Preto,  
dedico.*

## AGRADECIMENTOS

Ao Prof. Dr. Carlos Tomich pela oportunidade e privilégio de fazer Doutorado sob a sua orientação. Obrigado por ter me acolhido tão bem e estar sempre à disposição com seus aconselhamentos valiosos, seja para ciência ou para vida. Validei contigo que é possível uma relação orientador-orientando saudável, humana, motivadora, com respeito, confiança e admiração. Espero que esse seja só o início da nossa parceria.

Aos amigos e parceiros científicos do LCQF: João, Leo, Mari, Su, Rai, Isaque e Tamara. Em especial aos veteranos (professores/pós-doutores) Lorane, Breno e Dr. Federico, pelo tempo valioso que passamos juntos e pela propulsão que vocês deram em minha carreira científica. Sem vocês todos, nada disso teria sido possível!

A todos os amigos e funcionários dos laboratórios vizinhos e da copinha do M1 (bloco M, 1º andar), FCFRP-USP, destacando: Clau, Paixão, Carla, Fuku, Gabi, Carlis, Margot, Primona, Marina, Analu, Luiz, Filipe, Ariel e Dayan. Estendo ainda os meus agradecimentos aos demais amigos de outros labs da Farma, em especial: Bia, Débora, Rodrigoão, Duda, Thamires e Iara.

Ao Prof. Dr. Alexander Tropsha pela oportunidade e acolhimento em meu doutorado sanduíche no *MML, UNC Chapel Hill*, EUA. Que honra! Também, ao Prof. Dr. Eugene Muratov e, em especial, ao Vini, pelo caminho das pedras, suporte, conselhos e amizade verdadeira. Aos amigos de lab durante a estadia nos EUA: Dan, Josh, Holli, Jon, Marielle, Zoe, Cleber, Nyssa, Kat, Travis e Kostya. Também, à família de sanduícheiros que foi lá estabelecida: Rô, Carlos, Bia e Brian.

Aos demais professores e colaboradores que contribuíram e participaram (in)diretamente deste (ou de outros) projeto(s). Também, àqueles com quem tive trocas e discussões ao longo do meu Doutorado, seja via email, em sessões de pôster, ou em conversas em eventos, que me proporcionaram relevantes *insights* científicos.

A todos os funcionários da USP que me auxiliaram, em especial à Lâmia do Dpto. de Química (FFCLRP-USP), à Denise (FFCLRP-USP) e aos demais da FCFRP-USP.

Aos meus pais, José Alberto e Maria de Lourdes: minha base, meu porto-seguro. Eternamente grato por todo o suporte e amor. Obrigado por acreditarem em mim.

À minha amada irmã, Ana Luisa, por toda a compreensão, carinho e risadas ♥

Ao meu irmão, Vinicius, meu guia de vida, que me fez chegar até aqui. Minha referência e estímulo, minha harmonia e melodia. Também à minha cunha, Carol. Vai!

À Yasmine, minha companheira. Você revirou minha vida, mas ainda assim consegui terminar o Doutorado. Obrigado por toda leveza, apoio, paciência e amor, linda!

À minha filhota canina, Fulô, minha parceirinha, que torna os meus dias mais felizes e cheios de pelos.

A todos os meus verdadeiros amigos: os de longa data e os novos, próximos ou distantes, os de copo e os de ombro. Vocês nem imaginam o quanto é fundamental passar por momentos de descontração com vocês para enfrentar um Doutorado. Agradecimento especial ao Samuka!

À ciência, de um modo geral. Ciência praticada de forma rigorosa, honesta e baseada em evidências. Desde as exatas até as sociais. Sobretudo, à ciência que, com muito sufoco, prevaleceu em tempos negacionistas e atenuou uma pandemia. Viva a ciência! Aliás, aproveito aqui pra fazer uma dedicação especial aos que se foram e aos que perderam seus próximos durante essa pandemia, assim como aos nossos povos originários e demais minorias; à quem tanto se intensificou os prejuízos nos últimos anos.

Às artes e à cultura, sobretudo à música, propulsora de sentido, sem a qual eu não teria condições de me equilibrar e de me direcionar nessa busca eterna que é a vida.

Por fim, mas não menos importante, às agências de fomento FAPESP, CNPq e CAPES que, direta ou indiretamente, viabilizaram a realização deste Projeto de Pesquisa de Doutorado.

Agradeço, ainda, ao CAPES-PrInt pelo meu doutorado sanduíche. Também, ao PPG-Química(FFCLRP-USP)/CAPES-PROEX, ao *BrazMedChem & RSC*, ao SBQ e à empresa *BioSolveIT*, pelas premiações concedidas.

*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.*

*A ciência é feita de fatos, assim como as casas são feitas de pedras; mas um mero acúmulo de fatos não é mais ciência do que uma pilha de pedras é uma casa. (Henri Poincaré)*

*“Science is the poetry of reality.” (Richard Dawkins)*

*Navegar é preciso; viver não é preciso  
Quero para mim o espírito desta frase,  
transformada a forma para a casar como eu sou:  
Viver não é necessário; o que é necessário é criar  
(Fernando Pessoa)*

*“Science may set limits to knowledge, but should not set limits to imagination.” (Bertrand Russell)*

*Quanto mais formos capazes de descobrir por que somos aquilo que somos, tanto mais nos será possível compreender por que é que a realidade é o que é. (Paulo Freire)*

*“The greatest discoveries of science have always been those that forced us to rethink our beliefs about the universe and our place in it.” (Robert L. Park)*

*Tu és meu Brasil em toda parte  
Quer na ciência ou na arte  
Portentoso e altaneiro [...]  
Não querendo levá-los ao cume da altura  
Cientistas tu tens e tens cultura  
E neste rude poema destes pobres vates  
Há sábios como Pedro Américo e César Lattes  
(Cartola)*

*“You can be standing right in front of the truth and not necessarily see it, and people only get it when they’re ready to get it.” (George Harrison)*

*O conhecimento é uma aventura em aberto. O que significa que aquilo que saberemos amanhã é algo que desconhecemos hoje; e esse algo pode mudar as verdades de ontem. (Karl Popper)*



## RESUMO

Silva, G. M. **Planejamento computacional e avaliação biológica de inibidores alostéricos de glicogênio sintase cinase 3- $\beta$  com interesse na doença de Alzheimer**. 2023. Tese (Doutorado em Química) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

A doença de Alzheimer (DA) é uma doença neurodegenerativa que ainda não possui forma eficaz de tratamento e gera graves impactos econômicos e sociais. Dentre as suas principais características fisiopatológicas, destacam-se a formação de placas amilóides e de emaranhados neurofibrilares. A enzima Glicogênio Sintase Cinase 3-beta (GSK-3 $\beta$ ) está diretamente relacionada a estas características e, portanto, a busca por inibidores desta é uma estratégia terapêutica promissora para a DA. Mais precisamente, a sua inibição seletiva e branda por inibidores alostéricos deve ser ainda mais vantajosa, diante da menor chance de se produzirem efeitos adversos. Assim, neste Projeto de Doutorado, objetivou-se realizar um planejamento computacional e subsequente avaliação biológica de inibidores alostéricos de GSK-3 $\beta$ . Primeiramente, reforçamos os conhecimentos sobre tal estratégia aplicando diversas metodologias computacionais, como detecção de cavidades, mapeamento de superfície, análise do *shape* 3D de ligantes, *docking* e dinâmica molecular. Os resultados obtidos permitiram indicar consensualmente a cavidade alostérica de GSK-3 $\beta$  mais propensa a realizar interações favoráveis e forneceram uma pose bioativa robusta para o respectivo modulador alostérico de referência. Estes foram aproveitados para, em segundo, desenvolver duas campanhas de triagem virtual, usando bases com ~ 40 milhões de compostos, com ênfase em similaridade 3D baseada em forma (ou *shape*), QSAR via aprendizagem de máquina, previsões ADME/Tox e *docking*. Dezoito compostos foram selecionados e avaliados por ensaios biológicos, dos quais cinco demonstraram preliminar inibição da GSK-3 $\beta$  e ampla diversidade química-estrutural. Em adição, suas inibições alostéricas foram validadas *in silico* por *blind docking* e estudos de ocupação de cavidades. Estes inibidores (*hits*) serão aproveitados em propostas de otimização molecular com vista à expansão do pequeno espaço químico conhecido dos moduladores alostéricos de GSK-3 $\beta$  como agentes promissores para futuro tratamento seletivo da DA.

**Palavras-chave:** doença de Alzheimer, glicogênio sintase cinase 3- $\beta$ , inibidores alostéricos, planejamento computacional de fármacos, modelagem molecular, QSAR.

## ABSTRACT

Silva, G. M. **Rational design and biological evaluation of allosteric modulators of glycogen synthase kinase 3 beta (GSK-3 $\beta$ ) related to Alzheimer's disease.** 2023. Thesis (Ph.D. in Chemistry) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

Alzheimer's disease (DA) is a neurodegenerative disease that causes memory impairment and still has no effective treatment, generating serious economic and social impacts. Among the main pathophysiological characteristics of such disorder, the formation of amyloid plaques and neurofibrillary tangles stands out. The enzyme Glycogen Synthase Kinase 3-beta (GSK-3 $\beta$ ) is directly related to these two characteristics and, therefore, the search for inhibitors of this enzyme is a promising therapeutic strategy for DA. More precisely, its selective and subtle modulation by allosteric inhibitors should be even more advantageous, given the lower chance of producing adverse effects. Thus, the goal of this PhD project was to carry out a computational design and subsequent biological evaluation of GSK-3 $\beta$  allosteric inhibitors. First, we reinforced the knowledge about such strategy by application of several computational methodologies, such as cavity detection, surface mapping, 3D shape analysis of ligands, docking, and molecular dynamics. The results obtained allowed us to consensually indicate the GSK-3 $\beta$  allosteric cavity most prone to perform favorable interactions and provided a robust bioactive pose for the respective reference GSK-3 $\beta$  allosteric modulator. Secondly, these were leveraged to develop two virtual screening campaigns, using databases with ~ 40 million compounds, with an emphasis on shape-based 3D similarity, machine learning QSAR, ADME/Tox predictions and docking. Eighteen compounds were thus selected and evaluated by biological assays, of which five demonstrated preliminary inhibition of GSK-3 $\beta$  and a wide chemical-structural diversity. Moreover, their allosteric inhibitions have been validated by blind docking and pocket coverage studies. These inhibitors (hits) shall be further employed in molecular optimization proposals with the intention of expanding the small known chemical space of GSK-3 $\beta$  allosteric modulators as promising selective agents for the future treatment of DA.

**Keywords:** Alzheimer's disease, glycogen synthase kinase 3-beta, allosteric modulators, computer-aided drug design, molecular modeling, QSAR.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1.1</b> - (a) Tipos de demência [Fonte: adaptado de (ZAGÓRSKA et al., 2023)]. (b) Número de pessoas (em milhões) vivendo com demência no mundo todo em 2019, bem como estimativas para 2030 e 2050 [Fonte: adaptado de (WORLD HEALTH ORGANIZATION, 2021)].	22
<b>Figura 1.2</b> - Representação esquemática dos diversos alvos biológicos, proteínas (em azul/roxo), enzimas (em laranja), receptores (em verde), enfatizando-se a via metabólica de $\beta$ -amilóide ( $A\beta$ ) e receptores acoplados às proteínas G (GPCR's, texto em vermelho) na doença de Alzheimer. As setas pretas representam interação molecular direta, enquanto as setas azuis com um sinal de menos representam efeito inibidor. As setas verdes com um sinal de mais representam a estimulação agonista dos GPCR's ativando as respectivas enzimas [Fonte: adaptado de Li et al., 2018].	24
<b>Figura 1.3</b> - Representação das estruturas químicas de fármacos aprovados (ou não) para DA: acetilcolina (ACh), tacrina, donepezil, rivastigmina, galantamina e memantina.	26
<b>Figura 1.4</b> - Representação da estrutura terciária da GSK-3 $\beta$ complexada com AMP-PNP na cavidade 1 de ATP. Folhas- $\beta$ em verde, alfa-hélices em vermelho e o loop de ativação em amarelo. Código PDB 1PYX (BERTRAND et al., 2003) [Fonte: (POIANI, 2017)].	28
<b>Figura 1.5</b> - Representação das estruturas químicas de alguns inibidores da GSK-3 $\beta$ .	30
<b>Figura 1.6</b> - Representação das estruturas químicas de inibidores alostéricos da GSK-3 $\beta$ , bem como os respectivos valores de IC <sub>50</sub> (em $\mu$ M) e cavidades.	32
<b>Figura 1.7</b> - Representação das cavidades 1 – 7 encontradas na estrutura 3D da GSK-3 $\beta$ , sendo: 1 - sítio de ligação de ATP; 2 - sítio de ligação de substrato e, 3 – sítio de ligação de axin/fratide. As cavidades 4, 5, 6 e 7 representam potenciais sítios alostéricos (Fonte: (SILVA et al., 2021b)).	33
<b>Figura 1.8</b> - (a) Representação do VP07 acomodado na cavidade 7, sítio alostérico, da GSK-3 $\beta$ (estrutura PDB código 1PYX) [Fonte: (PALOMO et al., 2011)]; (b) Interações entre o VP07 e resíduos de aminoácidos presentes na cavidade da GSK-3 $\beta$ [Fonte: (PALOMO et al., 2017)].	34
<b>Figura 1.9</b> - Esquema ilustrativo do processo de descoberta e desenvolvimento de fármacos [Fonte: adaptado de (DUELEN et al., 2019)].	35
<b>Figura 1.10</b> - Esquema ilustrativo representando a geração de modelos de QSAR (ou QSPR) [Fonte: adaptado de (TROPISHA, 2010)].	45
<b>Figura 1.11</b> - Representação esquemática de uma típica simulação de docking molecular. O sítio de ligação é localizado no receptor, ao qual são preditas um conjunto de poses para o ligante. As poses consistem de possíveis modelos para representar o complexo ligante-receptor. Cada pose é ranqueada por seu valor de score ( $S_1, S_2, \dots$ ), que dizem respeito a predição de afinidade de ligação ligante-receptor dos respectivos complexos [Fonte: adaptado de (BASCIU et al., 2022)].	47
<b>Figura 3.1</b> - Campanha 1 de triagem virtual exibindo as metodologias aplicadas para a obtenção de hits como moduladores alostéricos GSK-3 $\beta$ .	72
<b>Figura 3.2</b> - Campanha 2 de triagem virtual exibindo as metodologias aplicadas para a obtenção de hits como moduladores alostéricos GSK-3 $\beta$ .	78
<b>Figura 4.1</b> - (a) Cavidade detectada pelo fpocket e representada por esferas marrons (grupos polares) e brancas (grupos apolares), com a enzima representada nas formas de Superfície e Cartoon; (b) Forma da cavidade predita pelo Superstar em mesh branco, com o mapa de potencial eletrostático da enzima mapeado em sua superfície; imagens produzidas no Pymol, representando a cavidade alostérica 7 da GSK-3 $\beta$ (1PYX.pdb).	85

<b>Figura 4.2</b> - Caverna melhor pontuada pelo <b>metaPocket</b> e representada em mesh branco (cluster de sítios indicados em consenso por 3 metodologias) e esferas rosas (centro de massa do respectivo sítio global), com a enzima representada na forma de Superfície. ....	86
<b>Figura 4.3</b> - Caverna detectada pelo <b>Sitemap</b> representada por esferas brancas (demais mapas contornados são discutidos na seção 4.1.3), com a enzima representada na forma de Superfície. ....	87
<b>Figura 4.4</b> - Caverna gerada pelo webserver <b>PARS</b> indicando a caverna 7 com maior potencial de representar um sítio alostérico na GSK-3 $\beta$ (1PYX.pdb), com a enzima representada na forma de Superfície. ....	88
<b>Figura 4.5</b> - Representação das poses de docking obtidas para o <b>composto 1</b> , nas cavernas alostéricas 4, 5, 6 e 7 da proteína GSK-3 $\beta$ (1PYX.pdb), usando os programas <b>GOLD</b> , <b>Glide</b> , <b>Autodock</b> e <b>FRED</b> , com a enzima representada na forma de Superfície e a caverna destacada na cor verde. ....	89
<b>Figura 4.6</b> - (a, b, c) Mapas de contorno gerados pelo software <b>Superstar</b> , usando a função propensity, com base nos bancos de dados CSD/PDB, e usando as respectivas probes/grupos funcionais; o pontilhado de vermelho indica uma chance 2x maior do que o aleatório de se encontrar o grupo da respectiva probe naquela posição, verde 4x, azul 8x. (d) Hotspots farmacofóricos gerados pelo <b>Superstar</b> com base nos bancos de dados CSD; esfera verde (C-H), esfera azul (N), esfera vermelha (O). Imagens são da estrutura da GSK-3 $\beta$ (1PYX.pdb); quando possível, a pose do <b>composto 1</b> gerada pelo programa de docking <b>Glide</b> foi inserida em sobreposição para análise. ....	91
<b>Figura 4.7</b> - Mapas de superfície gerados pelo software <b>GRID</b> , usando as respectivas probes, em diferentes faixas de energia de interação. ....	92
<b>Figura 4.8</b> - Mapas de superfície gerados pelo <b>Sitemap</b> ; amarelo indica região hidrofóbica, vermelho região aceitadora de ligação de hidrogênio, azul região doadora de ligação de hidrogênio. ....	93
<b>Figura 4.9</b> - Curvas ROC obtidas usando a série de 40 compostos para validação dos protocolos/programas de docking. ....	95
<b>Figura 4.10</b> - Validação de docking mostrando as curvas ROC obtidas para cada programa usando (a) série de 88 compostos, e (b) série de 1938 compostos (88 + decoys). ....	97
<b>Figura 4.11</b> - Pose mais bem classificada de <b>1</b> na caverna alostérica 7 da GSK-3 $\beta$ (1PYX.pdb), obtida usando o protocolo refinado de docking pelo <b>Glide</b> . (a) representação de superfície da pose (em verde) acomodada na caverna (mapa de potencial eletrostático mapeado na superfície e produzido pelo <b>Pymol</b> ); (b) representação em stick da pose, em potenciais interações com os principais resíduos (coloridos em amarelo) da caverna (imagens do <b>Pymol</b> ); (c) diagrama 2D de interações mostrando as interações inibidor-enzima preditas pelo <b>Maestro</b> (flecha rosa: ligação de hidrogênio; fita verde: interações eletrostáticas; fita azul: interações hidrofóbicas). ....	99
<b>Figura 4.12</b> - Validação <b>ROCS</b> mostrando as curvas ROC obtidas pelo uso de cada query dos respectivos programas usando (c) série de 88 compostos, e (d) série de 88 compostos junto aos decoys. ....	101
<b>Figura 4.13</b> - Conformação da pose <b>1</b> (predita pelo <b>Glide</b> ) e validada como query para o <b>ROCS</b> , mostrando seu shape (contorno esbranquiçado) e suas características químicas (color). Em verde, os anéis aromáticos; em vermelho, grupos aceitadores de ligação de hidrogênio; em azul, grupos doadores de ligação de hidrogênio; em amarelo, grupos hidrofóbicos). ....	102
<b>Figura 4.14</b> - Conformação da pose de docking obtida para o <b>composto 1</b> (em representação stick) pelo <b>GOLD</b> e pelo <b>Glide</b> , mostrando que uma está em conformação/orientação invertida em relação à outra. A enzima é aqui representada na forma de superfície (em cor azul), com destaque para a caverna alstérica (em cor verde). ....	103
<b>Figura 4.15</b> - Gráficos de DM obtidos a partir dos complexos formados entre as poses de docking de <b>1</b> <b>Glide</b> (preto) e <b>GOLD</b> (vermelho) no sítio alostérico 7 da GSK-3 $\beta$ . (a) SASA; (b) Rg; (c) RMSF para os complexos; (d) RMSD para os complexos; (e) RMSD apenas para os ligantes (poses). ....	104
<b>Figura 4.16</b> - Representação dos shapes obtidos pelo <b>ROCS</b> para três compostos, bem como as respectivas sobreposições com o shape do query ( <b>composto 1</b> , em representação stick, com forma/contorno colorida em cinza) e seus valores de ROCSTC. (a,b)	

<b>LCQFGS03</b> , em cor magenta; (c,d) <b>LCQFGS04</b> , em cor amarela; (e,f) <b>LCQFGS06</b> , em cor salmão. Os três compostos foram selecionados como hits pela Campanha 1- vide a frente seção 4.2.3. ....	108
<b>Figura 4.17</b> - Representação das estruturas químicas 2D dos 11 hits, <b>LCQFGS01 – 11</b> , obtidos pela Campanha 1 de triagem virtual. ....	109
<b>Figura 4.18</b> - (a) Mapa de calor da matriz de correlação considerando os 28 descritores (propriedades RDKit) para a construção do Modelo 1 de QSAR via aprendizagem de máquina. (b) Importância (média) de cada descritor considerado na construção do Modelo 1.....	110
<b>Figura 4.19</b> - Validação de aleatorização de Y: box plot's para cada métrica obtida para os correspondentes modelos 1, 2 e 3 de QSAR. Métricas: CCR (TCC), Sensibilidade, PPV (VPP), Especificidade e NPV (VPN). Outliers representados por círculos brancos. ....	112
<b>Figura 4.20</b> - Representação gráfica dos valores das métricas (características estatísticas) obtidas para os modelos de QSAR gerados por aprendizagem de máquina. Linha pontilhada em vermelho indica a nota de corte de 0,6. Métricas: CCR (TCC), Sensibilidade, PPV (VPP), Especificidade e NPV (VPN).....	113
<b>Figura 4.21</b> - Representação das estruturas químicas dos 7 hits, <b>LCQFGS12 – 18</b> , obtidos pela Campanha 2 de triagem virtual. ....	114
<b>Figura 4.22</b> - Interações de docking observadas entre o sítio alostérico 7 da GSK-3 $\beta$ (1PYX.pdb) e os hits <b>LCQFGS04</b> (em representação stick, com carbonos em cor amarela) e <b>LCQFGS15</b> (em representação stick, com carbonos em cor verde). (a,d) mostrando a acomodação dos ligantes em fase (ou de acordo) com o mapa de potencial eletrostático mapeado na superfície; (b,e) principais resíduos responsáveis pelas interações-chave com os ligantes são destacados, com a enzima em representação Cartoon; (c,f) diagrama 2D das interações inibidor-enzima.....	118
<b>Figura 4.23</b> - Interações de docking observadas entre o sítio alostérico 7 da GSK-3 $\beta$ (1PYX.pdb) e o hit <b>LCQFGS03</b> , (em representação stick, com carbonos em cor magenta). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em Cartoon; em (c), as interações inibidor-enzima são representadas em diagrama 2D. ....	119
<b>Figura 4.24</b> - Interações de docking observadas entre o sítio alostérico 7 da GSK-3 $\beta$ (1PYX.pdb) e o hit <b>LCQFGS06</b> , (em representação stick, com carbonos em cor salmão). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em Cartoon; em (c), as interações inibidor-enzima são representadas em diagrama 2D. ....	119
<b>Figura 4.25</b> - Interações de docking observadas entre o sítio alostérico 7 da GSK-3 $\beta$ (1PYX.pdb) e o hit <b>LCQFGS12</b> , (em representação stick, com carbonos em cor de laranja). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em Cartoon; em (c), as interações inibidor-enzima são representadas em diagrama 2D. ....	120
<b>Figura 4.26</b> - Interações inibidor-enzima com respeito às poses de docking obtidas, observadas por diagramas 2D de interações, entre o sítio alostérico 7 da GSK-3 $\beta$ (1PYX.pdb) e os hits remanescentes. ....	121
<b>Figura 4.27</b> - Gráfico de radar mostrando as similaridades entre o <b>composto 1</b> de referência e os cinco hits. Similaridades representadas em termos do cálculo dos respectivos índices Tanimoto usando dois fingerprints de similaridade (MACCS, em vermelho, e Morgan, em verde). ....	122
<b>Figura 4.28</b> – (a) Cavidades preditas pelo <b>DoGSiteScorer</b> , na estrutura da GSK-3 $\beta$ (representada em superfície), sendo a ortostérica (poc 1) representada por esferas azuis e a alostérica (pocs 2 & 3) por esferas verdes. (b) exemplo mostrando a melhor pose obtida por blind docking para <b>LCQFGS15</b> , permitindo a visualização da sua ocupação do sítio (PC, pocket coverage). ....	123

## LISTA DE TABELAS

<b>Tabela 1.1</b> - Metodologias e algoritmos de similaridade 3D via sobreposição das estruturas moleculares e os respectivos exemplos de softwares [Fonte: adaptado de (HÖNIG; LEMMEN; RAREY, 2022)].....	42
<b>Tabela 3.1</b> - Série de 88 inibidores alostéricos de GSK-3 $\beta$ compilados da literatura <sup>(a, b, c, d)</sup> e os correspondentes valores de IC <sub>50</sub> ( $\mu$ M).....	66
<b>Tabela 3.2</b> - Dezessete bases de dados utilizadas na Campanha 1 de triagem virtual. ....	74
<b>Tabela 3.3</b> - Duas bases de dados utilizadas na Campanha 2 de triagem virtual. ....	78
<b>Tabela 4.1</b> - Valores de score obtidos por docking pelos programas <b>GOLD</b> , <b>Glide</b> , <b>Autodock</b> e <b>FRED</b> , para o <b>composto 1</b> nas quatro cavidades alostéricas (4, 5, 6 e 7) da GSK-3 $\beta$ (1PYX.pdb).....	89
<b>Tabela 4.2</b> - Valores de AUC, TG e BEDROC obtidos a partir das correspondentes curvas ROC plotadas para as validações de docking, usando a série de dados de 40 compostos.....	95
<b>Tabela 4.3</b> - Valores de AUC obtidos a partir das correspondentes curvas ROC plotadas para as validações de docking ( <b>GOLD</b> , <b>Glide</b> , <b>Autodock</b> e <b>FRED</b> ), usando série de 88 compostos sem e com decoys. ....	97
<b>Tabela 4.4</b> - Valores de AUC obtidos nas correspondentes curvas ROC plotadas para as validações de shape similarity ( <b>ROCS</b> ), usando os queries obtidos para o <b>composto 1</b> pelos respectivos programas de docking, usando série de 88 compostos sem e com decoys.....	101
<b>Tabela 4.5</b> - Características estatísticas dos modelos de QSAR gerados por aprendizagem de máquina. ....	112
<b>Tabela 4.6</b> - Compostos hits obtidos por triagem virtual e correspondentes valores % de inibição de GSK-3 $\beta$ medidos pelos ensaios in vitro. Acrescidos de seus valores de ROCSTC, QSAR score, <b>Glide XP</b> score e propriedades físico-químicas e farmacocinéticas preditas.....	115
<b>Tabela 4.7</b> - Melhor pose de blind docking para cada composto, correspondentes valores de docking score, bem como valores de ocupação de sítio (PC) na cavidade ortostérica (poc 1) e alostérica (pocs 2 e 3) da enzima GSK-3 $\beta$ . ....	123

## LISTA DE EQUAÇÕES

<i>Equação 1.1</i>	40
<i>Equação 1.2</i>	43
<i>Equação 1.3</i>	43
<i>Equação 3.1</i>	59
<i>Equação 3.2</i>	61
<i>Equação 3.3</i>	61
<i>Equação 3.4</i>	61
<i>Equação 3.5</i>	62
<i>Equação 3.6</i>	75
<i>Equação 3.7</i>	75
<i>Equação 3.8</i>	76
<i>Equação 3.9</i>	76
<i>Equação 3.10</i>	76
<i>Equação 3.11</i>	81
<i>Equação 3.12</i>	81
<i>Equação 3.13</i>	81
<i>Equação 3.14</i>	81
<i>Equação 3.15</i>	81
<i>Equação 3.16</i>	82

## LISTA DE ABREVIATURAS E SIGLAS

*A $\beta$* : peptídeo  $\beta$ -amilóide (Amyloid-beta)

*ACh*: acetilcolina

*AChE*: acetilcolinesterase

*ADME/Tox*: biodisponibilidade expressa por propriedades farmacocinéticas e toxicológicas, relacionadas aos fenômenos de Absorção, Distribuição, Metabolismo, Excreção e Toxicidade

*AMW*: average molecular weight (peso molecular médio)

*ANN*: Artificial Neural Networks (redes neurais artificiais)

*APD*: Applicability domain (Domínio de Aplicabilidade)

*APP*: Amyloid-beta precursor protein (proteína precursora de amilóide)

*ATP*: adenosina trifosfato

*AUC*: Area Under the Curve

*BACE-1*: beta-secretase ou  $\beta$ -secretase

*BEDROC*: Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic (discriminação reforçada por Boltzman do ROC)

*Beta-catenina*:  $\beta$ -catenina; age como coativador transcricional à diversos genes.

*Blind docking*: refere-se ao acoplamento de um ligante em toda e qualquer cavidade na superfície de uma proteína sem especificação de um determinado sítio de ligação.

*CADD*: Computer-Aided Drug Design (Planejamento Computacional de Fármacos)

*CCD*: Chemical Component Dictionary

*CryoEM*: microscopia eletrônica criogênica

*DA*: doença de Alzheimer

*DAn*: Discriminant Analysis (análise discriminante)

*DM*: Dinâmica molecular

*EONTC*: ET\_combo - soma dos índices de similaridade ShapeTanimoto e ET\_pb

*FDA*: Food and Drug Administration (agência regulatória dos EUA)

*FN*: número de falsos negativos

*FP*: número de falsos positivos

*GA*: Genetic Algorithms (algoritmo genético)

*GS*: Glicogênio Sintase

*GSK-3 $\beta$* : Glicogênio sintase cinase 3-beta (glycogen synthase kinase 3-beta)

*HBA*: número de grupos aceptores de ligações de hidrogênio

*HBD*: número de grupos doadores de ligações de hidrogênio

*HOA*: absorção oral humana em % pelo QikProp

*IC<sub>50</sub>*: dose ou concentração molar de um ligante que reduz ou inibe determinada atividade biológica de um receptor em 50%.

*k-NN*: k nearest neighbor (k-vizinhos mais próximos; um método de aprendizado supervisionado não paramétrico)

*LBDD*: Ligand-Based Drug Design (planejamento de fármacos baseado em ligante)

*LBVS*: Ligand-Based Virtual Screening (triagem virtual baseada em ligante)

*logD*: coeficiente de partição em n-octanol/água dependente do pH do meio

*logP*: coeficiente de partição em n-octanol/água expressando a lipofilicidade



*MIFs: Molecular Interaction Fields*

*MLR: Multiple Linear Regression (regressão linear múltipla)*

*MMFFs: Merck molecular force field*

*MW: molecular weight (peso molecular, em g/mol)*

*NMDA: N-metil-D-aspartato*

*Notch: proteínas transmembrânicas que ao se ligarem com proteínas da célula vizinha ativam uma cascata bioquímica. A via de sinalização Notch é uma via de sinalização celular que atua principalmente no desenvolvimento de células embrionárias.*

*NPT: Simulação DM baseada em número (N) e temperatura (T) constantes, pressão (P) regulada*

*NVT: Simulação DM baseada em número (N), volume (V) e temperatura (T) constantes*

*OLS: Ordinary Least Squares (mínimos quadrados ordinários)*

*OPLS: Optimized potentials for liquid simulations*

*Outlier: uma observação que apresenta um grande afastamento, ou que é inconsistente, ou valor atípico, aos demais da série.*

*PBPK: Physiologically Based Pharmacokinetic (modelagem farmacocinética baseada em fisiologia)*

*PC: pocket coverage (cobertura/ocupação de sítio calculada pelo DoGSiteScorer)*

*PCA: Principal Component Analysis (análise de componente principal)*

*PCR: Principal Component Regression (regressão de componente principal)*

*PDB: Protein Data Bank*

*PLS: Partial Least-Square (regressão parcial de mínimos quadrados)*

*PSA: área de superfície polar (polar surface área)*

*<sup>QP</sup>logBB: valores do logaritmo do coeficiente de permeabilidade da barreira hematoencefálica*

*<sup>QP</sup>logP<sub>o/w</sub>: valores do logaritmo do coeficiente de partição em n-octanol/água preditos pelo QikProp*

*QSAR score: valores de confiança/confiabilidade das predições de dado modelo de QSAR*

*QSAR: Quantitative Structure Activity Relationship (Relação Quantitativa entre Estrutura Química e Atividade Biológica)*

*QSPR: Quantitative Structure-Property Relationships*

*RF: Random Forest*

*Rg: Radius of gyrate of the protein in space (Raio total de giro da proteína no espaço)*

*RMN: Ressonância magnética nuclear*

*RMSD: Root-mean-square deviation*

*RMSF: Root-Mean-Square Fluctuation*

*RNA: ribonucleic acid (ácido ribonucleico)*

*ROC: Receiver Operating Characteristic*

*ROCSTC: ROCS\_TanimotoCombo - soma dos índices de similaridade ShapeTanimoto e ColorTanimoto*

*SASA: Solvent Accessible Surface Area (área de superfície acessível ao solvente)*

*SBDD: Structure-Based Drug Design (planejamento de fármacos baseado em estrutura)*

*SBVS: Structure-Based Virtual Screening (triagem virtual baseada em estrutura)*

*SNC: Sistema nervoso central*

*TC: Tanimoto coeficiente (coeficiente ou índice de Tanimoto)*

*TCC: taxa de classificação correta (CCR, correct classification rate)*

*TDZD: classe de compostos heterocíclicas – thiadiazolidinones*

*TG: Total Gain (ganho total padronizado)*

*TP53: Gene (p53) que produz uma proteína que se encontra dentro do núcleo das células e desempenha um papel fundamental no controle da divisão celular e da morte celular.*

*VN: número de verdadeiros negativos*

*VP: número de verdadeiros positivos*

*VPN: valor de preditividade negativa (NPV, negative predictive values)*

*VPP: valor de preditividade positiva (PPV, positive predictive values)*

*VS: Virtual Screening (triagem virtual)*

*Wnt: via de sinalização da Wnt regula diversos eventos durante o desenvolvimento embrionário e está também relacionada a renovação de células-tronco.*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
<b>1.1</b>	<b>A doença de Alzheimer: uma doença neurodegenerativa</b>	<b>21</b>
1.1.1	Causas e tratamentos para a doença de Alzheimer	22
1.1.2	Alvos biológicos e hipóteses fisiopatológicas da doença de Alzheimer	23
<b>1.2</b>	<b>Glicogênio Sintase Cinase 3-beta (GSK-3<math>\beta</math>)</b>	<b>27</b>
1.2.1	Inibidores de GSK-3 $\beta$ com interesse na DA	29
1.2.2	Inibidores alostéricos de GSK-3 $\beta$	31
<b>1.3</b>	<b>Planejamento computacional de fármacos</b>	<b>35</b>
1.3.1	Triagem virtual	37
1.3.2	Planejamento baseado em ligante	38
1.3.2.1	Similaridade 2D	39
1.3.2.2	Similaridade 3D	41
1.3.2.3	QSAR	42
1.3.3	Planejamento baseado em estrutura	46
1.3.3.1	Docking	47
1.3.4	Predições ADME/Tox	50
<b>2</b>	<b>OBJETIVOS</b>	<b>52</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>53</b>
<b>3.1</b>	<b>PARTE I: Investigação acerca das cavidades de GSK-3<math>\beta</math> e das poses bioativas dos respectivos inibidores alostéricos</b>	<b>54</b>
3.1.1	Detecção e predição de cavidades	54
3.1.2	Simulações preliminares de <i>docking</i>	56
3.1.3	Geração de mapas de contorno e superfície	64
3.1.4	Compilação da série de inibidores alostéricos	65
3.1.5	Refinamento dos protocolos de <i>docking</i>	65
3.1.6	Validação dos protocolos de <i>docking</i>	68
3.1.7	Validação de <i>query(ies)</i> para uso em similaridade 3D	70
3.1.8	Estudos de dinâmica molecular	70

<b>3.2</b>	<b>PARTE II: Triagens virtuais para a identificação de inibidores alostéricos de GSK-3<math>\beta</math></b>	<b>72</b>
3.2.1	Campanha 1: Triagem baseada em forma e em estrutura _____	72
3.2.1.1	Bases de dados (Campanha 1) _____	73
3.2.1.2	Similaridade 3D pelo ROCS e EON _____	74
3.2.2	Campanha 2: Triagem baseada em QSAR via aprendizagem de máquina e em estrutura _____	77
3.2.2.1	Bases de dados (Campanha 2) _____	78
3.2.2.2	Modelagem de QSAR _____	79
3.2.3	Predições ADME/Tox _____	82
3.2.4	Simulações de <i>docking</i> _____	83
3.2.5	Ensaio de atividade biológica com a GSK-3 $\beta$ _____	83
3.2.6	Outras análises computacionais _____	84
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> _____	<b>85</b>
<b>4.1</b>	<b>PARTE I: Reforçando a proposta de cavidades e de poses bioativas para os inibidores alostéricos de GSK-3<math>\beta</math></b> _____	<b>85</b>
4.1.1	Detecção de cavidades e predição do potencial dos sítios alostéricos _____	85
4.1.2	Estudo preliminar de <i>docking</i> : perspectiva da cavidade _____	88
4.1.3	Avaliando a pose de 1 dentro da cavidade alostérica 7 da GSK-3 $\beta$ _____	90
4.1.4	Refinamento e validação dos protocolos de <i>docking</i> : perspectiva da pose _____	94
4.1.5	Otimização das validações de <i>docking</i> _____	96
4.1.6	Validação das poses como <i>query</i> para <i>shape similarity</i> _____	99
4.1.7	Estudos de dinâmica molecular _____	103
4.1.8	Considerações finais – <i>PARTE I</i> _____	106
<b>4.2</b>	<b>PARTE II: Triagens virtuais para identificação de inibidores alostéricos de GSK-3<math>\beta</math></b>	<b>107</b>
4.2.1	Triagem virtual usando a Campanha 1 _____	107
4.2.2	Triagem virtual usando o Campanha 2 _____	109
4.2.3	Avaliação biológica e análise dos <i>hits</i> obtidos pelas triagens virtuais _____	114
4.2.3.1	Relações estrutura-atividade e diversidade estrutural _____	117
4.2.3.2	Validação <i>in silico</i> da inibição alostérica de GSK-3 $\beta$ _____	122
4.2.4	Considerações finais – <i>PARTE II</i> _____	124
<b>5</b>	<b>CONCLUSÕES E PERSPECTIVAS</b> _____	<b>126</b>
	<b>REFERÊNCIAS</b> _____	<b>129</b>
	<b>Produção Científica</b> _____	<b>142</b>

## 1 INTRODUÇÃO

### 1.1 A doença de Alzheimer: uma doença neurodegenerativa

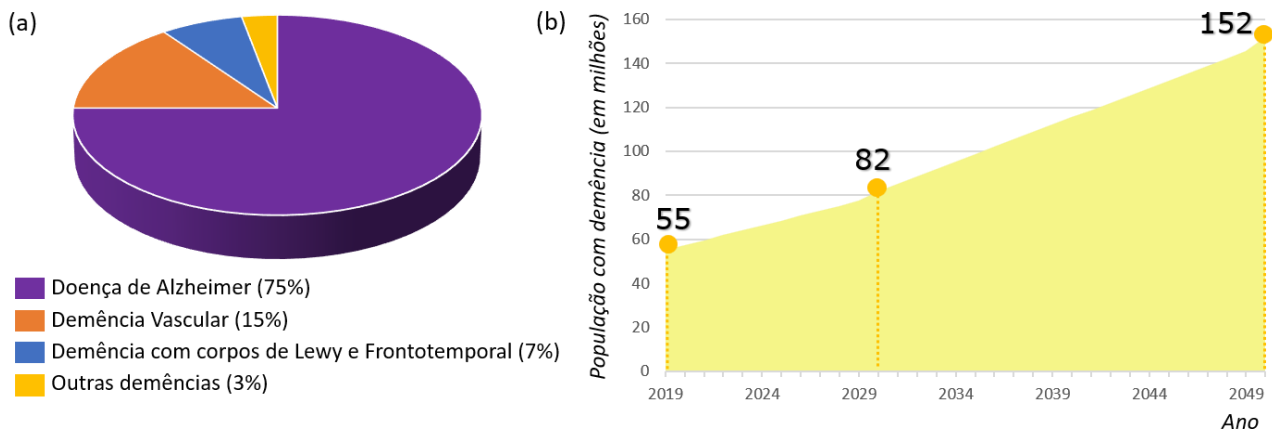
As doenças neurodegenerativas são patologias que afetam o Sistema Nervoso Central (SNC), compreendendo condições médicas que vão desde uma enxaqueca comum até patologias mais graves e de difícil tratamento, como a doença de Alzheimer, a doença de Parkinson, esclerose múltipla, outros tipos de demência, etc. (DUMURGIER; TZOURIO, 2020). Com o envelhecimento da população mundial, as doenças neurodegenerativas vem se tornando, cada vez mais, uma causa prevalente de mortalidade em idosos (STEPHENSON et al., 2018). Estima-se que existem mais de 55 milhões de pessoas vivendo com demência em todo o mundo, com um total de 9 milhões de mortes por ano causadas por doenças neurodegenerativas (ERKKINEN; KIM; GESCHWIND, 2018; CHEN; JIAO; ZHANG, 2022).

A doença de Alzheimer (DA) foi descrita pela primeira vez pelo Dr. Alois Alzheimer em 1906 e desde então tem sido considerada uma das patologias neurodegenerativas mais representativas e intrigantes do SNC (MARTINEZ; PEREZ; GIL, 2013; BUDD HAEBERLEIN; HARRIS, 2015; CARUSO et al., 2018). Representa o principal tipo de demência, como pode ser visto na **Figura 1.1**, e geralmente se manifesta clinicamente com amnésia inicial, caracterizada pela incapacidade de formar memórias recentes, bem como com comprometimento cognitivo (JAGUST, 2018). É um distúrbio crônico, de longo prazo e lentamente progressivo que também pode trazer sintomas como dificuldade de resolver problemas lógicos e declínio nas habilidades de linguagem (KIVIPELTO; MANGIALASCHE; NGANDU, 2018).

Em alguns casos, dentre 60 a 90% dos pacientes com DA, sintomas incluindo sinais de psicose, agressão verbal ou física, agitação, ansiedade, alteração de humor/comportamento e depressão são também verificadas (SHINTANI; UCHIDA, 1997; FERRERO et al., 2018). Geralmente, a progressão da DA, desde o seu diagnóstico até a aparição de sintomas mais severos, acarreta restrições na funcionalidade autônoma do paciente, que pode acabar necessitando de cuidado e vigilância constante e, assim, gerando um desgaste nos seus laços familiar e social (JANKOWSKA et al., 2018).

Estima-se que mais de 55 milhões de pessoas no mundo todo são afetadas atualmente pela DA e este número pode triplicar até 2050 (**Figura 1.1**).

Representando, portanto, um grave problema de saúde social e econômico para a humanidade, especialmente considerando-se o alto custo mundial, que fora estimado em cerca de 1 trilhão de dólares americanos em 2019. É previsto ainda que este quadro possa subir para 2 trilhões de dólares até 2030 (PATTERSON, 2018; VARESI et al., 2023).



**Figura 1.1** - (a) Tipos de demência [Fonte: adaptado de (ZAGÓRSKA et al., 2023)]. (b) Número de pessoas (em milhões) vivendo com demência no mundo todo em 2019, bem como estimativas para 2030 e 2050 [Fonte: adaptado de (WORLD HEALTH ORGANIZATION, 2021)].

### 1.1.1 Causas e tratamentos para a doença de Alzheimer

A incidência da DA possui uma forte relação com o fator envelhecimento. As chances de desenvolver DA aumenta exponencialmente após os 65 anos de idade, quase dobrando, a cada 5 anos, entre 65 e 90 anos de idade (KIVIPERTO; MANGIALASCHE; NGANDU, 2018). De acordo com a literatura (PATTERSON, 2018) uma em cada quatro pessoas com 85 anos ou mais pode apresentar sintomas de DA (FERRERO et al., 2018). Além do envelhecimento, existem outros potenciais fatores de risco associados à DA, como hipertensão, distúrbios metabólicos, obesidade na meia-idade, diabetes, depressão, sedentarismo e baixos níveis de escolaridade (JANKOWSKA et al., 2018).

Hoje, depois de mais de um século desde que a DA foi descrita pela primeira vez, ainda carecemos de terapêutica eficaz para controlar e tratar a progressão desta doença (CARUSO et al., 2018). Além da terapia medicamentosa, diversos estudos reportam sobre abordagens não farmacológicas, tais como tratamentos orientados para a cognição, exercícios físicos não invasivos de estimulação cerebral e

intervenções relacionadas ao estilo de vida (CHEN; JIAO; ZHANG, 2022; VARESI et al., 2023).

Neste contexto, um importante estudo com relação a possíveis ações preventivas em DA foi publicado por um grupo de brasileiros em 2019 (LOURENCO et al., 2018). Evidenciou-se que os níveis do hormônio chamado irisina estão de fato reduzidos nos cérebros de pacientes diagnosticados com DA. Foram realizados, portanto, experimentos com camundongos, nos quais verificou-se que a reposição deste hormônio no cérebro dos animais leva a uma melhora da memória; enfatizando-se que tal hormônio é um intermediário entre o efeito benéfico da prática de atividade física e a melhora de memória.

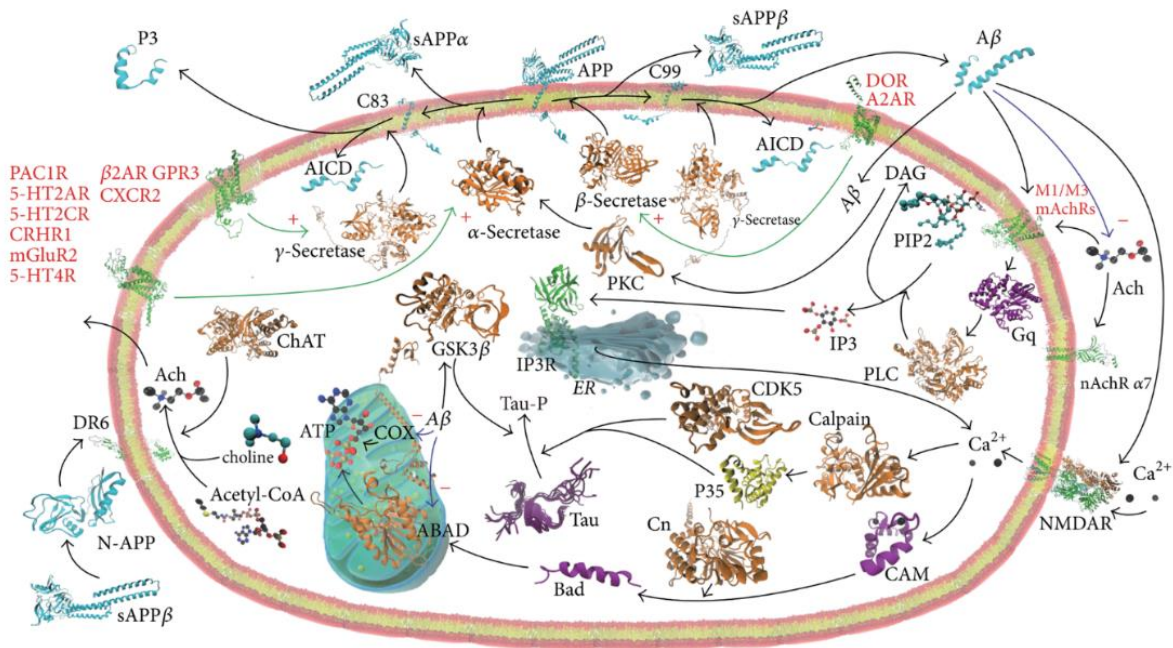
Em todo caso, as terapias de DA baseadas em fármacos são, até o momento, capazes de fornecer resultados paliativos, que apenas trazem alívio aos principais sintomas, mas que infelizmente também podem acarretar possíveis efeitos colaterais (REICHMAN, 2003; DI MARTINO et al., 2016). Além disso, os poucos fármacos aprovados para o tratamento de DA apresentam mecanismos de ação muito similares (Martinez *et al.*, 2013). A seguir, serão abordados alguns aspectos sobre as hipóteses fisiopatológicas e os principais alvos biológicos em se tratando de DA.

### **1.1.2 Alvos biológicos e hipóteses fisiopatológicas da doença de Alzheimer**

A área de conhecimento denominada fisiopatologia, visa estudar a natureza das alterações morfológicas, funções anormais ou patológicas dos vários órgãos, aparelhos e tecidos de determinado organismo. A fisiopatologia da DA é bastante complexa, especialmente devido ao seu caráter multifatorial (DE SIMONE et al., 2021; GARCÍA-MORALES et al., 2021). Entretanto, existem biomarcadores, hoje, considerados características histopatológicas primárias no cérebro de pacientes com DA, a saber: presença de placas amiloides e emaranhados neurofibrilares (discutidos na 3ª e 4ª hipóteses adiante). Embora grandes avanços vem sendo verificados em se tratando da exploração da etiologia da DA, especificidades da fisiopatologia da doença ainda carecem de elucidações (GUO et al., 2022).

Vale lembrar que as hipóteses fisiopatológicas estão geralmente relacionadas a um (ou múltiplos) alvo(s) biológico(s) - por exemplo representados por determinada enzima, proteína, receptor, etc. - que desempenha um papel importante ou está relacionado a alguma via de sinalização que ocasiona os sintomas vinculados ao

distúrbio. Além dos alvos biológicos aqui mencionados, muitos outros também têm sido relacionados às hipóteses fisiopatológicas de DA e, assim, têm sido utilizados em estudos de planejamento de fármacos para a DA, incluindo-se aqueles envolvendo métodos computacionais (VAN DER SCHYF, 2011; BUDD HAEBERLEIN; HARRIS, 2015; IBRAHIM; GABR, 2019).



**Figura 1.2** - Representação esquemática dos diversos alvos biológicos, proteínas (em azul/roxo), enzimas (em laranja), receptores (em verde), enfatizando-se a via metabólica de  $\beta$ -amilóide ( $A\beta$ ) e receptores acoplados às proteínas G (GPCR's, texto em vermelho) na doença de Alzheimer. As setas pretas representam interação molecular direta, enquanto as setas azuis com um sinal de menos representam efeito inibidor. As setas verdes com um sinal de mais representam a estimulação agonista dos GPCR's ativando as respectivas enzimas [Fonte: adaptado de Li *et al.*, 2018].

Existem propostas de diversas hipóteses fisiopatológicas, mecanismos e alvos biológicos para explicar as causas da DA, como observado na **Figura 1.2** (LI *et al.*, 2018). Essas tem sido, geralmente, propostas a partir dos sintomas nos pacientes ou de lesões histopatológicas encontradas no cérebro dos pacientes com DA (MARTINEZ; PEREZ; GIL, 2013). Os sintomas clínicos mais comuns da DA são característicos da degeneração cerebral e disfunção neuronal. A disfunção neuronal na DA é um processo que pode envolver problemas fisiopatológicos desde a sinapse até conectividade neuronal (SHINTANI; UCHIDA, 1997; FERRERO *et al.*, 2018).

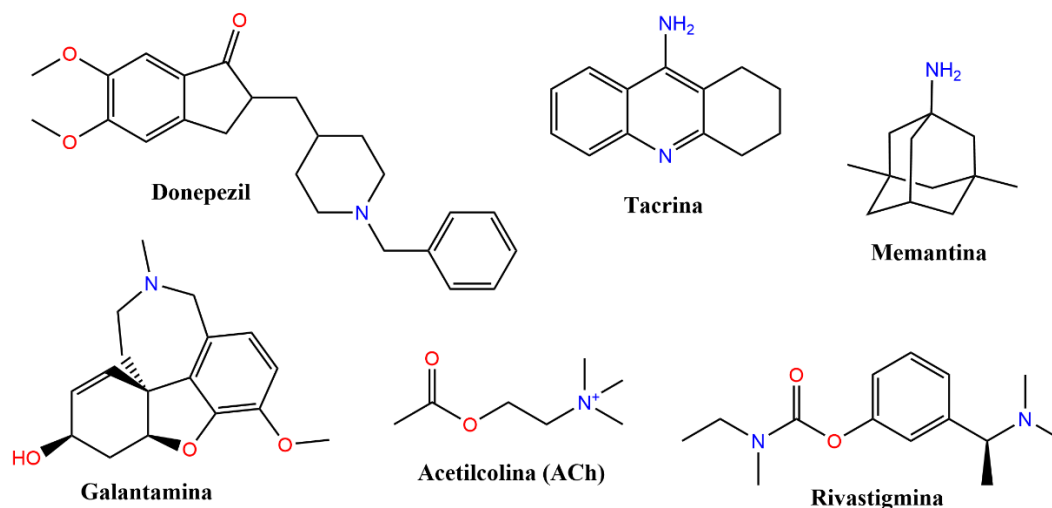
A primeira hipótese fisiopatológica atribuída a DA foi a hipótese colinérgica. Nesta, foi feita uma associação entre a morte ou disfunção dos neurônios colinérgicos,



progressivamente com o avanço da doença. Em outras palavras, considerou-se a doença como consequência do déficit de **acetilcolina (ACh)**. Desta forma, buscou-se a obtenção de fármacos que atuassem no aumento da eficiência de neurotransmissão colinérgica e, de fato, as formas disponíveis de tratamento de DA, atualmente, são baseadas neste mecanismo. Isto é possível através da inibição da enzima acetilcolinesterase (AChE), a qual atua nas sinapses colinérgicas do sistema nervoso central e periférico, sendo responsável pela interrupção da transmissão do impulso nervoso colinérgico através da hidrólise da **ACh** (GIACOBINI, 1998; DA SILVA; CARVALHO; TAFT, 2007; DE ALMEIDA et al., 2018).

Dentre os fármacos utilizados para o tratamento de DA estão a **tacrina** e o **donepezil** (vide fórmulas estruturas na **Figura 1.3**). A **tacrina** foi o primeiro inibidor de AChE licenciado pelo *FDA (Food and Drug Administration)* para o tratamento da DA. Trata-se de um inibidor reversível, não competitivo e não seletivo para AChE. O seu uso desencadeia efeitos colaterais hepatotóxicos e geralmente há a necessidade do uso de até quatro doses diárias, por isso foi descontinuada em 2013. O **donepezil** é o fármaco anticolinesterásico mais utilizado para DA. É um inibidor reversível, não competitivo e, ao contrário da **tacrina**, seletivo para a AChE. Os efeitos colaterais proporcionados pelo uso do **donepezil** são de natureza gastrointestinal, estando diretamente relacionados à dosagem administrada. Neste contexto, menciona-se ainda a **rivastigmina** e a **galantamina** (REICHMAN, 2003; GOULART et al., 2021).

Uma segunda característica observada em pacientes com DA é a disfunção do glutamato, processo esse que ocorre por meio do receptor *N*-metil-*D*-aspartato (NMDA). Em determinadas condições celulares ocorre uma ativação excessiva deste receptor, que é responsável, entre outras coisas, pela regulação do balanço eletrolítico celular glutamato-dependente. Este processo estaria desregulado em DA, contribuindo para aumento na toxicidade e consequente morte neuronal. Assim um antagonismo de NMDA seria responsável por reverter este processo. Um fármaco aprovado para esta hipótese é a **memantina**, que apesar de trazer melhoras cognitivas nos pacientes com DA pode também acarretar efeitos colaterais (GREENAMYRE et al., 1988; JANKOWSKA et al., 2018).



**Figura 1.3** - Representação das estruturas químicas de fármacos aprovados e/ou neurotransmissores relacionados a DA: acetilcolina (ACh), tacrina, donepezil, rivastigmina, galantamina e memantina.

Uma terceira característica da DA, pode ser observada, microscopicamente, pela presença de placas amiloides. As placas amilóides são compostas pelo peptídeo  $\beta$ -amilóide ( $A\beta$ ), o qual é derivado de clivagens sequenciais da proteína precursora de amilóide (*APP*). A produção excessiva de  $A\beta$  seria responsável pela neurotoxicidade que acarreta a progressiva morte neuronal em DA. Um possível alvo biológico neste contexto seria a  $\beta$ -secretase (*BACE-1*) que representa a primeira protease a ser clivada no processo de degradação da *APP* e leva a formação de  $A\beta$ . Como exemplo, o composto de nome **semagacestat** foi planejado com o objetivo de bloquear a  $\gamma$ -secretase (que cliva *APP* junto com a *BACE-1*), no entanto o seu desenvolvimento foi interrompido nas fases de testes clínicos. Estudos sobre a influência e a relação da  $A\beta$  com os sintomas da DA ainda não são totalmente conhecidos, mas existem demais estudos em andamento com base nesta hipótese (MARTINEZ; PEREZ; GIL, 2013; PINHIRO et al., 2015; DE ALMEIDA et al., 2018).

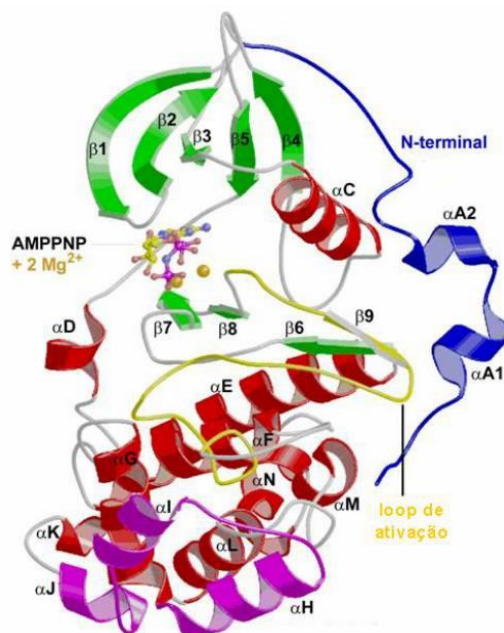
Vale aqui citar sobre o **aducanumab** que foi aprovado em 2021 para o tratamento de DA, pelo *FDA* nos EUA (DHILLON, 2021). Tal fármaco é, na verdade, um anticorpo monoclonal - e não um fármaco tradicional (molécula pequena) - que atua apenas via a remoção de proteínas  $\beta$ -amilóide acumuladas extracelularmente no cérebro; enquanto que, atualmente, é desejável uma ação em ao menos duas vias, ou duas alterações patológicas, para fármacos em DA. Além disso, os dados clínicos disponíveis indicam que tal fármaco é, na melhor das hipóteses, apenas marginalmente eficaz, falha em proteger os pacientes do declínio funcional cognitivo e traz riscos significativos (30 – 40%) - incluindo inchaço cerebral (GUO et al., 2022).

Uma quarta hipótese fisiopatológica para DA se baseia no aparecimento de emaranhados neurofibrilares, constituídos de proteína tau hiperfosforilada. Em condições normais a proteína tau deve estabilizar os microtúbulos, mediando a polimerização de tubulina, entretanto na DA ela perde esta função por se encontrar hiperfosforilada e acabar formando os emaranhados neurofibrilares, levando à desintegração dos microtúbulos (HANGER et al., 1992; MARTINEZ; PEREZ; GIL, 2013; DI MARTINO et al., 2016; RODRIGUES; SILVA, 2017). Esta desintegração ocasiona a disfunção neuronal que está associada aos sintomas de DA. A enzima Glicogênio Sintase Cinase 3-beta (GSK-3 $\beta$ ) surge como um importante alvo biológico para esta hipótese, pois é responsável pela hiperfosforilação de proteínas tau. Desta maneira, a sua inibição deve levar a um controle da fosforilação exagerada das proteínas tau e, conseqüentemente, redução de degeneração de neurônios (PALOMO et al., 2011; FERRERO et al., 2018), conforme será visto a seguir em detalhes.

## 1.2 Glicogênio Sintase Cinase 3-beta (GSK-3 $\beta$ )

A GSK-3 $\beta$  é expressada em uma variedade de tecidos que predominam no cérebro, estando localizada especialmente nos neurônios. Representa uma das isoformas mais abundantes da GSK-3, possui 47 kDa e é composta de 420 aminoácidos. Esta enzima é uma proteína serina-treonina cinase que, em resposta a eventos de sinalização celular, modifica as funções de uma variedade de proteínas por fosforilação de serinas ou treoninas (WOODGETT, 1990; STANDAERT; YOUNG, 2005; PALOMO; MARTINEZ, 2017; DE SIMONE et al., 2021).

A estrutura da GSK-3 $\beta$  apresenta os dois domínios típicos presentes em proteínas cinases: um domínio de folhas- $\beta$  N-terminal (resíduos 35 a 134) e um domínio  $\alpha$ -helicoidal C-terminal (resíduos 139 a 343). O domínio N-terminal consiste de cinco folhas- $\beta$  estendidas ( $\beta$ 1 -  $\beta$ 5) formando uma estrutura na forma de barril, que é interrompida entre as folhas  $\beta$ 3 e  $\beta$ 4 por uma curta  $\alpha$ -hélice ( $\alpha$ C, resíduos 96-102), que empacota na direção contrária ao barril de folhas- $\beta$ . Esta  $\alpha$ -hélice é conservada em todas as proteínas cinases e dois de seus resíduos (Arg96 e Glu97) exercem papel importante na atividade catalítica da enzima. O *loop* de ativação (resíduos 200-226) situa-se ao longo da superfície da fenda de ligação do substrato (HANKS; QUINN; HUNTER, 1988; POIANI, 2017) – vide **Figura 1.4**.



**Figura 1.4** - Representação da estrutura terciária da GSK-3 $\beta$  complexada com AMP-PNP na cavidade 1 de ATP. Folhas- $\beta$  em verde, alfa-hélices em vermelho e o *loop* de ativação em amarelo. Código PDB 1PYX (BERTRAND et al., 2003) [Fonte: (POIANI, 2017)].

O sítio de ligação de ATP está localizado na interface entre os domínios  $\alpha$ -helicoidal e o das folhas- $\beta$  estendidas e está envolvido por um *loop* rico em glicinas e uma estrutura na forma de dobradiça. A GSK-3 $\beta$  possui dois sítios de fosforilação que influenciam a atividade catalítica da proteína: Ser9 localizada no N-terminal que diminui a atividade catalítica da enzima, e Tyr216 localizada no *loop* de ativação, que aumenta a atividade catalítica da enzima. A principal diferença estrutural entre a GSK-3 $\beta$  na sua forma nativa e a fosforilada é a conformação da Tyr216. A fosforilação provoca uma rotação de aproximadamente 120° na cadeia lateral da Tyr216 e reposicionamento das cadeias laterais da Arg220 e Arg223, para formar uma ligação de hidrogênio com o fosfato da Tyr216. Contudo, o principal *loop* de fosforilação mantém-se inalterado (BERTRAND et al., 2003; MÜLLER; KNAPP, 2017).

A atividade catalítica de proteínas cinase depende da correta justaposição dos grupos catalíticos, o que contribui para a transferência do grupo fosfato- $\gamma$  do ATP para a cadeia lateral de um resíduo Ser ou Thr do substrato. Um segundo fator são a acessibilidade e o correto posicionamento dos grupos que constituem o sítio ativo do substrato, os quais proporcionam afinidade e especificidade ao substrato (JOHNSON; NOBLE; OWEN, 1996; BERTRAND et al., 2003).

Em se tratando da regulação/modulação da GSK-3, deve-se notar que existem vários mecanismos complexos em estudo, devido ao seu papel em uma ampla gama

de processos bioquímicos e vias de sinalização (CORMIER; WOODGETT, 2017). Vale lembrar que essa cinase pode ser inibida por diversos substratos endógenos (SUTHERLAND, 2011), como íons bivalentes de zinco (ILOUZ et al., 2002); além de estar envolvida ativa ou passivamente em diferentes vias - por exemplo, Wnt/beta-catenina, TP53, *Notch* e outras (MCCUBREY et al., 2017). Em adição, a GSK-3 pode estar relacionada a mecanismos transcricionais e/ou traducionais, uma vez que essas vias são reguladas por micro-RNAs (MCCUBREY et al., 2017; CANDIDO et al., 2019) e também por ser capaz de modular modificações de histonas (BEUREL; GRIECO; JOPE, 2015), por exemplo.

Vale lembrar, ainda, que como a GSK-3 $\beta$  participa do metabolismo da glicose, inibindo Glicogênio Sintase (GS) e sendo inibida por insulina, a sua inibição pode acarretar hipoglicemia como efeito adverso. Portanto, indica-se a busca e aplicação de inibidores brandos, entre 30 – 40% da atividade da enzima (MARTINEZ; PEREZ; GIL, 2013).

Estudos recentes revelam que o aumento da atividade da GSK-3 $\beta$  compromete vários eventos, como a fosforilação da proteína tau em sítios específicos, fazendo com que a tau perca a afinidade de se ligar aos microtúbulos, favorecendo sua deposição aberrante nos neurônios das regiões afetadas do sistema nervoso, desestabilizando os microtúbulos e levando à progressiva degeneração neuronal. Além disso, a GSK-3 $\beta$  facilita a transformação da APP, resultando no aumento da produção de A $\beta$  e consequentemente favorecendo a formação das placas amilóide (HANGER et al., 1992; DI MARTINO et al., 2016; MEDINA, 2018). Contudo, enfatizamos que esta introdução focará nos aspectos relativos à modulação de GSK-3 $\beta$  frente a inibidores usuais, ou seja, compostos químicos ou moléculas pequenas.

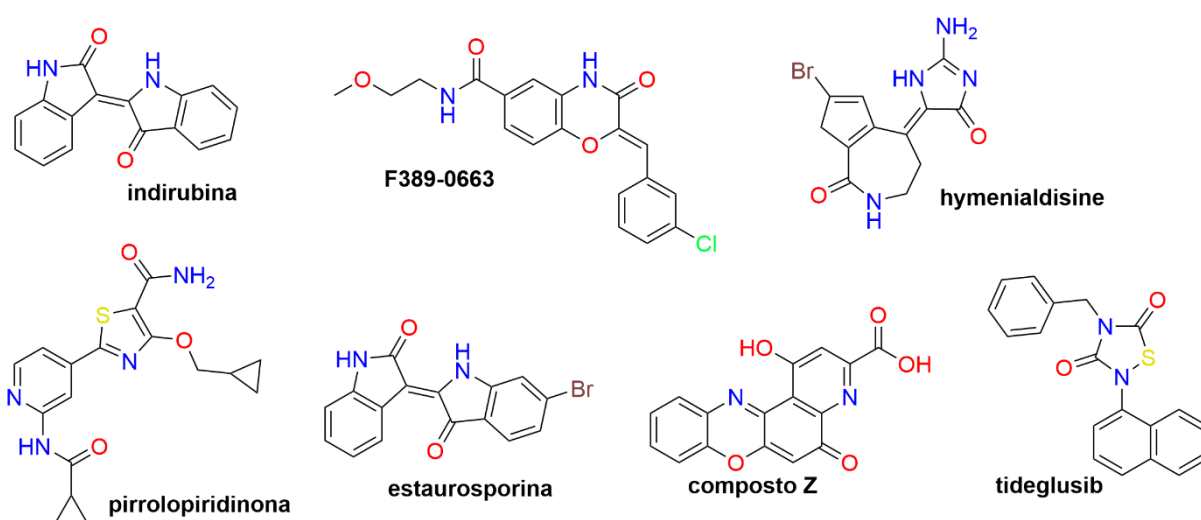
### **1.2.1 Inibidores de GSK-3 $\beta$ com interesse na DA**

Atualmente, não há dúvida de que potenciais inibidores seletivos de GSK-3 $\beta$ , os quais evitem a hiperfosforilação da tau e diminuam a produção de A $\beta$ , podem representar uma abordagem terapêutica efetiva para o tratamento de DA, bem como de demais doenças neurodegenerativas (JANKOWSKA et al., 2018; DE SIMONE et al., 2021).

Um dos primeiros inibidores da GSK-3 $\beta$  a ser reportado na literatura (KLEIN; MELTON, 1996) foi o lítio (Li<sup>+</sup>), que já era utilizado no tratamento de transtorno bipolar.

O  $\text{Li}^+$  é classificado como um inibidor fraco de GSK-3 $\beta$  e seu mecanismo de ação ainda é desconhecido. Entretanto, por atingir diversos processos metabólicos e possuir baixa seletividade, apresenta muitos efeitos adversos (MEDINA, 2018).

Vários grupos de pesquisa têm identificado alternativamente pequenas moléculas orgânicas como potentes inibidores da GSK-3 $\beta$ , tais como derivados da **estaurosporina** ( $\text{IC}_{50} = 3\text{-}7\text{ nM}$ ) (HERS; TAVARÉ; DENTON, 1999), derivados da **indirubina** ( $\text{IC}_{50} = 5\text{-}50\text{ nM}$ ) (LECLERC et al., 2001), **hymenialdisine** ( $\text{IC}_{50} = 10\text{ nM}$ ) (MEIJER et al., 2000), **F389-0663** com  $\text{IC}_{50} = 1,6\text{ }\mu\text{M}$  (JOSHI et al., 2017), **composto Z** com  $\text{IC}_{50} = 2,19\text{ }\mu\text{M}$  (ZHANG et al., 2018) e o *scaffold* da **pirrolopiridinona** com  $\text{IC}_{50} = 0,29\text{ nM}$  (SIVAPRAKASAM et al., 2015) - conforme ilustrado na **Figura 1.5**.



**Figura 1.5** - Representação das estruturas químicas de alguns inibidores da GSK-3 $\beta$ .

Esses inibidores, no entanto, atuam por mecanismo competitivo com o ATP para se ligar ao sítio ativo da GSK-3 $\beta$ . Isto levanta duas principais desvantagens: a necessidade de competir com o ATP para se ligar ao seu sítio ativo, assim como a falta de seletividade destes compostos pela GSK-3 $\beta$ , frente às demais enzimas cinases presentes no sistema biológico (DARSHIT et al., 2014; PALOMO; MARTINEZ, 2017).

Como a GSK-3 $\beta$  está envolvida em vários caminhos de sinalização celular, sugere-se que inibidores competitivos de ATP da enzima apresentam maior chance de gerar efeitos adversos, e inclusive de ser o motivo de baixo sucesso nos testes clínicos dos inibidores já revelados para esta. De modo geral, a busca por maior seletividade se tornou um grande desafio neste contexto e no atual cenário (AVRAHAMI et al., 2013; DI MARTINO et al., 2016; DE SIMONE et al., 2021).

Considerando, então, as vantagens de se buscar inibidores não competitivos de ATP para a GSK-3 $\beta$ , principalmente pela questão da seletividade, existe na literatura alguns trabalhos que comprovaram experimentalmente a inibição desta enzima por compostos que não competem com o ATP. O exemplo mais representativo é o da família de pequenas moléculas heterocíclicas thiadiazolidinones (*TDZD*) (MARTINEZ et al., 2002), onde se descobriu a primeira classe de compostos que inibem a GSK-3 $\beta$  sem competir com ATP. A partir deste estudo, destaca-se o composto **tideglusib** (vide **Figura 1.5**) que é, atualmente, o único composto inibidor da GSK-3 $\beta$  que alcançou fase II de testes clínicos (MEDINA, 2018). Descobriu-se, mais especificamente, que o **tideglusib** inibe a GSK-3 $\beta$  por mecanismo não competitivo com ATP e de modo irreversível (DOMÍNGUEZ et al., 2012).

Outro mecanismo interessante que foi recentemente reportado é a inibição alostérica da GSK-3 $\beta$  (BIDON-CHANAL et al., 2013). Considera-se que inibidores alostéricos da GSK-3 $\beta$  são preferenciais por apresentar maior seletividade, se ligando a regiões únicas na superfície da enzima, bem como por serem úteis na superação da resistência desenvolvida pelo uso de fármacos que competem com ATP (MCINNES; FISCHER, 2005; MÜLLER; KNAPP, 2017). Em adição, estes provocam uma inibição na atividade da enzima de modo mais sutil, do que simplesmente bloqueando a entrada do ATP no sítio ativo (PALOMO et al., 2017).

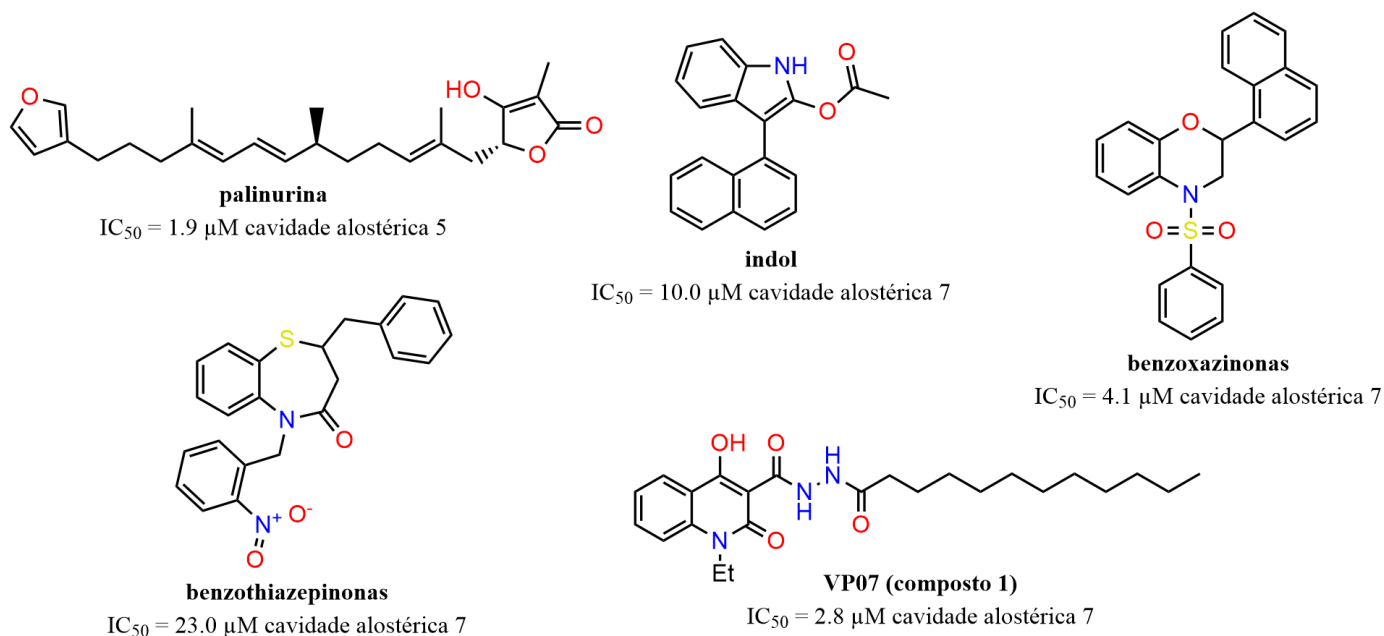
A inibição mais branda da GSK-3 $\beta$  por inibidores alostéricos deve permitir a recuperação do seu equilíbrio homeostático sem interferir em outra sinalização celular (PALOMO et al., 2017). Em complemento, é reconhecido que uma inibição mais branda de determinada proteína-alvo é preferencial do que a sua inibição completa, pois provoca resultados mais favoráveis em tratamentos de doenças crônicas ou de longo prazo, como a DA (AVRAHAMI et al., 2013).

### 1.2.2 Inibidores alostéricos de GSK-3 $\beta$

O alosterismo é reconhecido como uma propriedade inerente de diversas proteínas e enzimas de relevância biológica. De modo geral, é definido como uma interação indireta entre sítios de ligação topograficamente diferentes em uma proteína, mediados por mudança conformacional. Esta mudança conformacional pode ser ocasionada por um composto que se liga em um sítio alostérico.

É um conceito que vem sendo explorado há mais de 50 anos e existem diversos fármacos aprovados e amplamente comercializados que atuam por mecanismos alostéricos nos respectivos alvos. Alguns exemplos: o benzodiazepínico diazepam (Valium®) usado como calmante, e o trametinib (Mekinist®) usado como agente anticancerígeno (MÜLLER; KNAPP, 2017).

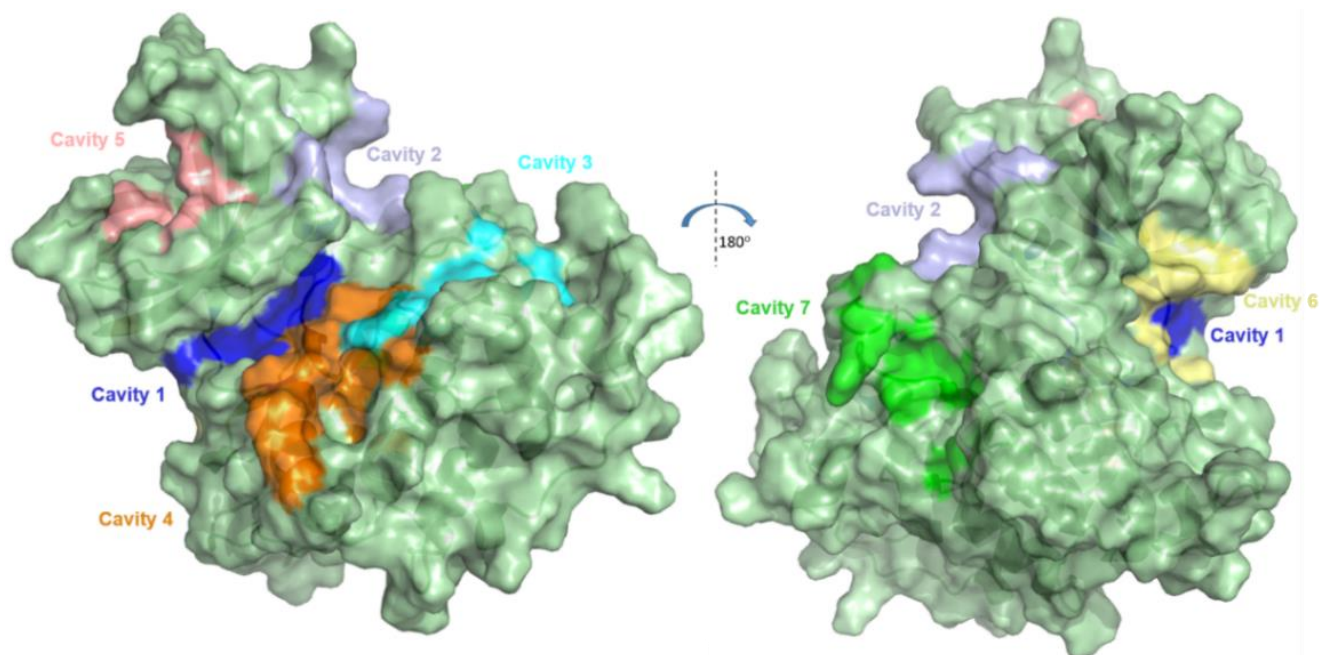
Contudo, para a enzima GSK-3 $\beta$ , poucas informações relacionadas ao conceito de alosterismo são encontradas na literatura (SILVA et al., 2019). Atualmente, poucos inibidores alostéricos da GSK-3 $\beta$  foram revelados, dentre os quais compilamos alguns exemplos representativos das classes conhecidas, conforme apresentado na **Figura 1.6**.



**Figura 1.6** - Representação das estruturas químicas de inibidores alostéricos da GSK-3 $\beta$ , bem como os respectivos valores de IC<sub>50</sub> (em  $\mu$ M) e cavidades.

O primeiro, aqui mencionado, está descrito no trabalho de Bidon-Chanal *et al.* (2013), onde mostra o composto **palinurina** (um produto natural da classe dos sesquiterpenos extraído de esponjas marinhas) como capaz de inibir a GSK-3 $\beta$  pelo mecanismo alostérico. Mais especificamente, foi mostrado que a **palinurina** interage com alta seletividade em uma cavidade alostérica específica, localizada no domínio N-terminal da enzima, gerando uma mudança conformacional do *loop* rico de glicina que restringe o acesso do ATP ao sítio ativo e, conseqüentemente, inibindo brandamente a atividade da GSK-3 $\beta$ . Ainda, correlacionando com outro estudo prévio (PALOMO et al., 2011), foi descoberto que a cavidade de interação para a **palinurina**, representa a cavidade de número 5 na **Figura 1.7**.





**Figura 1.7** - Representação das cavidades 1 – 7 encontradas na estrutura 3D da GSK-3 $\beta$ , sendo: 1 - sítio de ligação de ATP; 2 - sítio de ligação de substrato e, 3 – sítio de ligação de axin/fratide. As cavidades 4, 5, 6 e 7 representam potenciais sítios alostéricos (Fonte: (SILVA et al., 2021b)).

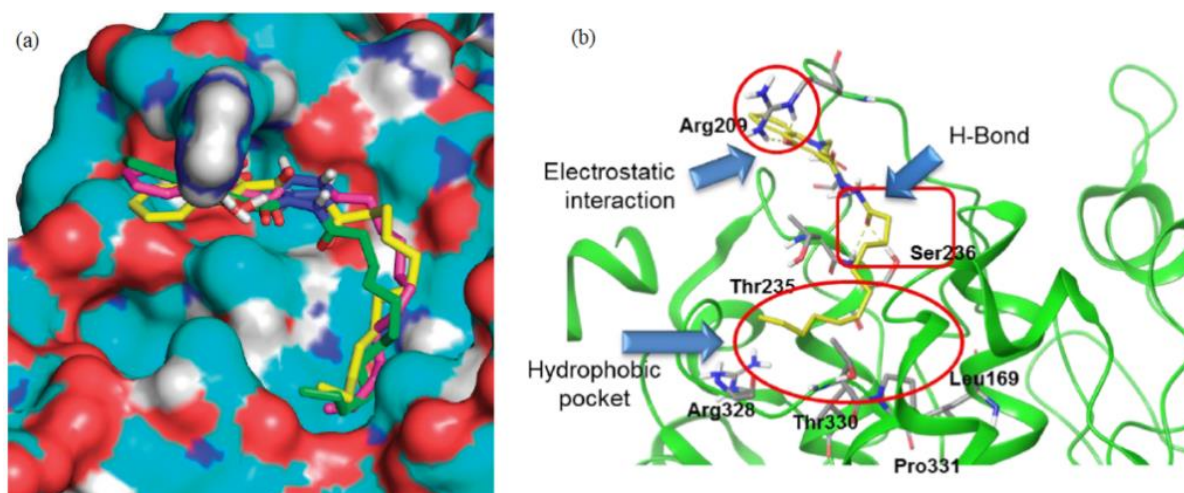
Palomo et al. (2011) realizou um estudo envolvendo modelagem molecular com o objetivo de analisar a estrutura da proteína GSK-3 $\beta$ , localizando precisamente as cavidades em sua superfície que representam potenciais sítios alostéricos. Para isso, utilizou o programa *fpocket* (LE GUILLOUX; SCHMIDTKE; TUFFERY, 2009) para mapear a 25 diferentes estruturas 3D da GSK-3 $\beta$ , assim revelando 7 cavidades em comum (consensuais) a estas, sendo que 4 dessas cavidades representam potenciais sítios alostéricos (**Figura 1.7**).

Em adição, foi verificado que o composto quinolínico (**VP07, composto 1**, vide estrutura na **Figura 1.6**) apresentou inibição frente ao sítio alostérico número 7 da GSK-3 $\beta$  (**Figura 1.7**), sendo capaz de promover uma mudança em seu *loop* de ativação e levando, assim, a diminuição da atividade da enzima.

Mais recentemente, outro trabalho (PALOMO et al., 2017) foi reportado detalhando que o **composto 1** interage com resíduos de aminoácidos específicos presentes na cavidade alostérica da GSK-3 $\beta$ , conforme mostrado na **Figura 1.8**. Seis principais aminoácidos neste sítio alostérico são destacados, a saber: Arg209, Ser236, Thr235, Arg328, Thr330, Pro331 e Leu169.

Adicionalmente, outros compostos análogos do **composto 1** foram sintetizados e testados frente a GSK-3 $\beta$ , revelando-se algumas características estruturais imprescindíveis para que ocorra inibição alostérica da enzima, na cavidade de

interesse. Pode-se citar, por exemplo, a importância da longa cadeia alifática de **1** que realiza interação hidrofóbica com a região hidrofóbica da cavidade, que é representada pelos aminoácidos Leu169, Arg328, Thr330 e Pro331 (vide **Figura 1.8**).



**Figura 1.8** - (a) Representação do **VP07** acomodado na cavidade 7, sítio alostérico, da GSK-3 $\beta$  (estrutura *PDB* código 1PYX) [Fonte: (PALOMO et al., 2011)]; (b) Interações entre o **VP07** e resíduos de aminoácidos presentes na cavidade da GSK-3 $\beta$  [Fonte: (PALOMO et al., 2017)].

Demais estudos também revelaram outras poucas classes de inibidores alostéricos de GSK-3 $\beta$  possivelmente se ligando a cavidade 7, e.g. classe das **benzoxazinonas** (BROGI et al., 2017), do **indol** (BROGI et al., 2017) e das **benzothiazepinonas** (GAO et al., 2018). Estas classes estão ilustradas por exemplos representativos de correspondentes estruturas na **Figura 1.6**.

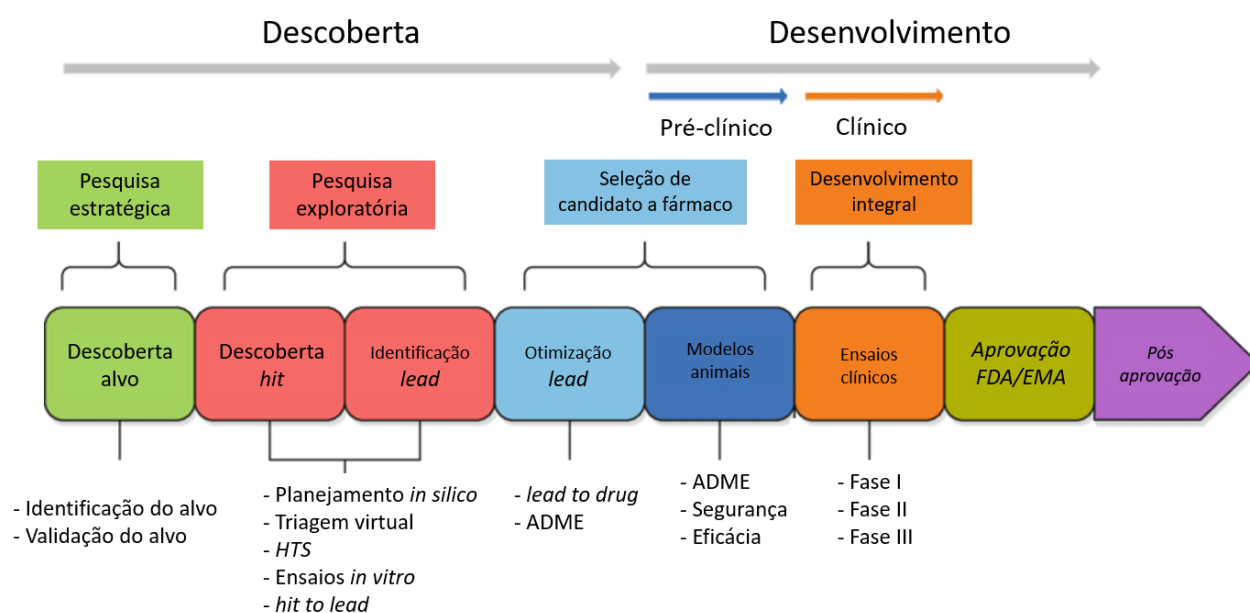
Com base nestas informações surge a possibilidade de se estudar mais a fundo possíveis interações entre outros compostos e esta cavidade da GSK-3 $\beta$ , bem como dos demais potenciais sítios alostéricos mencionados. Em especial, pode-se expandir estas investigações com o uso de estratégias computacionais.

Especula-se, por exemplo, a aplicação de métodos baseados em estrutura considerando-se os aminoácidos presentes nestes sítios, ou ainda, com o uso de métodos baseados em ligante considerando-se as características estruturais dos (poucos) inibidores alostéricos da enzima previamente reportados. Além disso, até o momento, não foi publicado nenhum trabalho envolvendo o planejamento racional de potenciais candidatos a fármacos em DA, que consiste na aplicação de triagem virtual (*Virtual Screening, VS*) para buscar novos compostos, ou novas classes de compostos, que se apresentem como potenciais inibidores alostéricos da GSK-3 $\beta$ .

### 1.3 Planejamento computacional de fármacos

A descoberta e o desenvolvimento de fármacos é um processo longo e complexo, principalmente no que diz respeito ao gasto de dinheiro e tempo. Estima-se um custo total na faixa de 2 – 3 bilhões de dólares e um tempo de 13 – 15 anos para se fazer um fármaco chegar à prateleira do mercado, ou seja, considerando-se todo o processo desde as etapas iniciais da sua descoberta como uma molécula até as demais etapas de sua aprovação (**Figura 1.9**) (ABRAHAM, 2005; KANNT; WIELAND, 2016; ZHANG et al., 2022).

Além do mais, com os avanços tecnológico e com o progresso na aquisição de dados, há um número cada vez maior de possíveis estruturas moleculares para candidatos a fármacos, assim como de possíveis alvos biológicos a serem avaliados. Ficando, assim, inviável trabalhar em projetos de pesquisa utilizando somente técnicas experimentais (TROPSha; GOLBRAIKH, 2007; CHERKASOV et al., 2014; BASITH et al., 2018).



**Figura 1.9** - Esquema ilustrativo do processo de descoberta e desenvolvimento de fármacos [Fonte: adaptado de (DUELEN et al., 2019)].

Visando melhorar este cenário, o uso da química computacional vem se tornando, a cada dia, mais indispensável no desenvolvimento de fármacos, pois reduz consideravelmente o custo e o tempo gastos em pesquisas tanto acadêmicas-universitárias como na indústria farmacêutica (KUBINYI, 2007; BROWN et al., 2017; CHANG et al., 2022). Pode-se dizer, portanto, que o planejamento computacional (ou

*in silico*) de fármacos é um processo que compreende o uso de diversas estratégias metodológicas *in silico* para identificar novas moléculas como potenciais candidatos a fármacos. Existem diversos autores que enfatizam a importância deste processo, bem como inúmeros casos de sucesso na literatura (PEITSCH, 2004; ABRAHAM, 2005; KELLICI et al., 2015; KANNT; WIELAND, 2016; KAPLAN et al., 2022; VEMULA et al., 2023).

Vale observar que, diante do mesmo objetivo em comum, diferentes termos, categorias e/ou designações de áreas do conhecimento correlacionam-se e são usados para se referir aos estudos envolvendo o 'Planejamento Computacional de Fármacos' ou o "Planejamento Racional de Fármacos" - que, na verdade, vem do inglês *Computer-Aided Drug Design (CADD)* e *Rational Drug Design*, respectivamente. Pode-se dizer que (de modo geral e aqui no Brasil) tais estudos estão contidos na área de Química Medicinal Computacional, a qual paralelamente contempla ou está relacionada a outras áreas/subáreas correlatas: Química Computacional como uma subárea da Química que emprega métodos teóricos/semiempíricos para se calcular computacionalmente propriedades estruturais de moléculas; Quiminformática como o uso de técnicas computacionais para explorar bases de dados químicas e descobrir compostos com atividade biológica ou propriedades físico-químicas específicas; Modelagem Molecular como a utilização de programas gráficos 3D para modelar ou mimetizar o comportamento de moléculas, em vista de suas naturezas e mecanismos frente a proteínas (NOGUEIRA; MONTANARI; DONNICI, 2009; BARREIRO; FRAGA, 2015; ALVES et al., 2017; DO AMARAL et al., 2017).

Existem outras subdesignações, mas estas mencionadas acima já nos permite inferir, portanto, que os os estudos *CADD* englobam as tarefas de identificar e selecionar *hits* ou *leads*, otimizar a estrutura molecular de *leads* (com respeito à afinidade, especificidade, eficácia e efeitos colaterais), predizer a biodisponibilidade de ligantes, obter compostos candidatos a testes clínicos, bem como estabelecer relações entre a estrutura química e atividade biológica (ABRAHAM, 2005; BROWN et al., 2017). Em complemento a esta última menção, pode-se, por exemplo, predizer a eficácia terapêutica e também elucidar detalhes do mecanismo de ação de determinado ligante, por meio de *softwares* que simulam interações ligante-receptor (CHERKASOV et al., 2014; CHANG et al., 2022).

De modo geral, as principais estratégias metodológicas envolvidas em *CADD* podem se dividir em duas: baseadas em ligante (do inglês *Ligand-Based Drug Design*, *LBDD*) e baseadas em estrutura (do inglês *Structure-Based Drug Design*, *SBDD*). Como ambas também podem ser utilizadas e incorporadas em campanhas de prospecção, abordaremos, em antemão, alguns aspectos sobre triagem virtual.

### 1.3.1 Triagem virtual

A seleção de compostos se tornou uma necessidade no planejamento racional de fármacos, antes que estes sejam submetidos a ensaios biológicos e técnicas experimentais, principalmente considerando-se o grande número de compostos disponíveis e os altos custos das etapas de desenvolvimento de um novo fármaco (BLEICHER et al., 2003; SCIOR et al., 2012). Esta seleção de compostos passou a ser feita de modo virtual e aplicando-se filtros de diferentes naturezas. Em outras palavras, o objetivo principal da triagem virtual (*VS*) é reduzir um grande número de compostos de uma base de dados a um número menor de compostos, que apresentem potencial atividade frente à proteína-alvo em estudo e, assim, tendo maior chance destes se tornarem candidatos a fármacos (KUBINYI, 2007; PICCIRILLO; DO AMARAL, 2018).

Geralmente, com relação aos filtros usados em uma campanha trivial de *VS*, é conveniente utilizar métodos com menor custo computacional nas primeiras etapas e métodos que requerem maior capacidade de processamento nas etapas finais. Ao final da triagem, após a seleção dos *hits*, estes podem ser obtidos fisicamente (comercialmente ou por síntese orgânica) e ter suas atividades confirmadas/validadas por ensaios biológicos. Quando convir, os *hits* selecionados podem ser submetidos a estudos de otimização molecular (e.g. *hit to lead*) visando o aumento de potência e de propriedades (FEDERICO et al., 2021a).

Cabe aqui reforçar as definições, de acordo com a *IUPAC* (BUCKLE et al., 2013): *hit* é um composto que produz atividade/efeito biológico reproduzível, acima de um valor de corte definido para determinado ensaio biológico, e cuja identidade estrutural foi estabelecida; *lead* é um composto que satisfaz critérios mínimos predefinidos para prosseguir para etapas de otimização de estrutura e atividade.

Estudos de *CADD* envolvendo campanhas de *VS* são geralmente realizados com a expectativa de se descobrir ou identificar *hits* altamente potentes (e.g.

composto bioativo com valor de  $IC_{50}$  na escala nM) (BLEICHER et al., 2003). Entretanto, é importante lembrar que também é de grande valor alcançar, por exemplo, diversidade química dos *hits* (LIONTA et al., 2014; GORGULLA et al., 2022). Isto é particularmente interessante quando se trabalha com um alvo novo/específico, pois assim também se amplia a respectiva biblioteca de inibidores bem como de dados sobre as respectivas relações estrutura-atividade. Para tanto, diversos fatores em VS podem ser levados em conta, dos quais enfatizaremos dois.

Em *primeiro lugar*, a escolha de bases de dados é fundamental para iniciar uma campanha VS (BLEICHER et al., 2003; WARR et al., 2022), especialmente se considerarmos o tamanho estimado do espaço químico de  $10^{60}$  moléculas orgânicas (BOHACEK; MCMARTIN; GUIDA, 1996). Neste contexto, surge um paradigma com relação ao tamanho das bases: há os que sugerem o uso de bibliotecas maiores (ultra grandes) para aumentar as taxas de *hits* em VS (GLORIAM, 2019; GORGULLA et al., 2020), enquanto outros sugerem o uso de bibliotecas modestas como o suficiente para a obtenção de *hits* com ‘estruturas belas’ (CLARK, 2020).

Sendo assim, em geral, sugere-se fazer um balanceamento para a escolha das bases, considerando também quais objetivos são almejados e, em *segundo lugar*, quais metodologias serão escolhidas para aplicação na campanha de VS.

Enfatizamos, desta forma, que há de se objetivar campanhas de VS para além de somente a obtenção de *hits* potentes, mas que também contemple a identificação de *hits* com diversidade estrutural. Além disso, junto ao uso de diferentes/amplas/versáteis bibliotecas químicas, também é de grande interesse explorar o uso de metodologias distintas e em consenso para tal finalidade (DOS SANTOS MAIA et al., 2020; VÁZQUEZ et al., 2020). Isso, pois, assim se torna viável rumar a proposição de novas abordagens metodológicas que possam ser aproveitadas em outros contextos, que também carecem de campanhas de VS mais eficientes.

### 1.3.2 Planejamento baseado em ligante

O planejamento computacional de fármacos via metodologias baseadas em ligantes, *LBDD*, é usado quando não se conhece a estrutura da proteína-alvo ou receptor (HAWKINS; SKILLMAN; NICHOLLS, 2007; HÖNIG; LEMMEN; RAREY, 2022). É possível e recomendável agregar o uso de abordagens metodológicas de

*LBDD* em campanhas de triagem virtual, por isso, comumente, utiliza-se também o termo *Ligand-Based Virtual Screening (LBVS)* (ECKERT; BAJORATH, 2007; HIMMAT et al., 2016).

Estudos *LBVS* se baseiam no uso de informações químicas e estruturais de um (ou vários) ligante(s) com atividade biológica conhecida, para se buscar ou prever novas estruturas moleculares que apresentem características físico-químicas/estruturais similares com tais ligantes de referência (KUBINYI, 2007; NAYARISSERI et al., 2021).

Em geral, os então denominados ‘estudos baseados em similaridade’ se fundamentam no princípio de que moléculas semelhantes apresentam atividades biológicas semelhantes (KLOPMAND, 1992), podendo ser subclassificados em métodos 2D ou 3D (FEDERICO et al., 2021a). Dentre as mais diversas metodologias *LBDD*, serão aqui abordados alguns aspectos sobre similaridade 2D e 3D, bem como sobre *QSAR*.

### **1.3.2.1 Similaridade 2D**

As medidas de similaridade estrutural se dão em três principais etapas: a representação estrutural da molécula, o esquema de ponderação para determinar o valor da importância relativa da representação estrutural e o coeficiente de similaridade (BERO et al., 2017; KUMAR; ZHANG, 2018).

Nas metodologias de similaridade 2D, as moléculas são descritas na forma de um conjunto de caracteres denominado *fingerprint* molecular (i.e., “impressão digital” molecular), que torna possível a análise computacional da complexa e subjetiva comparação por similaridade (MAGGIORA et al., 2014; ALVES et al., 2017).

Os *fingerprints* consistem de representações da molécula na forma de *bits* (valores binários de 0 e 1), de acordo com a ausência (0) ou presença (1) de determinados fragmentos/subestruturas/grupos. Estes, por sua vez, podem ser analisados computacionalmente de maneira objetiva (WILLETT, 2006; CERETO-MASSAGUÉ et al., 2015). Existem diversos métodos para geração de *fingerprints* 2D, dentre os quais destacam-se: subestrutura-chave, e.g., *MACCS* (DURANT et al., 2002); topológicos, e.g., *The Daylight fingerprint* (DAYLIGHT, 2023); circulares, e.g., *ECFPs* ou *Morgan* (MORGAN, 1965).

Após as etapas de representação estrutural e ponderação, a análise comparativa das similaridades estruturais deve ser feita por meio de um método matemático que viabilize a quantificação das mesmas. Mais comumente empregado é o coeficiente de Tanimoto (TC) (ROGERS; TANIMOTO, 1960), definido pela seguinte equação:

$$TC = \frac{c}{a + b - c}$$

Equação 1.1

Em que, dada duas moléculas A e B, 'a' corresponde ao número de características presentes na molécula A (número de *bits* iguais a 1 em A), 'b' corresponde ao número de características presentes na molécula B (número de *bits* iguais a 1 em B), e 'c' corresponde ao número de características compartilhadas pelas duas moléculas (número de *bits* iguais a 1 em ambas moléculas A e B). Dessa forma, o TC tem seu valor calculado entre 0 e 1, sendo que 0 indica uma estrutura que não apresenta nenhuma similaridade com a molécula de referência e, quanto mais próximo de 1, maior é a similaridade entre a molécula analisada e a molécula de referência (WILLETT, 2006; BERO et al., 2017).

Para além do conhecimento clássico sobre similaridade 2D, citam-se metodologias sofisticadas, como por exemplo, o *software FTrees* (BOEHM et al., 2008) que faz buscas por similaridade com base nas características farmacofóricas e na topologia global das moléculas. Para tanto, é feito um alinhamento ou sobreposição do *scaffold* das moléculas, permitindo o reconhecimento de novas estruturas, pertencentes a classes químicas diferentes, mas com propriedades parecidas. Em especial, o **FTrees** é ultra rápido, sendo capaz de fazer buscas em bases de dados (espaços químicos) com bilhões de moléculas em apenas alguns segundos.

Adicionalmente, esforços vêm sendo feitos para aprimorar a efetividade e busca por similaridade em *LBVS*, considerando-se a necessidade de se livrar de atributos irrelevantes (ou com menor importância/peso) que constituem os *fingerprints* 2D - que usualmente levam a altas taxas de insucesso quando se utiliza um grande volume de dados. Neste sentido, algoritmos bastante sofisticados têm sido usados, e.g., algoritmo genético (BERRHAIL; BELHADEF, 2020), rede neural convolucional (ALTALIB; SALIM, 2022), inteligência artificial e aprendizagem de máquina (NAYARISSERI et al., 2021), *deep learning* (NASSER et al., 2022), etc.



Contudo, a similaridade 2D, apesar de ser uma metodologia fundamentada em conceitos clássicos simples, ainda é bastante visada e aplicada, diante, principalmente, do surgimento de tais métodos mais robustos. A baixa necessidade de processamento computacional para realização de triagens, bem como o rápido desenvolvimento das mesmas, faz com que estas metodologias persistam e predominem nos estudos *CADD* (FEDERICO et al., 2021a).

### 1.3.2.2 Similaridade 3D

Os estudos envolvendo similaridade 3D surgiram diante do aumento da quantidade de informações sobre as estruturas das moléculas, ausentes nos métodos 2D. Estes consideram a distribuição espacial das propriedades e/ou a estrutura 3D das moléculas, ou seja, levam em conta a conformação/orientação e o volume/forma que as moléculas ocupam no espaço (KORTAGERE; KRASOWSKI; EKINS, 2009; FAN; HIRTE; KIRCHMAIR, 2022).

Assim, em analogia ao conceito de 'pose supostamente bioativa' (KUBINYI, 2007), se baseiam no pressuposto de que moléculas que adotam uma forma (*shape*) similar à uma molécula de referência (com atividade conhecida frente a um sítio de ligação de dado receptor) possuem grandes chances de similarmente se ligar no mesmo receptor (KOSHLAND, 1958; KUMAR; ZHANG, 2018).

Várias abordagens envolvendo modelagem 3D são amplamente utilizadas a partir deste conceito. Como exemplo, citam-se: *QSAR-3D*, modelos farmacofóricos, similaridade de forma e eletrostática, dentre outras. Existem subclassificações a partir destas abordagens, assim como inúmeros *softwares* que as empregam.

Com relação a sobreposição em similaridade 3D, por um lado, existem aqueles que não necessitam da superposição/sobreposição/alinhamento entre as moléculas, e.g. as que empregam métodos baseados em distância atômica. Estas utilizam de descritores geométricos, e assim, descrevem a forma de uma molécula por meio da posição relativa dos átomos que a compõe (BALLESTER; FINN; RICHARDS, 2009). Por outro lado, existem abordagens de similaridade 3D que dependem de uma sobreposição adequada entre moléculas, para efetuar uma aproximação à geometria bioativa do ligante referência (NICHOLLS et al., 2010).

Conforme revisado recentemente por Hönig et al. (2022), os estudos de similaridade 3D via sobreposição das estruturas moleculares podem ser divididos nas

seguintes subcategorias: baseadas em forma ou volume; baseada em campo; baseada em gráficos; e baseados em distância. A título de exemplo, a **Tabela 1.1** mostra alguns *softwares* de tais subdivisões de metodologias.

Na seção 3.2.1.2, detalharemos mais sobre os métodos dos *softwares* **ROCS** e **EON** que empregam similaridade 3D via sobreposição de forma e de potencial eletrostático, respectivamente.

**Tabela 1.1** -. Metodologias e algoritmos de similaridade 3D via sobreposição das estruturas moleculares e os respectivos exemplos de *softwares* [Fonte: adaptado de (HÖNIG; LEMMEN; RAREY, 2022)].

Metodologia/algoritmo	Software	Ano e referências
Sobreposição de volume ou forma	<b>ROCS</b>	1995 <sup>(a), (b)</sup>
	<b>FlexS</b>	1997 <sup>(c)</sup>
	<b>MOE</b>	2001 <sup>(d)</sup>
	<b>MolAlign</b>	2017 <sup>(e)</sup>
	<b>eSim</b>	2019 <sup>(f)</sup>
Baseado em campo	<b>APF</b>	2008 <sup>(g)</sup>
	<b>FLAP</b>	2007 <sup>(h)</sup>
Baseado em distância	<b>Catalyst</b>	1996 <sup>(i)</sup>
	<b>GALAHAD</b>	2006 <sup>(j)</sup>
	<b>LS-align</b>	2018 <sup>(k)</sup>
	<b>InterLig</b>	2020 <sup>(l)</sup>
Baseado em gráficos	<b>Disco</b>	1993 <sup>(m)</sup>
	<b>LigandScout</b>	2005 <sup>(n)</sup>
	<b>PharmaGist</b>	2008 <sup>(o)</sup>
	<b>FKCOMBU</b>	2014 <sup>(p)</sup>

<sup>(a)</sup> (GRANT; GALLARDO; PICKUP, 1996). <sup>(b)</sup> (HAWKINS; SKILLMAN; NICHOLLS, 2007). <sup>(c)</sup> (LEMMEN; LENGAUER; KLEBE, 1998). <sup>(d)</sup> (LABUTE et al., 2001). <sup>(e)</sup> (CHAN, 2017). <sup>(f)</sup> (CLEVES; JOHNSON; JAIN, 2019). <sup>(g)</sup> (TOTROV, 2007). <sup>(h)</sup> (BARONI et al., 2007). <sup>(i)</sup> (BARNUM et al., 1996). <sup>(j)</sup> (RICHMOND et al., 2006). <sup>(k)</sup> (HU et al., 2018). <sup>(l)</sup> (MIRABELLO; WALLNER, 2020). <sup>(m)</sup> (MARTIN et al., 1993). <sup>(n)</sup> (WOLBER; LANGER, 2005). <sup>(o)</sup> (SCHNEIDMAN-DUHOVNY et al., 2008). <sup>(p)</sup> (KAWABATA; NAKAMURA, 2014).

### 1.3.2.3 QSAR

O efeito que uma substância química exerce sobre um sistema biológico pode ser elucidado pela análise de suas interações e pela inspeção de suas relações representadas pelas respectivas propriedades físico-químicas/estruturais e biológicas. Ao estabelecer tais relações de maneira quantitativa, esta pode ser denominada Relações Quantitativas entre Estrutura Química e Atividade Biológica (do inglês *Quantitative Structure-Activity Relationship*, QSAR) (ANDREWS; CRAIK; MARTIN, 1984; KUBINYI, 1993; WORTH, ANDREW; FUART-GATNIK, MOJCA;

LAPENNA, SILVIA; SERAFIMOVA, 2011; VUKOVIC; GADALETA; BENFENATI, 2019).

Os estudos de QSAR estendem-se a diferentes campos de aplicação, incluindo o desenvolvimento/descoberta/otimização de fármacos, a predição do potencial de toxicidade de compostos e também a avaliação dos mecanismos de ação de fármacos e agroquímicos (HANSCH; LEO, 1995; RUSYN et al., 2012; BAUMANN; BAUMANN, 2014). Além disso, o QSAR se apresenta hoje como uma técnica robusta, de enorme interesse da academia e da indústria farmacêutica (CRAMER, 2012; DE et al., 2022).

A publicação pioneira de QSAR ocorreu em 1964 por Hansch e Fujita (C. HANSCH, 1964), cuja proposição ficou conhecida como abordagem extratermodinâmica, consistindo de um modelo matemático-estatístico correlacionando a atividade biológica de uma série de moléculas em função de suas propriedades físico-químicas/estruturais. As Equações 1.2 e 1.3 exemplificam de modo genérico e de modo ilustrativo, respectivamente, de como a abordagem extratermodinâmica pode ser compreendida:

$$\textit{Atividade biológica} = f(\textit{descritores moleculares})$$

Equação 1.2

$$\log(1/C) = k_1 \log P - k_2 \sigma + k_3 Es + k_4$$

Equação 1.3

Em que, **C** é a concentração (dose) molar de composto que produz determinada atividade/resposta biológica; **logP**, **σ**, **Es** são parâmetros físico-químicos que expressam as propriedades hidrofóbicas, eletrônicas e estéricas, respectivamente; **k<sub>1</sub>**, **k<sub>2</sub>**, **k<sub>3</sub>** são os coeficientes que expressam as contribuições relativas de cada parâmetro, e **k<sub>4</sub>** é uma constante.

A abordagem extratermodinâmica clássica foi baseada na extensão de conceitos de química orgânica (físico-química orgânica) e nos princípios de relações lineares de energia livre (HAMMETT, 1937), que permite a análise de possíveis relações entre propriedades cinéticas (expressas por parâmetros biológicos, e.g., IC<sub>50</sub>) e termodinâmicas (expressas por parâmetros obtidos experimentalmente ou calculados) (KUBINYI, 1993; GERTZEN; GOHLKE, 2012).

Além do mais, tal abordagem era limitada ao uso de pequenas séries congêneres de compostos análogos, com pequenas variações estruturais de grupos

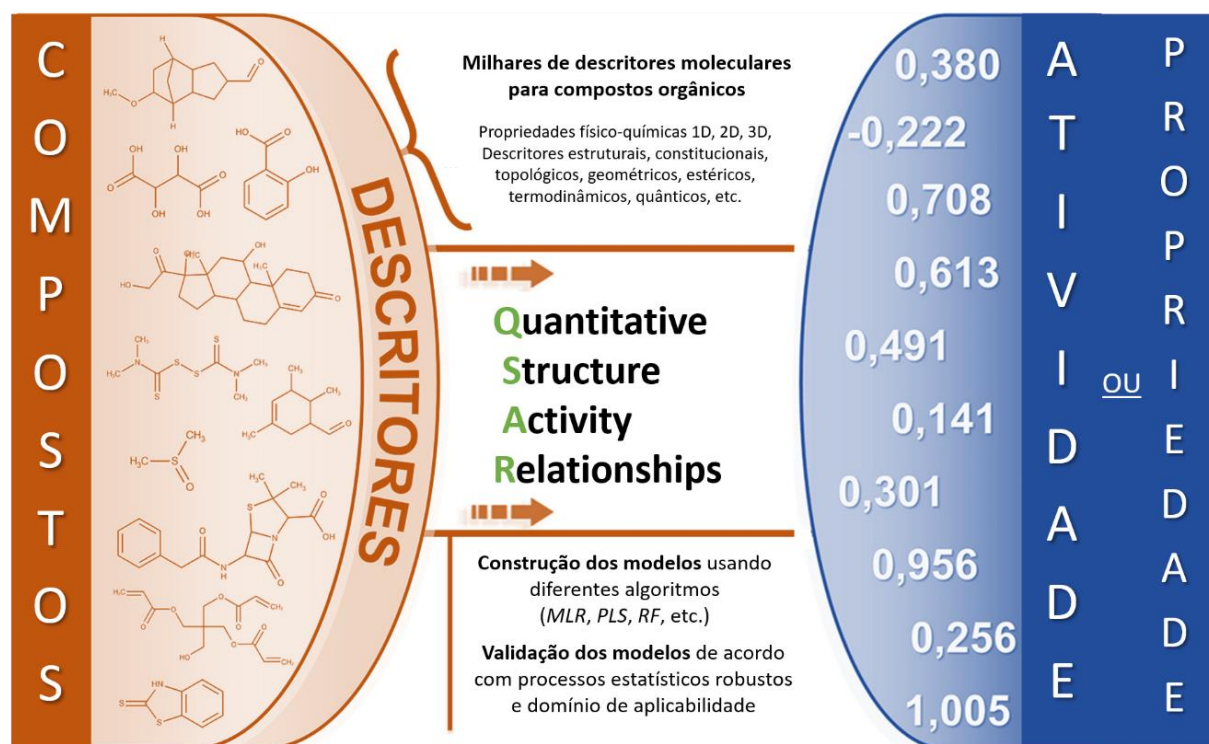
substituintes ou fragmentos – podendo também ser expressas pela variável indicadora de Free-Wilson (FREE; WILSON, 1964). Ainda, ao uso de poucos descritores químicos, i.e., parâmetros físico-químicos/estruturais medidos ou calculados, como  $\log P$ ,  $\sigma$ ,  $\pi$ , etc. Assim, a avaliação das contribuições relativas de cada parâmetro para a atividade biológica só poderia ser estabelecida por modelos matemáticos simples e lineares (HANSCH et al., 1962; WILLETT, 2011; ALEXANDER; TROPSHA; WINKLER, 2015). Mas logo surgiu a proposição de relações mais rebuscadas por meio de modelos não lineares, como a parabólica (C. HANSCH, 1964) e a bilinear (KUBINYI, 1993).

Décadas correram e os algoritmos e as ferramentas estatísticas para propor modelos de QSAR evoluíram e se sofisticaram, juntamente com a quantidade de dados a serem manipulados. Citam-se aqui alguns exemplos: regressão linear múltipla (*MLR, Multiple Linear Regression*) (ENSLEIN; CRAIG, 1978; MARTIN; LILAVOIS; BARRON, 2017), mínimos quadrados ordinários (*OLS, Ordinary Least Squares*), análise e regressão de componente principal (*PCA e PCR, Principal Component Analysis e Regression*), análise discriminante (*DAn, Discriminant Analysis*), regressão parcial de mínimos quadrados (*PLS, Partial Least-Square*) (SAZONOVAS; JAPERTAS; DIDZIAPETRIS, 2010), *k-NN (k nearest neighbor)*, algoritmo genético (*Genetic Algorithms, GA*), redes neurais artificiais (*Artificial Neural Networks, ANN*) (HAMADACHE et al., 2016), *RF (Random Forest)* (LEI et al., 2016) e assim por diante (ROY; KAR; DAS, 2015; SLIWOSKI; MENDENHALL; MEILER, 2016).

No que diz respeito aos descritores químicos, atualmente, existem índices topológicos, estéricos, eletrostáticos e também índices que codificam informações geométricas 3D sobre a molécula. Estes podem ser calculados fácil e rapidamente (TODESCHINI; CONSONNI, 2000; TETKO; TANCHUK, 2002; SCHWAB, 2010; YAP, 2011). Existe uma grande variedade destes devido ao crescente interesse em identificar formas adicionais de descrever as estruturas de moléculas, em diversas perspectivas dimensionais, bem como à enorme expansão de dados gerados nas últimas décadas (TROPSHA; GOLBRAIKH, 2007; KAUSAR; FALCAO, 2018).

Aproveitamos aqui para citar referências que tratam do conteúdo histórico e da trajetória do QSAR (KUBINYI, 2007; HANSCH, 2011; CRAMER, 2012; GERTZEN; GOHLKE, 2012; CHERKASOV et al., 2014; GRAMATICA, 2020; MURATOV et al., 2020; TOROPOV; TOROPOVA, 2020).

Trazendo para os tempos de hoje e buscando simplificar: para se desenvolver um modelo de QSAR são necessários descritores químicos e a atividade biológica, conforme mostra a **Figura 1.10**. De modo análogo, para se desenvolver os denominados QSPR (*Quantitative Structure-Property Relationships*), substitui-se o parâmetro biológico, que expressa atividade biológica definida, por determinado parâmetro que então expresse a correspondente propriedade definida.



**Figura 1.10** - Esquema ilustrativo representando a geração de modelos de QSAR (ou QSPR) [Fonte: adaptado de (TROPSHA, 2010)].

Em se tratando dos métodos de validação de modelos QSAR, existem muitas diretrizes e critérios disponíveis na literatura (GRAMATICA, 2007; OECD, 2007; DEARDEN; CRONIN; KAISER, 2009; TROPSHA, 2010; CHERKASOV et al., 2014; ROY; KAR; DAS, 2015). Segundo (GRAMATICA, 2020), um modelo de QSAR robusto e confiável, além de útil, deve ser o adequado para o objetivo proposto. Demais detalhes sobre estes métodos de validação são apresentados na seção 3.2.2.2 desta Tese.

Ao longo do tempo, os estudos de QSAR mostraram que além de interações específicas entre o ligante e o receptor, poderia haver influência de efeitos de campo por interações intermoleculares não covalentes na vizinhança espacial dos ligantes (CHERKASOV et al., 2014). Isso levou ao aprofundamento de estudos relacionados a descritores moleculares tridimensionais e, portanto, ao surgimento do QSAR-3D

(CRAMER; PATTERSON; BUNCE, 1988), que incorpora informações tridimensionais sobre ligantes, requerindo alinhamento e escolha conformacional dos mesmos (DOWEYKO, 2004).

Vale ainda ressaltar que demais progressos no campo de modelagem de QSAR culminaram no desenvolvimento de outras abordagens, tais como: *HQSAR* (LOWIS, 1997), *4D-QSAR* (VEDANI et al., 2000), *5D-QSAR* (VEDANI; DOBLER, 2002), *QSAR multi-target* e multiobjetivo (PRADO-PRADO et al., 2009; LAMBRINIDIS; TSANTILIKAKOULIDOU, 2018), *aiQSAR* (VUKOVIC; GADALETA; BENFENATI, 2019), dentre outros.

### 1.3.3 Planejamento baseado em estrutura

O planejamento computacional de fármacos baseado em estrutura, *SBDD*, é uma metodologia muito eficiente que se baseia no uso da estrutura do receptor como alvo biológico. A informação estrutural do receptor/proteína permite o entendimento de suas possíveis interações com ligantes, ou seja, da formação do complexo ligante-receptor (ABRAHAM, 2005; DUTTA; BOSE, 2021).

Estudos *SBDD* estão diretamente associados aos avanços das técnicas de RMN, cristalografia de raios X de proteínas e, mais recentemente, de microscopia eletrônica criogênica (*CryoEM*), que permitiram a identificação e a caracterização da estrutura 3D de inúmeras proteínas-alvo de interesse terapêutico (ANDERSON, 2003; APLIN et al., 2022). Estruturas 3D de proteínas podem ser encontradas em diversas bases de dados, das quais há maior destaque para o *Protein Data Bank* (*PDB*, <https://www.rcsb.org/>).

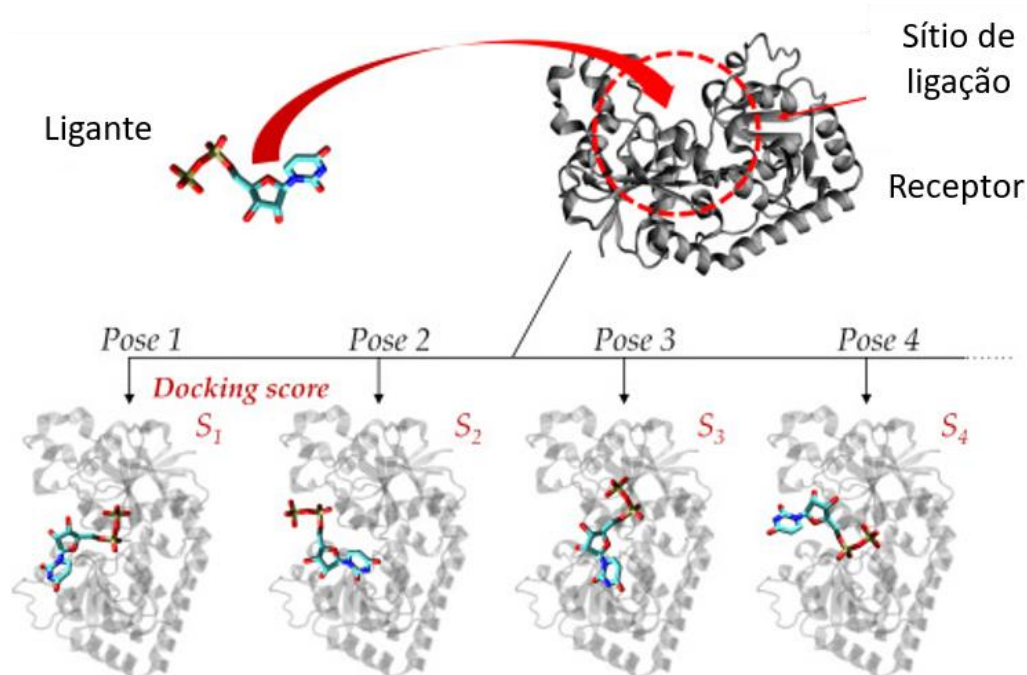
O fato de lidar com a doença em nível molecular associado a uma estrutura 3D da proteína-alvo, atribui ao *SBDD* um caráter de maior especificidade, eficiência e rapidez na descoberta e otimização de fármacos (LIONTA et al., 2014). De acordo com o trabalho de revisão de Dutta et al. (2021), o *SBDD* desempenhou um papel crucial no desenvolvimento de uma série de medicamentos que foram lançados com sucesso no mercado, e.g., antiretroviral ritonavir, inibidor de protease de HIV saquinavir, anticancerígeno raltitrexede, tofacitinib para tratar artrite reumatóide, etc.

Dentre as abordagens *SBDD* mais comuns, pode-se citar o *docking* e a dinâmica molecular. Como ambos são comumente utilizados como filtros em triagens

virtuais, utiliza-se também o termo *Structure-Based Virtual Screening (SBVS)*. Exploraremos a seguir alguns aspectos sobre o *docking* molecular.

### 1.3.3.1 Docking

O *docking* molecular consiste de uma metodologia de *SBDD* que visa investigar as possíveis poses que determinada molécula assume frente ao sítio de ligação de um receptor (proteína-alvo), baseando-se em sua estrutura 3D (KITCHEN et al., 2004). Vide representação esquemática na **Figura 1.11**. Destaca-se como um dos filtros mais importantes em *SBVS* (LI; SHAH, 2017; NEVES et al., 2021).



**Figura 1.11** - Representação esquemática de uma típica simulação de *docking* molecular. O sítio de ligação é localizado no receptor, ao qual são preditas um conjunto de poses para o ligante. As poses consistem de possíveis modelos para representar o complexo ligante-receptor. Cada pose é ranqueada por seu valor de *score* ( $S_1, S_2, \dots$ ), que dizem respeito a predição de afinidade de ligação ligante-receptor dos respectivos complexos [Fonte: adaptado de (BASCIU et al., 2022)].

Conforme reportado na literatura (LENGAUER; RAREY, 1996; CHANG et al., 2022), três principais objetivos podem ser alcançados por meio do *docking*: (i) predição de pose (em termos de orientação e conformação) para prever como um ligante pode se ligar ao receptor; (ii) triagem virtual para buscar novos candidatos a fármacos a partir de bibliotecas químicas; (iii) predição de afinidade de ligação usando funções de pontuação para estimar a afinidade de ligantes ao receptor.

De forma simplificada, os métodos de *docking* constituem-se da combinação de dois componentes: um algoritmo de geração conformacional e uma função de *score*/pontuação (HUANG; ZOU, 2010b; FEDERICO et al., 2021a).

Para a geração de conformações do ligante, considerando-se este flexível, destacam-se o algoritmo de busca sistemática e o algoritmo estocástico (BASCIU et al., 2022; VEMULA et al., 2023). A busca sistemática consiste na análise dos graus de liberdade do ligante, sendo muitas vezes utilizado a construção incremental baseada em fragmentos para tal tarefa (RAREY et al., 1996; LORBER; SHOICHET, 1998; GORELIK; GOLDBLUM, 2008). O algoritmo estocástico realiza a busca conformacional modificando aleatoriamente os parâmetros dos ligantes, gerando um conjunto de conformações moleculares e preenchendo uma ampla gama de cenários energéticos (VERDONK et al., 2003; SOUSA; FERNANDES; RAMOS, 2006; MORRIS et al., 2009).

Para então classificar as poses/conformações geradas, emprega-se funções de pontuação. Estas tem como objetivo calcular a afinidade da ligação entre o ligante e o receptor, avaliando as correspondentes forças de interações moleculares, dessolvatação e efeitos entrópicos (quando possível), envolvidos na formação do ligante-receptor (KITCHEN et al., 2004; VEMULA et al., 2023). Destacam-se três principais tipos de funções de pontuação: empíricas, de campo de força e baseadas no conhecimento.

As funções de pontuação empíricas analisam a afinidade do ligante com o receptor com base em seus números de átomos e derivando-se a soma de vários termos de energia empírica (com pesos específicos atribuídos), tais como energia de Van der Waals, energia eletrostática, energia de ligação de hidrogênio, termo de dessolvatação, termo de entropia, termo de hidrofobicidade, etc. (GUEDES; PEREIRA; DARDENNE, 2018).

O cálculo da pontuação de campo de força é estimado com base na decomposição da energia do complexo ligante-receptor em termos individuais para as interações intermoleculares (e.g. energias eletrostáticas, Van der Waals, energias de estiramento e torção, dentre outras) usando um conjunto de parâmetros derivados de campos de força *AMBER* ou *CHARMM*, por exemplo (MENG; SHOICHET; KUNTZ, 1992; WANG et al., 2001).

A função de pontuação baseada no conhecimento depende de parâmetros potenciais derivados a partir da observação de características estruturais verificadas



em complexos ligante-receptor determinados experimentalmente, utilizando-se do princípio do potencial de força média (MUEGGE, 2006; HUANG; ZOU, 2010a).

Contudo, os programas de *docking* são capazes de predizer com sucesso a pose do ligante no sítio de ligação do receptor, porém, frequentemente, não reproduzem de forma satisfatória a energia de interação absoluta do complexo. Nesse sentido, os programas vêm buscando contornar desafios relacionados, por exemplo, com a consideração de efeitos entrópicos e de dessolvatação nas respectivas funções de pontuação (RUVINSKY, 2007; MYSINGER; SHOICHET, 2010; LLANOS et al., 2021).

Vale mencionar que as funções de pontuação são frequentemente comparadas em diversos estudos de *benchmark* (WARREN et al., 2006; SPITZER; JAIN, 2012; SU et al., 2019), apresentando diferentes taxas de precisão a depender de condições específicas e, assim, mostrando que uma não pode substituir à outra (FERREIRA et al., 2015). Desta forma, o uso simultâneo de diferentes *softwares* visando a obtenção de resultados (poses e pontuação) em consenso tem sido cada vez mais explorado (CHARIFSON et al., 1999; CLARK et al., 2002; FEHER, 2006; DOS SANTOS MAIA et al., 2020).

Atualmente, outra questão diz respeito à aplicação de técnicas de inteligência artificial, sobretudo de *deep learning*, como uma forma de contornar o problema do tempo que se leva para rodar simulações de *docking* (DUTTA; BOSE, 2021; NAYARISSERI et al., 2021). Principalmente diante da expansão imensa das bases de dados disponíveis, para bilhões de moléculas - que fornecem acesso a áreas inéditas/desconhecidas do espaço químico -, surge o *deep docking* (GENTILE et al., 2020). A ideia é propor *workflows* que melhorem progressivamente a predição de *score* de *docking*, focando rapidamente nas moléculas mais promissoras dentro de uma biblioteca e descartando as de baixo *score*, sem que haja necessidade de, de fato, rodar as simulações de *docking*. Diversas metodologias vêm surgindo nesse sentido e parecem ser bastante promissoras (GORGULLA et al., 2020; BERENGER et al., 2021; FASSIO et al., 2022).

Mencionamos por fim que, na seção 3.1.2 de materiais e métodos, detalharemos mais sobre os algoritmos incorporados nos métodos/programas de *docking* **GOLD**, **Glide**, **FRED** e **Autodock**, que foram utilizados nesta Tese.

### 1.3.4 Predições ADME/Tox

Durante o planejamento racional de um novo fármaco também é possível estimar as propriedades *drug-like*, ou farmacocinéticas e toxicológicas, ou ADME/Tox (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade) para os compostos de interesse, selecionando, durante as etapas de VS, somente aqueles com potencial de se tornarem fármacos com biodisponibilidade apropriada (ABRAHAM, 2005; PAWAR et al., 2019).

São de extrema importância, pois a descoberta precoce destas informações pode reduzir as falhas decorrentes de propriedades ADME/Tox inapropriadas nos ensaios clínicos - que fazem parte das etapas mais custosas no desenvolvimento de fármacos (LAGORCE et al., 2010; REN; SAI; CHEN, 2019).

Além disso, predizer se um fármaco será atóxico, específico, ou ainda, se sofrerá algum tipo de biometabolismo até chegar ao seu alvo receptor, contribui reduzindo as etapas de planejamento e, assim, agilizando o processo de descoberta de um novo fármaco (AMASHITA; ASHIDA, 2004; BROWN et al., 2017; SILVA et al., 2021a).

Os estudos *in silico* de propriedades ADME/Tox objetivam, em geral, prever o comportamento *in vivo* de moléculas candidatas a fármacos. Atualmente, existem diversas metodologias e programas computacionais destinados a predição dos respectivos parâmetros definidos de ADME/Tox, e.g.: solubilidade aquosa (importante para a taxa de biodisponibilidade oral), constante de dissociação ácida  $pK_a$ , lipofilicidade expressa pelo coeficiente de partição em *n*-octanol/água  $\log P$  ou  $\log D$ , permeabilidade oral, gastrointestinal ou de barreira hematoencefálica (*BBB*, *blood-brain barrier*) (ALQAHTANI, 2017; DULSAT et al., 2023).

Um marco na utilização de regras/critérios *drug-like* que estimam relações entre características estruturais e biodisponibilidade oral, foi o trabalho de Lipinski et al. (1997), também conhecido como a 'Regra dos 5' (LIPINSKI et al., 1997; KUBINYI, 2007). Neste, considera-se que, para uma boa absorção ou permeabilidade, os potenciais fármacos devem apresentar peso molecular  $\leq 500$ ,  $\log P \leq 5$ , número de grupos doadores de ligações de hidrogênio (HBD)  $\leq 5$  e número de grupos aceptores de ligações de hidrogênio (HBA)  $\leq 10$ .

Após a Regra dos 5, diversas outras regras e subregras surgiram (TEAGUE et al., 1999; GOLDBERG et al., 2015) para auxiliar os pesquisadores durante o

desenvolvimento de *leads* e de fármacos, entretanto, foram evidenciadas as limitações das mesmas (LIPINSKI, 2016; DEGOEY et al., 2018). Ainda assim, pode-se sugerir, de modo geral, a aplicação de filtros *drug-like* durante as etapas do planejamento de fármacos, desde que estes sejam considerados de modo criterioso e com considerações pertinentes ao *sistema específico* estudado (YUSOF; SEGALL, 2013; MIGNANI et al., 2016).

Naturalmente, em vista de tais limitações e em conjuntura com a enorme expansão de dados experimentais/computacionais, novas abordagens surgiram para se avaliar a *drug-likeness* dos compostos (TIAN et al., 2015), i.e., suas capacidades de apresentarem propriedades ADME/Tox favoráveis, diante do *sistema* em questão.

Um exemplo é o *Quantitative Estimate of Drug-likeness (QED)* que incorpora o conceito de desejabilidade (*desirability*), via o método de otimização multiparamétrica (*MPO, Multi Parameter Optimisation*), para relacionar a *drug-likeness* de dado composto com as de fármacos orais com base em propriedades moleculares/estruturais simples (BICKERTON et al., 2012). O *QED* foi então extrapolado para o *sistema específico* de compostos relacionados ao SNC, sendo bastante interessante no escopo de doenças neurodegenerativas como a DA (RANKOVIC, 2015; WAGER et al., 2016).

Outros exemplos incluem, ainda, o *Relative Drug Likelihood* (YUSOF; SEGALL, 2013), o uso de otimizações multiobjetivos via *QSAR* (LAMBRINIDIS; TSANTILIKAKOULIDOU, 2018), o uso de relações entre *drug-likeness* e fragmentos moleculares (OVER et al., 2013), o uso de modelos preditivos via aprendizagem de máquina (TIAN et al., 2015) e a modelagem farmacocinética baseada em fisiologia (*Physiologically Based Pharmacokinetic, PBPK*) (WANG; OUYANG, 2022).

## 2 OBJETIVOS

São objetivos gerais deste Projeto de Pesquisa de Doutorado o planejamento computacional e subsequente avaliação biológica de inibidores alostéricos de GSK-3 $\beta$ , que representa uma proteína-alvo de interesse na busca por novos candidatos a fármacos em doença de Alzheimer, com vistas a otimizar as propriedades moleculares dos candidatos a fármacos assim planejados.

Para tanto, incluem-se os seguintes objetivos específicos:

- Avaliar e reforçar estudos prévios com relação às cavidades da GSK-3 $\beta$  e as poses observadas para os respectivos inibidores alostéricos;
- Obter e prever uma pose bioativa bem validada para uso em triagem virtual via metodologias baseadas em ligante, i.e., por similaridade 3D;
- Desenvolver protocolos de *docking* bem validados e de alta performance tanto para classificar os inibidores alostéricos de GSK-3 $\beta$  com maior atividade bem como para aplicação em triagem virtual;
- Desenvolver modelos preditivos e robustos de *QSAR* via aprendizagem de máquina para ser aplicado em triagem virtual;
- Selecionar e preparar bases de dados, de fontes e tamanhos diversos, para a execução das campanhas de triagem virtual, mesclando metodologias baseadas em ligante e em estrutura;
- Propor critérios robustos de propriedades farmacocinéticas e de toxicidade (ADME/Tox) para filtrar os compostos durante as triagens;
- Aplicar protocolos de *docking* validados em triagem virtual, para inspecionar as respectivas poses, interações e *scores*, selecionando, assim, *hits* em potencial;
- Conduzir ensaios biológicos *in vitro* (em colaboração) para a validação experimental dos *hits* adquiridos com respeito à inibição da GSK-3 $\beta$ ;
- Propor relações estrutura-atividade e analisar a diversidade estrutural verificada para os *hits* mais ativos;
- Validar *in silico* a habilidade dos *hits* mais ativos, de fato, se ligarem à cavidade alostérica 7 da GSK-3 $\beta$ .

### 3 MATERIAIS E MÉTODOS

Enfatizamos que ao longo do desenvolvimento do Projeto de Pesquisa desta Tese de Doutorado, foram surgindo novas ideias de estudos, entretanto sob a mesma temática e com os mesmos objetivos. Assim, diferentes publicações foram feitas na forma de comunicações em pôster, artigos e etc., como pode ser verificado na seção final Produção Científica.

Na presente Tese, todavia, descrevemos os resultados mais relevantes desta Pesquisa de forma organizada e correspondente a dois manuscritos (um já publicado e outro em submissão). Assim sendo, esta encontra-se subdividida em duas partes:

- *PARTE I* corresponde ao manuscrito publicado “Revisiting the Proposition of Binding Pockets and Bioactive Poses for GSK-3 $\beta$  Allosteric Modulators Addressed to Neurodegenerative Diseases” (SILVA et al., 2021b).
- *PARTE II* corresponde ao manuscrito em submissão “Discovery of Potential GSK-3 $\beta$  Allosteric Modulators for Alzheimer’s Disease by Integrated Computational Strategies”.

Assim, esperamos facilitar a leitura e compreensão desta Tese, conforme serão apresentados os respectivos métodos, resultados e discussões.

Destacamos que a maior parte deste Doutorado foi conduzido utilizando-se da infraestrutura e dos computadores situados no Laboratório Computacional de Química Farmacêutica (LCQF), liderado pelo Prof. Dr. Carlos Tomich, da Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo.

Outra parte dos trabalhos foram conduzidos durante período de Doutorado Sanduíche, junto ao grupo do Prof. Dr. Alexander Tropsha, pesquisador principal do *Molecular Modeling Laboratory (MML)* da *Eshelman School of Pharmacy* da *University of North Carolina at Chapel Hill, North Carolina, USA*.

### 3.1 **PARTE I: Investigação acerca das cavidades de GSK-3 $\beta$ e das poses bioativas dos respectivos inibidores alostéricos**

#### 3.1.1 **Detecção e predição de cavidades**

Foram utilizados os *softwares* **fpocket** (LE GUILLOUX; SCHMIDTKE; TUFFERY, 2009), **SuperStar** (SUPERSTAR, 2018), **metaPocket** (ZHANG et al., 2011), **Sitemap** (HALGREN, 2009) e **PARS** (PANJKOVICH; DAURA, 2014) para investigar as cavidades/sítios presentes na superfície da GSK-3 $\beta$ . Em geral, estes *softwares* empregam algoritmos para a detecção e mapeamento de cavidades em proteínas, permitindo a visualização da forma/volume das mesmas, juntamente com a predição dos correspondentes valores de pontuação, de acordo com o potencial da cavidade em interagir com potenciais ligantes e/ou drogas (*druggability score*).

Vale ressaltar que todos os *softwares* foram executados utilizando-se a estrutura cristalográfica da GSK-3 $\beta$  depositada no *Protein Data Bank* sob a identificação de PDB ID 1PYX, como *input*. Quando determinado *software* não forneceu a opção de automaticamente realizar um pré-processamento padrão, a referida estrutura foi preparada usando-se o *software* **Maestro** (SCHRÖDINGER, 2021) para adicionar hidrogênios, remover moléculas de água e outros ligantes/cofatores do complexo, e excluir a cadeia B da estrutura do dímero.

O **fpocket** é um *webserver* para a detecção de cavidades em proteínas que se baseia no método de tesselação de *Voronoi*. Brevemente, tal método matemático consiste de um tipo de decomposição de um dado espaço métrico, determinado pela distância para uma determinada família de objetos/sítios no espaço. Cada sítio está associado à “Célula de *Voronoi*” correspondente, isto é, um conjunto de todos os pontos no dado espaço o qual a distância para o dado sítio não é maior que sua distância para os outros objetos (AURENHAMMER, 1991). Então, o **fpocket** realiza a identificação dos sítios através de um algoritmo controlado por diversos parâmetros. Por exemplo, o ‘*Alpha Sphere Filtering*’ é um parâmetro que reúne todas as esferas alfa da estrutura da proteína e filtra as esferas alfa com raio  $m < R < M$ , onde  $m = 3,0$  e  $M = 6,0$  Å. Um segundo parâmetro é o ‘*Alpha Sphere Clustering*’, que identifica aglomerados de esferas próximas, ou seja, com centros de massa próximos, agregando-os em um único baricentro. Por fim, o ‘*Pocket Pruning Parameters*’ calcula diversas propriedades dos átomos do sítio, a fim de pontuar cada um com base em

uma abordagem de agrupamento de ligação múltipla, removendo pequenos sítios e pontuando e classificando os demais (LE GUILLOUX; SCHMIDTKE; TUFFERY, 2009).

O **SuperStar** utiliza um algoritmo homólogo ao do *software* **Ligsite** para realizar a detecção de cavidades na estrutura da proteína. Tal detecção pode ser realizada pela seleção de um resíduo ou de um ponto no espaço, mas também pode ser realizada para toda a proteína se nenhuma seleção for feita. O algoritmo pontua todos os pontos de *grid* de acordo com a profundidade em que estão enterrados, ou seja, valores altos indicam um ponto mais enterrado na cavidade. Assim, ele pode determinar o sítio de ligação retendo todos os pontos de *grid* com valores iguais ou superiores a 4 – valor de corte padrão (*default*) para cavidades normais. Esse valor também pode ser definido selecionando o tipo de cavidade como: rasa, rasa/normal, normal, normal/enterrada e enterrada (SUPERSTAR, 2018).

O **metaPocket** é um *webserver* que aplica oito algoritmos/*softwares* diferentes e independentes para detectar e prever consensualmente os possíveis sítios presentes na superfície da proteína. Esses algoritmos são divididos em dois tipos: baseados em geometria ou baseados em energia. O **Ligsite**, **GHECOM**, **Concavity**, **SURFNET**, **PASS**, **POCASA** e **fpocket** consistem de métodos baseados em geometria. O **Q-sitefinder** consiste de método baseado em energia – calculada por pequenas sondas moleculares (*probes*) e respectivas interações com regiões da proteína. O *webserver* roda cada um dos oito algoritmos, gerando os respectivos resultados. É calculada a métrica *z-score* para cada sítio e por cada método, uma vez que as funções de pontuação são específicas de cada algoritmo. Desta forma, são escolhidos apenas os três sítios mais bem pontuados por cada algoritmo, resultando em um conjunto total de 24 sítios. Estes, com o *z-score* calculado anteriormente, servem como a função de pontuação final para ranquear os sítios finais preditos pelo **metaPocket**. Tais sítios finais preditos pelo programa são representados pelo centro de massa de cada *cluster* final calculado (ZHANG et al., 2011).

O **Sitemap** opera de maneira análoga ao algoritmo **GRID** de Goodford. O seu cálculo começa pela seleção e agrupamento dos pontos relevantes do sítio com base em suas propriedades geométricas e energéticas. Desta forma, um *grid* é formado e os pontos são agrupados em conjuntos, de acordo com os vários critérios para se definir os sítios e, a seguir, são construídos mapas de contorno. Por fim, são gerados os arquivos de *output* compatível com o *software* **Maestro**, incluindo a visualização

dos mapas de contorno bem como das propriedades dos sítios, e.g. regiões hidrofílicas, hidrofóbicas, doadoras, aceitadoras, etc. (HALGREN, 2009; SCHRÖDINGER, 2021).

O **PARS** é um *webserver* que prediz, especificamente, sítios alostéricos na estrutura de proteínas. Este emprega a Análise de Modo Normal (*Normal Mode Analysis, NMA*) como método computacional para avaliar a dinâmica de proteínas e prever a localização de sítios alostéricos. Primeiramente, é executado o algoritmo **Ligsite** para prever possíveis sítios de ligação na estrutura da proteína. Em seguida, o *NMA* é executado na estrutura da proteína *apo* (i.e., sem ligante), bem como para cada potencial sítio de ligação definido previamente. No último caso, o *NMA* pode ser executado de duas maneiras diferentes: usando uma representação simplificada do ligante do complexo proteína-ligante ou usando um ligante cocrystalizado para o correspondente sítio de ligação. Então, as cavidades são classificadas de acordo com seu potencial como sítios alostéricos, considerando ainda as métricas calculadas de *p-value* e de conservação estrutural – em que o *p-value* está relacionado a uma mudança estatisticamente significativa observada na flexibilidade geral da proteína após a ligação do ligante, enquanto que os valores de conservação estrutural são relacionados a sítios alostéricos estruturalmente conservados dentro de uma família de proteínas (PANJKOVICH; DAURA, 2014).

### 3.1.2 Simulações preliminares de *docking*

Em antemão a execução das simulações de *docking* tanto os ligantes como as proteínas passaram por procedimentos de preparo.

Com relação aos ligantes (em especial ao ligante principal, **composto 1**, usado como principal referência neste trabalho) o procedimento de preparo consistiu de, inicialmente, obter as estruturas das moléculas no formato *SMILES* (e.g. desenhando suas estruturas no *software ChemDraw* (PERKINELMER INFORMATICS INC., 2021)). As estruturas em *SMILES* foram então importadas no **Ligprep** (SCHRÖDINGER, 2018a), em que foi aplicado o campo de força *MMFFs*, bem como cálculo de cargas parciais dos átomos e avaliação de possíveis estados de ionização por meio do *Epik* em  $\text{pH} = 7,0 \pm 1,0$ . Vale ressaltar que nos casos em que houve geração de diferentes espécies tautoméricas, foi mantida somente aquela com maior



probabilidade de predominância no referido pH. Por fim, as estruturas dos ligantes foram exportadas no formato *mol2* para uso.

Com relação às proteínas (em especial a enzima GSK-3 $\beta$ , de código *PDB* 1PYX) o preparo consistiu de inicialmente importar a estrutura do *Protein Data Bank* (*PDB*, <https://www.rcsb.org/>) para o *software Protein Preparation Wizard* (MADHAVI SASTRY et al., 2013; SCHRÖDINGER, 2018b). Então, foram aplicadas as seguintes funções: atribuição de ordens de ligação usando o banco de dados *CCD* (*Chemical Component Dictionary*) (WESTBROOK et al., 2015); adição de hidrogênios; geração de pontes dissulfeto; uso do *Prime* para preencher loops e cadeias laterais ausentes; remoção de moléculas de água; apenas o monômero A foi mantido, excluindo o monômero B; remoção de demais ligantes, cofatores e metais; átomos pesados foram convergidos e minimizados para 0,30 Å de *RMSD* usando o campo de força *OPLS3*.

Neste trabalho, cada cavidade alostérica da GSK-3 $\beta$  avaliada pelas metodologias de *docking* foi definida pelos seguintes centróides, em termos de coordenadas espaciais: cavidade 4 (x = 30,09, y = -3,98, z = 30,83), cavidade 5 (x = 46,11, y = 20,74, z = 31,39), cavidade 6 (x = 40,28, y = 10,02, z = 46,49) e cavidade 7 (x = 11,78, y = 13,15, z = 38,33). Vale notar que esta enumeração sequencial de cavidades corresponde à representação anterior da literatura (PALOMO et al., 2011), a fim de facilitar a interpretação (vide **Figura 1.7**, na introdução).

As simulações preliminares de *docking* foram realizadas usando configurações padrão em cada um dos 4 *softwares*. Mais detalhes sobre o refinamento dos protocolos de *docking* são mostrados na seção 3.1.5. A seguir, citamos apenas as particularidades de cada configuração padrão em cada um dos 4 *softwares* de *docking*, sendo eles **GOLD** (VERDONK et al., 2003; CCDC, 2022), **Glide** (HALGREN et al., 2004; SCHRÖDINGER, 2019), **AutoDock** (MORRIS et al., 2009; THE SCRIPPS RESEARCH INSTITUTE, 2020) e **FRED** (MCGANN, 2011; OPENEYE, 2020a).

No **GOLD** foi usada a função de pontuação *CHEMPLP* e um raio de esfera de 10 Å centralizado no centróide de cada cavidade.

No **Glide** foi utilizada a função de pontuação de *Glide XP* e *grids* de 10x10x10 Å centralizados em cada centróide.

No **Autodock** as configurações padrão foram usadas junto com *grids* de 40x40x40 Å centralizados em cada centróide.

No **FRED** primeiramente usamos **spruce4docking** (OPENEYE, 2020b) para processar a estrutura apo de GSK-3 $\beta$  (1PYX.pdb; previamente preparado conforme descrito acima), a fim de gerar os 'receptors' para as respectivas cavidades; indicadas por um resíduo de referência em cada cavidade. Em seguida, os ligantes foram processados pela **OMEGA** (OPENEYE, 2020c) para geração de 300 confôrmeros; assim, por fim, os *receptors* de cada cavidade e os sets de confôrmeros de cada ligante foram submetidos à execução de *docking* em modo padrão.

Como nesta Tese fizemos uso intenso destes 4 *softwares* de *docking*, aproveitamos aqui para apresentar resumidamente alguns detalhes sobre os respectivos métodos.

### **GOLD**

O **GOLD** (*Genetic Optimisation for Ligand Docking*) é um programa automatizado de *docking* que emprega algoritmo genético para explorar toda a gama de flexibilidade conformacional do ligante, incluindo-se flexibilidade parcial do receptor (i.e., dos resíduos que constituem o respectivo sítio de ligação). Em vista do seu pioneirismo (JONES; WILLETT; GLEN, 1995), ele vem sendo usado até hoje como um dos mais robustos métodos em *docking*.

O **GOLD** utiliza algoritmo estocástico via algoritmo genético para geração de confôrmeros e, assim, posicionar os ligantes no sítio do receptor. Inicialmente, são geradas populações consistindo de conformações aleatórias do ligante, em que cada indivíduo dessa população é codificado e chamado de 'cromossomo'. Então, tais cromossomos contém informações sobre os átomos dos ligantes, assim como sobre liberdades translacionais, rotacionais e conformacionais - chamadas de 'genes'.

Com vista à geração de conformações 'mais evoluídas', os cromossomos utilizam mecanismos denominados operadores genéticos, nos quais analisam-se possíveis mutações e recombinações. Desta forma, após o repetimento deste procedimento por inúmeras vezes, obtém-se a população final, evoluída a partir dos 'seus ancestrais' e sempre tendenciada a geração de poses com maior afinidade frente ao sítio do receptor (JONES et al., 1997; FEDERICO et al., 2021a).

Após gerada a tal população final, cada cromossomo (conformação do ligante) é *dockado* várias vezes no sítio em diferentes orientações e cada pose é pontuada usando as funções de *score* disponíveis: *GoldScore*, *ChemScore* e *CHEMPLP*

(LIEBESCHUETZ; COLE; KORB, 2012). Tais funções de *score* se baseiam no método empírico que, de modo geral, pode ser descrito pela Equação 3.1.

$$\Delta G = \sum_i W_i \cdot \Delta G_i$$

Equação 3.1

Em que,  $\Delta G_i$  representa os termos individuais de energia empírica e os correspondentes coeficientes  $W_i$  são determinados reproduzindo os dados de afinidade de ligação de uma série de treinamento de complexos proteína-ligante com estruturas 3D conhecidas usando o ajuste de mínimos quadrados (HUANG; ZOU, 2010b).

A função de pontuação *GoldScore* trata termos de dispersão e de contato (*clash*) considerando forças de Van der Waals, termos de ligação de hidrogênio e envolvendo metal-metal, com base em informações derivadas de estruturas cristalográficas (LIEBESCHUETZ; COLE; KORB, 2012).

A função de pontuação *Chemscore* é parametrizada a partir de 82 complexos conhecidos (em termos de afinidade de ligação), possuindo um termo de área de contato lipofílico-lipofílico, um termo de ligação de hidrogênio geometricamente restrito e um termo para penalizar a flexibilidade excessiva (VERDONK et al., 2003).

A função de pontuação *CHEMPLP* considera contatos neutros e repulsivos com um potencial linear por partes (*Piecewise Linear Potential, PLP*), o qual incorpora termos atrativos e repulsivos para contatos neutros e somente termos repulsivos para contatos anticomplementares (doador-doador, doador de metal e aceitador-aceitador). Os termos de ligação de hidrogênio e de energia interna são 'emprestados' da função *Chemscore*. Ressalta-se que o *CHEMPLP* é calculado mais rapidamente do que o *GoldScore* (KORB; STÜTZLE; EXNER, 2009).

### **Glide**

O software **Glide**, implementado no pacote/suite da *Schrödinger* (<https://www.schrodinger.com/products/glide>), consiste de um dos métodos de *docking* mais bem respaldados na literatura, desde a sua criação em 2004 (FRIESNER et al., 2004; HALGREN et al., 2004). Tal método emprega o algoritmo de busca sistemática para a geração de conformações do ligante e funções de *score*

empíricas, em ambas as opções: *Glide SP (standard precision)* e *Glide XP (extra precision)*.

Num primeiro passo, constrói-se um *grid* no sítio de interação do receptor para se obter informações sobre os possíveis tipos de interações que podem ocorrer nessa região. A forma e as propriedades do sítio do receptor são representadas neste *grid* por diferentes campos, permitindo então a pontuação progressivamente mais precisa da pose do ligante. Em seguida, as conformações para o ligante são geradas a partir de uma enumeração exaustiva de variações mínimas no ângulo de torção do ligante.

Estas são sistematicamente extrapoladas no espaço do *grid* e quando há então uma correspondência entre a conformação e o sítio de interação, esta é pré selecionada. Tal conformação é avaliada de acordo com seu diâmetro e possíveis rotações, assim como de acordo com seus átomos no espaço, a fim de evitar choques estéricos com os resíduos do sítio. Depois, o ligante é minimizado no sítio do receptor usando a função de energia de mecânica molecular *OPLS-AA* em conjunto com um modelo dielétrico dependente da distância. Ainda, para orientar adequadamente grupos periféricos, aplica-se o algoritmo de Monte Carlo. A função de pontuação *GlideScore* (ou *Glide SP*) - que corresponde a uma atualização do *ChemScore* (VERDONK et al., 2003) - é aplicada em combinação com a energia de interação ligante-receptor estimada por mecânica molecular e a energia de deformação do ligante para, assim, ranquear e selecionar as melhores poses de *docking*.

A função de *score Glide SP*, inclui termos de recompensa, por exemplo, o termo de recompensa para contatos entre os átomos hidrofóbicos do ligante e da proteína, em função da distância entre eles. Ainda, termos que recompensam tipos específicos de ligações de hidrogênio, em função do ângulo e da distância entre os átomos. Em adição, para computar interações eletrostáticas, emprega potencial de Coulomb (GUEDES; PEREIRA; DARDENNE, 2018)

Em contraste ao *Glide SP*, a função de *score Glide XP* (FRIESNER et al., 2006) contabiliza categorias distintas de ligações de hidrogênio, como interações neutro-neutro, carregado-carregado e neutro-carregado. Em adição, possui termos de penalização específicos implementados e desenvolvidos por meio de abordagens distintas para lidar com a dessolvatação de grupos polares e carregados de ligantes e proteínas. Outro diferencial do *Glide XP* diz respeito a capacidade de identificar grupos ou fragmentos estruturais específicos que fornecem contribuições excepcionalmente grandes para afinidade de ligação.

A atribuição precisa das penalidades de dessolvatação e do reconhecimento de grupos moleculares requer uma metodologia de amostragem expandida para desempenho ótimo. Assim, o *Glide XP* foi elaborado visando a otimização simultânea de ambos algoritmos de amostragem e de função de *score*. A função de *score* *Glide XP* é descrita a seguir pelas equações:

$$Glide\ XP = E_{coul} + E_{vdW} + E_{bind} + E_{penalty}$$

Equação 3.2

$$E_{bind} = E_{hydenclosure} + E_{hb_{nn}motif} + E_{hb_{cc}motif} + E_{PI} + E_{hb_{pair}} + E_{phobic_{pair}}$$

Equação 3.3

$$E_{penalty} = E_{desolv} + E_{ligand_{strain}}$$

Equação 3.4

Em que,  $E_{coul}$  e  $E_{vdW}$  diz respeito as energias de interações eletrostáticas com vista ao potencial de Coulomb e de forças intermoleculares de van der Waals, respectivamente. Enfatizamos que a descrição para tais termos, aqui apresentada, é resumida e pode ser aprofundada na literatura (FRIESNER et al., 2006).

$E_{bind}$  (Equação 3.3) diz respeito a um termo principal envolvendo as energias de interação e de recompensa, responsáveis por formar o complexo ligante-proteína, sendo descrito pelos termos:  $E_{hyd\_enclosure}$  (recompensa do compartimento/bolsão hidrofóbico),  $E_{hb\_nn\_motif}$  (termo para ligações de hidrogênio neutro-neutro em um ambiente hidrofóbico compartimentado),  $E_{hb\_cc\_motif}$  (termo para ligações de hidrogênio carregado-carregado),  $E_{PI}$  (recompensa pi-stacking/pi-cation),  $E_{hb\_pair}$  (termo de ligações de hidrogênio para um dado par átomo-átomo, semelhante ao *ChemScore*),  $E_{phobic\_pair}$  (termo de energia lipofílica para um dado par átomo-átomo).

$E_{penalty}$  (Equação 3.4) consiste do termo principal envolvendo energias impeditivas (penalidades) para a formação do complexo, sendo descrito pelos termos:  $E_{desolv}$  (penalidade de dessolvatação calculada via 'dockagem rápida de moléculas de água', i.e., um modelo de água explícito que captura a física básica da solvatação dentro dos limites da região do sítio ativo do complexo ligante-receptor) e  $E_{ligand\_strain}$  (penalidade de contato calculada nos casos graves em que as poses geradas possuem demasiados contatos/*clashes* internos, em vista da energia de deformação do ligante ao se acoplar no receptor).

### **Autodock**

O *software* de *docking* **Autodock** emprega o algoritmo de busca estocástica via algoritmo genético Lamarckiano para a geração de conformações, e uma função de pontuação empírica para o cálculo da energia livre de ligação (MORRIS et al., 1998, 2009). Em sua versão mais atual (usada nesta Tese), o **AutoDock** 4.2 utiliza um campo de força semi-empírico de energia livre para avaliar o acoplamento das poses (conformações previamente geradas) frente ao sítio do receptor.

Vale mencionar que tal campo de força foi previamente parametrizado usando um grande número de complexos ligante-receptor, para os quais tanto a estrutura quanto as constantes de inibição são experimentalmente conhecidas (AUTODOCK, 2014).

A avaliação de *docking* pelo **Autodock** se dá, portanto, em duas etapas: (i) as energias de interações intramoleculares são estimadas para o ligante em seu estado livre assim como para a conformação do ligante-receptor em seu estado complexado; (ii) avalia-se a energia intermolecular da combinação do ligante e da proteína em sua conformação complexada. A função de pontuação que incorpora o campo de força supracitado é descrita pela seguinte equação:

$$\Delta G = (V_{bound}^{L-L} - V_{unb}^{L-L}) + (V_{bound}^{P-P} - V_{unb}^{P-P}) + (V_{bound}^{P-L} - V_{unb}^{P-L} + \Delta S_{conf})$$

Equação 3.5

Em que,  $V$  corresponde a cada um dos termos, das seis avaliações em pares, incluindo energias relacionadas a dispersão/repulsão, ligação de hidrogênio, eletrostática e dessolvatação;  $L$  corresponde ao ligante;  $P$  corresponde ao receptor (proteína); *bound* traduzido do inglês para 'ligado', aqui sinônimo de complexado; *unb*, abreviação em inglês de *unbound*, tradução para não ligado, aqui sinônimo de estado livre;  $\Delta S_{conf}$  corresponde a uma estimativa da entropia conformacional perdida na formação do complexo ligante-receptor.

### **FRED**

A empresa OpenEye (<https://www.eyesopen.com/>) possui atualmente um pacote/*suite* de *softwares* de *docking* chamada **OEDocking** (OPENEYE, 2023a). Dentro desses, possui o **FRED** (MCGANN, 2011), o **HYBRID** (MCGANN, 2012) e o **POSEIT** (KELLEY et al., 2015).

Basicamente, o **POSEIT** é mais adequado para quando se usa uma série de complexos de ligantes cocristalizados, pois roda *docking* por meio da sobreposição de moléculas semelhantes com poses conhecidas, fornecendo, para cada pose, uma probabilidade robusta de que a pose dada seja razoável, via uma análise heurística simples. O **HYBRID** requer que a estrutura de um ligante ligado (nativo) seja conhecida, uma vez que usa informações tanto da estrutura do receptor quanto do ligante ligado (em termos de *shape*) para melhorar a performance do *docking*. O **FRED** consiste de um *software* tradicional de *docking*, que prediz as poses e afinidades de 1 ou mais moléculas em um sítio de receptor, usando um algoritmo de busca exaustivo.

Vale ressaltar que, para os 3 *softwares*, as multiconformações para o ligante a ser *dockado* tem de ser pré-computada em antemão, utilizando-se, e.g., o **OMEGA** (HAWKINS et al., 2010).

Para avaliar as multiconformações de um (ou vários) ligantes frente ao receptor no **FRED**, realiza-se uma busca exaustiva pelas várias conformações do ligante dentro do sítio de ligação, variando-se sistematicamente as respectivas magnitudes rotacionais e translacionais. Em sequência, as poses mais bem pontuadas são otimizadas e recebem uma pontuação final.

As funções usadas para pontuar e otimizar tais poses consistem de *Chemgauss3* e *Chemgauss4*, respectivamente. A função de pontuação *Chemgauss3* emprega potenciais gaussianos suavizados para mensurar a complementaridade das poses do ligante no sítio ativo, reconhecendo os seguintes tipos de interações: forma, interações do tipo metal-quelante, interações de ligação de hidrogênio entre o ligante e a proteína, bem como com solvente implícito.

Todos os potenciais de interação do *Chemgauss3* são inicialmente obtidos por meio de sub-funções para descrever a interação de pares de átomos em função da distância. Tais interações são mapeadas em um *grid*, visando suavizar o potencial em direção a uma menor sensibilidade por variações de posição do ligante. Assim, as simulações de *docking* podem se dar em resoluções menores que seria requerido, se o *score* não fosse suavizado. Também, reduz o erro associado com a aproximação de proteína rígida, considerando a habilidade do receptor em se rearranjar para acomodar o ligante.

A diferença da função de pontuação *Chemgauss4*, usada na etapa final de otimização da pose, é somente em relação ao aprimoramento dos termos de interação de ligação de hidrogênio e metal-quelante. Os novos termos de ligação de

hidrogênio e quelante de metal possuem uma melhor percepção da direcionalidade dessas interações e também consideram os efeitos de rede de ligações de hidrogênio.

### 3.1.3 Geração de mapas de contorno e superfície

Foram utilizados os *softwares* **Superstar** (SUPERSTAR, 2018), **GRID** (GOODFORD, 1985) e **Sitemap** (HALGREN, 2009) para gerar mapas de contorno e superfície usando a estrutura GSK-3 $\beta$  (PDB ID 1PYX). A estrutura da proteína foi preparada conforme descrito na [seção 3.1.2](#) antes da execução destes *softwares*.

O **Superstar** é um *software* que usa o método empírico para calcular mapas que representam a tendência (*propensity*) de um grupo funcional (sonda/*probe*) em se ligar em diferentes posições ao redor de um sítio de ligação de proteína. O *software* usa dados do *Cambridge Structural Database (CSD)* e *Protein Data Bank (PDB)*, usando a base de conhecimento *IsoStar* como intermediário. A vantagem de utilizar um método baseado em conhecimento é a possibilidade de obter resultados influenciados por um modelo ajustado aos dados reais, porém, a desvantagem é que pode haver algum viés de uma amostra super-representada na base de dados. A fonte de dados *IsoStar*, além de cobrir uma ampla gama de interações, também apresenta bom desempenho para uma grande variedade de complexos proteína-ligante de acordo com suas validações (SUPERSTAR, 2018).

O **GRID** permite detectar sítios de ligação energeticamente favoráveis para grupos funcionais em macromoléculas, empregando campos de interação molecular (*molecular interaction fields, MIFs*) e diferentes sondas/*probes* químicas. Seu método é baseado em “pontos de grade” (*grid points*) que são sobrepostos na proteína e a energia potencial de cada sonda é calculada a partir de um grupo químico pré-determinado. Para cada posição da sonda, a energia de interação não ligada  $E_{xyz}$  é calculada usando uma função de energia empírica. Existem mais de 60 sondas disponíveis no **GRID**, das quais as mais importantes incluem NH (amina doador de ligação de H), O (aceptor de ligação de H), C=O (carbonila), DRY (hidrofóbica) e CH<sub>3</sub> (grupo metil). As superfícies de contorno podem ser analisadas por uma interface de visualização molecular implementada no próprio *software* (GOODFORD, 1985).

O **Sitemap** realiza o cálculo de mapas de contorno identificando regiões hidrofóbicas, doadoras de ligação de hidrogênio, aceitadoras de ligação de hidrogênio e de ligação de metal na proteína, também usando um *grid* de pontos de localização.



Para o caráter hidrofóbico e hidrofílico, a média da energia de interação de van der Waals - sobre o sítio original e pontos estendidos da sonda-receptor - é computada e para as propriedades de doar/aceitar ligações de hidrogênio; assim, o algoritmo é capaz de calcular e sugerir quais regiões são preferenciais (HALGREN, 2009; KRISHNAN et al., 2018).

#### 3.1.4 Compilação da série de inibidores alostéricos

Como ponto de partida, usamos o conjunto conhecido de 40 inibidores alostéricos GSK-3 $\beta$  (análogos do **composto 1**, de 1 a 40 na **Tabela 3.1**), com valores de IC<sub>50</sub> medidos, conforme reportado na literatura (PALOMO et al., 2017).

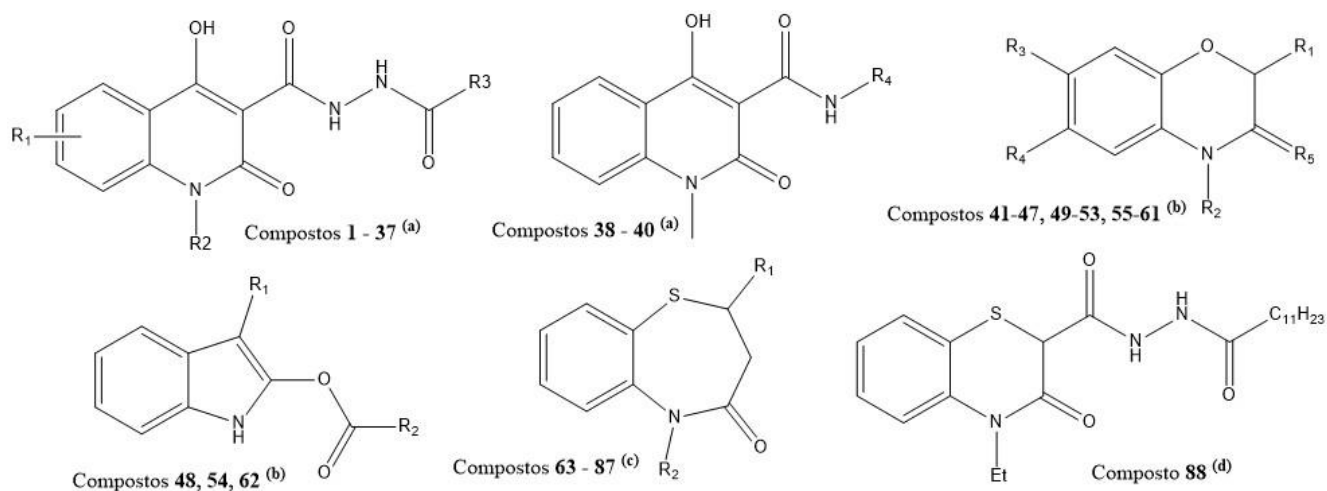
Em uma pesquisa preliminar na literatura, verificamos que não há muitos compostos relatados especificamente como inibidores alostéricos da GSK-3 $\beta$  - apesar da quantidade considerável de inibidores ortostéricos para tal enzima. Além disso, muitos inibidores são descritos como não competitivos de ATP, porém carecem de informações sobre qual cavidade alostérica, especificamente, deve hospedar tais inibidores, não permitindo o seu uso.

Assim, após uma busca exaustiva na literatura para se encontrar mais compostos, acabamos encontrando mais 48 compostos que podem ser incluídos ao nosso *dataset* de forma confiável, conforme mostrado na **Tabela 3.1**. Totalizando um conjunto de dados compilados de 88 inibidores alostéricos GSK-3 $\beta$  que devem atuar especificamente na cavidade alostérica 7.

#### 3.1.5 Refinamento dos protocolos de *docking*

Conforme mencionado, preliminarmente, foram usados protocolos padrões para as simulações de *docking*. Enfatiza-se que, com o termo protocolo nos referimos a uma combinação de ajustes de parâmetros (*settings*) em cada *software* e para cada proteína (aqui, no caso, somente a GSK-3 $\beta$  - 1PYX.pdb).

**Tabela 3.1** - Série de 88 inibidores alostéricos de GSK-3 $\beta$  compilados da literatura (a, b, c, d) e os correspondentes valores de IC<sub>50</sub> ( $\mu$ M).



	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	IC <sub>50</sub> ( $\mu$ M)		R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	IC <sub>50</sub> ( $\mu$ M)
1	H	Et	C <sub>11</sub> H <sub>23</sub>	-	-	2,80	45	1-naftil	4-OMe-bn	H	H	O	8,10
2	H	H	C <sub>11</sub> H <sub>23</sub>	-	-	4,50	46	1-naftil	4-F-bn	OCH <sub>2</sub> O	H	O	8,10
3	H	Me	C <sub>11</sub> H <sub>23</sub>	-	-	7,30	47	2-naftil	4-F-bn	H	H	O	8,70
4	H	Me	C <sub>7</sub> H <sub>15</sub>	-	-	8,70	48	1-naftil	Me	-	-	-	10,00
5	H	isoprenyl	C <sub>11</sub> H <sub>23</sub>	-	-	7,78	49	1-naftil	4-F-bn	1-morpholin	H	O	10,10
6	H	isoprenyl	C <sub>7</sub> H <sub>15</sub>	-	-	7,11	50	1-naftil	4-F-bn	F	H	O	21,20
7	H	CH <sub>2</sub> -cyclopropyl	C <sub>11</sub> H <sub>23</sub>	-	-	5,25	51	1-naftil	Et	OCH <sub>2</sub> O	H	O	25,60
8	H	Bn	C <sub>11</sub> H <sub>23</sub>	-	-	5,51	52	1-naftil	4-CH <sub>3</sub> -bn	H	H	O	50,00
9	6-F	H	C <sub>11</sub> H <sub>23</sub>	-	-	5,80	53	phenyl	Bn	OCH <sub>2</sub> O	H	O	50,00
10	6-Cl	H	C <sub>11</sub> H <sub>23</sub>	-	-	3,81	54	4-ch3-bn	Me	-	-	-	50,00
11	6-Cl	Me	C <sub>11</sub> H <sub>23</sub>	-	-	5,95	55	1-naftil	benzil	H	H	O	>50
12	6-Cl	Me	C <sub>7</sub> H <sub>15</sub>	-	-	8,48	56	1-naftil	4-Cl-bn	H	H	O	>50
13	6-Cl	Et	C <sub>11</sub> H <sub>23</sub>	-	-	3,18	57	1-naftil	4-NO <sub>2</sub> -bn	H	H	O	>50
14	6-Cl	Et	C <sub>7</sub> H <sub>15</sub>	-	-	9,42	58	1-naftil	benzoyl	H	H	HH	>50
15	6-Br	H	C <sub>11</sub> H <sub>23</sub>	-	-	3,54	59	1-naftil	Me	OCH <sub>2</sub> O	H	O	>50
16	6-Br	Me	C <sub>11</sub> H <sub>23</sub>	-	-	4,28	60	phenyl	Me	OCH <sub>2</sub> O	H	O	>50
17	6-Br	Me	C <sub>7</sub> H <sub>15</sub>	-	-	6,28	61	phenyl	Et	OCH <sub>2</sub> O	H	O	>50
18	6-Br	Et	C <sub>11</sub> H <sub>23</sub>	-	-	2,01	62	1-naftil	NEt <sub>2</sub>	-	-	-	>50
19	6-Br	Et	C <sub>7</sub> H <sub>15</sub>	-	-	7,34	63	2-Thienyl	Bn	-	-	-	47,50
20	7-Cl	H	C <sub>11</sub> H <sub>23</sub>	-	-	3,12	64	2-Furyl	Bn	-	-	-	77,20
21	7-Cl	H	C <sub>7</sub> H <sub>15</sub>	-	-	9,00	65	Ph	Bn	-	-	-	42,70
22	7-Cl	Me	C <sub>11</sub> H <sub>23</sub>	-	-	4,03	66	Ph	2-NO <sub>2</sub> -Bn	-	-	-	25,00
23	7-Cl	Me	C <sub>7</sub> H <sub>15</sub>	-	-	6,66	67	Ph	2-Br-Bn	-	-	-	73,90
24	7-Cl	Et	C <sub>11</sub> H <sub>23</sub>	-	-	2,48	68	Ph	2-CH <sub>3</sub> -Bn	-	-	-	76,10
25	7-Cl	Et	C <sub>7</sub> H <sub>15</sub>	-	-	4,83	69	Ph	4-CH <sub>3</sub> O-Bn	-	-	-	73,70
26	7-Cl	Bn	C <sub>11</sub> H <sub>23</sub>	-	-	5,99	70	Ph	3-COOMe-Bn	-	-	-	27,80
27	7-Cl	Bn	C <sub>7</sub> H <sub>15</sub>	-	-	5,96	71	Ph	3-Cl-PhCOCH <sub>2</sub>	-	-	-	37,80
28	H	H	C <sub>7</sub> H <sub>15</sub>	-	-	>20	72	PhCH <sub>2</sub>	2-NO <sub>2</sub> -Bn	-	-	-	23,00
29	H	Me	Ph	-	-	>20	73	4-F-Ph	2-NO <sub>2</sub> -Bn	-	-	-	81,50
30	H	Me	Me	-	-	>20	74	4-Cl-Ph	2-NO <sub>2</sub> -Bn	-	-	-	71,30
31	H	Me	Bn	-	-	>20	75	4-Br-Ph	2-NO <sub>2</sub> -Bn	-	-	-	67,80
32	H	Me	(CH <sub>2</sub> ) <sub>2</sub> -indol-3-yl	-	-	>20	76	2-Thienyl	Et	-	-	-	>100
33	H	Et	C <sub>7</sub> H <sub>15</sub>	-	-	>20	77	2-Thienyl	i-Pr	-	-	-	>100
34	H	Et	(CH <sub>2</sub> ) <sub>2</sub> -indol-3-yl	-	-	>20	78	2-Thienyl	n-Bu	-	-	-	>100
35	H	Et	(CH <sub>2</sub> ) <sub>2</sub> -(4-OH-Ph)	-	-	>20	79	2-Thienyl	Cyclohexthylmethyl	-	-	-	>100
36	H	isobutyl	C <sub>11</sub> H <sub>23</sub>	-	-	>20	80	2-Thienyl	Benzoyl	-	-	-	>100
37	6-F	H	C <sub>7</sub> H <sub>15</sub>	-	-	>20	81	H	Bn	-	-	-	>100
38	-	-	-	C <sub>6</sub> H <sub>13</sub>	-	>20	82	Me	Bn	-	-	-	>100
39	-	-	-	CH <sub>3</sub> CONHCH <sub>2</sub> COOBn	-	>20	83	3-Pyridyl	Bn	-	-	-	>100
40	-	-	-	NH-(CH <sub>2</sub> ) <sub>2</sub> -CN	-	>20	84	Ph	2-CN-Bn	-	-	-	>100
41	1-naftil	SO <sub>2</sub> -phenyl	H	H	HH	4,10	85	Ph	2-F-Bn	-	-	-	>100
42	1-naftil	4-F-bn	H	H	O	5,40	86	Ph	2-Cl-Bn	-	-	-	>100
43	1-naftil	bn	OCH <sub>2</sub> O	H	O	6,40	87	Ph	3-COOH-Bn	-	-	-	>100
44	2-naftil	bn	OCH <sub>2</sub> O	H	O	6,90	88	-	-	-	-	-	8,00

\* Em verde, os compostos considerados como ativo e em vermelho os inativos, de acordo com o valor de corte de 20  $\mu$ M para valores de IC<sub>50</sub>. (a) (PALOMO et al., 2017); (b) (BROGI et al., 2017); (c) (ZHANG et al., 2013); (d) (ZHANG et al., 2014).

Em adição, para todos os *softwares* aqui utilizados, ou seja, **GOLD**, **Glide**, **Autodock** e **FRED**, os resultados foram programados para gerar 10 poses de *docking* para cada molécula e, em alguns casos, menos poses foram geradas (de acordo com o limiar (de corte) de energia estabelecido por cada *software*).

As poses obtidas para cada molécula foram inspecionadas visualmente e, assim, separadas em “clusters” (conjuntos/famílias) de poses aparentemente mais sobrepostas, que assim representavam *clusters* de poses homogêneas. Dentre as famílias com maior número de poses homogêneas, aquela com o melhor valor de pontuação foi selecionada como a pose representativa. Quando os resultados das poses foram muito dispersos (baixa homogeneidade entre as poses), a pose com a melhor pontuação foi simplesmente selecionada como a representativa daquela molécula. Além disso, destacamos que não foram utilizadas *constraints* em nenhum dos *softwares*.

Para o refinamento, diferentes centróides (coordenadas x, y e z) foram avaliados em relação à distância entre os principais resíduos que constituem a cavidade 7. Adicionalmente, para **GOLD** e **Glide**, também foram avaliadas mudanças em relação à flexibilidade das cadeias laterais dos resíduos Arg209 e Ser236 da GSK-3 $\beta$  (1PYX.pdb).

Os seguintes parâmetros foram sistematicamente variados no **GOLD**: tamanho da esfera de 7 a 12 Å, funções de pontuação *CHEMPLP*, *ChemScore* e *GoldScore*, tamanho da população e operações em “configurações de GA” e permitindo (ou não) a inversão de N (amida) durante a geração de confôrmeros. O protocolo de *docking* final selecionado no **GOLD** foi: centroide de x=11,78, y=13,15, z=38,33, raio de esfera de 10 Å, 4 rotâmeros permitidos para Ser236, e função de pontuação *CHEMPLP*.

No **Glide**, variou-se: o comprimento de cada lado do *grid* de 10 a 20 Å, as funções de *score* *Glide SP* e *Glide XP*, permitindo (ou não) a inversão de N (amida) durante a geração dos confôrmeros. O protocolo de *docking* final selecionado no **Glide** foi: mesmo centróide que **GOLD**, função de pontuação *Glide XP*, não permitindo inversão de N, mantendo o resíduo Ser236 flexível.

No **Autodock**, além do protocolo padrão, também tentamos executá-lo conforme descrito na literatura (PALOMO et al., 2011). O protocolo de *docking* final selecionado foi então: centróide x=10,508, y=13,154, z=37,338, *grid* de 40 x 40 x 38 Å com espaçamento de 0,375 Å, GA Lamarckiano (número de corridas de GA = 200,

tamanho da população = 200, número máximo de *evals* = 2500000), e resíduo Ser236 flexível.

No **FRED**, executamos simulações de *docking* nas resoluções *Standard*, *High* e *Low*. A função de *score* padrão *Chemgauss4* foi mantida. O protocolo de *docking* final selecionado foi: indicação do resíduo Ser236 como referência para indicar a cavidade 7 durante a geração do *receptor* pelo *spruce4docking*, seguido pela execução das simulações de *docking* na resolução padrão/*Standard*.

### 3.1.6 Validação dos protocolos de *docking*

Para avaliar a performance de cada protocolo de *docking*, desenvolvido e refinado em cada *software*, aplicamos um procedimento de validação de acordo com as suas capacidades de discriminar compostos ativos e inativos (ou *decoys*), analisados por curvas *ROC* (*Receiver Operating Characteristic*) e respectivas métricas. Preliminarmente, foi realizada uma validação mais simples ([seção 4.1.4](#)) usando apenas o pequeno *dataset* de 40 inibidores alostéricos de GSK-3 $\beta$  (13 inativos contra 27 ativos), conhecidos da literatura (PALOMO et al., 2017). Posteriormente, para otimização ([seção 4.1.5](#)), usou-se o *dataset* com 88 compostos (51 inativos contra 37 ativos), incluindo e não incluindo *decoys* adicionais gerados pelo **DUD-E** (MYSINGER et al., 2012) - vide **Tabela 3.1** para melhor entendimento dos *datasets*.

*Decoys* são estruturas potencialmente inativas geradas computacionalmente, com vista a apresentar propriedades físico-químicas semelhantes, mas estruturas químicas e topologias diferentes. Estes *decoys* foram gerados a partir do *input* de compostos ativos no *webserver* **DUD-E**, ou seja, usando as estruturas dos 37 ativos. Tal *webserver* possibilita a geração de 50 *decoys* por ativo, portanto foram gerados 1850 *decoys* adicionais. Assim, somou-se tais *decoys* com os 51 compostos inativos, totalizando 1901 'inativos', e estes foram usados contra os 37 ativos conhecidos; por fim obtendo um *dataset* totalizando aproximadamente 1938 compostos.

Desta forma, validou-se os protocolos refinados dos *softwares* **GOLD**, **Glide**, **Autodock** e **FRED**, considerando os respectivos valores de *docking score* obtidos para cada *dataset*. Os valores de *score* foram ranqueados (ordenados do melhor ao pior) de acordo com os códigos binários correspondentes - indicando ativo/inativo para cada composto, ou seja, 1 para ativo e 0 para inativo. As curvas *ROC* foram construídas usando o *webserver* *Screening Explorer*, <http://stats.drugdesign.fr/>

(EMPEREUR-MOT; ZAGURY; MONTES, 2016), e foram avaliadas pelos respectivos valores de *AUC* (*Area Under the Curve*).

As validações através da análise de curvas *ROC* permitem avaliar a capacidade de uma dada metodologia/modelo classificar corretamente compostos ativos (ou seja, compostos com atividade biológica conhecida e demonstrada por resultados experimentais) dentro de um *dataset* contendo compostos inativos. Em geral, os compostos são classificados em relação às razões verdadeiro/falso positivo/negativo e são representados em termos de sensibilidade e especificidade. Assim, para a construção das curvas *ROC*, a sensibilidade (eixo y) é correlacionada em função da 1-especificidade (eixo x), sendo a sensibilidade definida pela razão de verdadeiros positivos e a 1-especificidade pela razão de falsos positivos classificados pelo modelo. É possível avaliar curvas *ROC* pelos respectivos valores de *AUC*, sendo que valores próximos a 1,0 indicam idealidade e melhor desempenho do modelo, enquanto o valor de 0,5 (linha diagonal no gráfico) indica uma classificação aleatória pelo modelo (TRIBALLEAU et al., 2005; NICHOLLS, 2008).

Outras duas métricas relacionadas as validações por curvas *ROC* são o ganho total padronizado (*TG*, *Total Gain*) (BURA; GASTWIRTH, 2001) e a discriminação reforçada por Boltzman do *ROC* (*BEDROC*, *Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic*) (TRUCHON; BAYLY, 2007). Ambas são obtidas a partir das curvas de predição e de enriquecimento, respectivamente, que por sua vez são derivadas de análises de curvas *ROC* (EMPEREUR-MOT et al., 2015).

O *TG* resume a discriminação dos compostos ativos em relação a variação das pontuações ao longo do *dataset* de moléculas e seus valores variam de 0 a 1. Em geral, um valor de  $TG > 0,25$  (com  $AUC > 0,5$ ) indica variações de pontuação relevantes para discriminar os compostos ativos, enquanto que  $TG > 0,4$  (com  $AUC > 0,5$ ) indica um bom desempenho e reprodutibilidade em condições experimentais semelhantes para o método/modelo avaliado. O *BEDROC* também varia de 0 a 1, e é usado para quantificar o reconhecimento precoce dos compostos ativos usados na validação. Ou seja, a probabilidade de um composto ativo ser classificado antes de um composto aleatoriamente selecionado (EMPEREUR-MOT; ZAGURY; MONTES, 2016).

### 3.1.7 Validação de *query(ies)* para uso em similaridade 3D

As poses de *docking* do **composto 1** - obtida pelo melhor protocolo de cada *software* -, foi validada para verificar seu potencial uso como *query* (molécula de referência) na busca por similaridade 3D por forma (*shape similarity*) para o estudo de prospecção virtual (Campanha 1 de triagem virtual seção 3.2.1).

Para tanto, utilizou-se o *software* **vROCS** diante de seu procedimento mais facilitado para validação, contemplando visualização gráfica bem como análise de curva ROC e de valores de *AUC*. Vale notar que o **vROCS** é uma versão alternativa da OpenEye ao **ROCS**, porém que opera de maneira idêntica.

Também, em adição às 4 poses geradas pelos *softwares* **GOLD**, **Glide**, **Autodock** e **FRED**, foi utilizada como *query* a pose de conformação mínima de energia para o **composto 1** gerada pelo **OMEGA** - estratégia recomendada pelos desenvolvedores do **ROCS** para a validação e seleção de *queries* em estudos de prospecção (HAWKINS; NICHOLLS, 2012; OPENEYE, 2020b).

Conforme descrito anteriormente, dois *datasets* foram usados. Um primeiro de 88 compostos, consistindo de 37 ativos e 51 inativos (vide **Tabela 3.1**); e um segundo consistindo de 37 ativos e ~ 1901 'inativos' (incluindo *decoys*). Ambos *datasets* foram previamente submetidos à geração de 300 confôrmeros usando o *software* **OMEGA** (OPENEYE, 2020c).

Para classificar/discriminar as moléculas, foram considerados os respectivos valores de pontuação de *ROCSTC* obtidos pelo **vROCS**. Tais valores variam de 0 a 2, sendo que valores mais próximos de 2 indicam uma alta sobreposição conformacional entre dada molécula e o *query* utilizado, em termos de forma (*ShapeTanimoto*) + características químicas (*ColorTanimoto*). A partir dos resultados gerados, curvas ROC foram então geradas para a obtenção dos respectivos valores de *AUC*, de modo análogo ao descrito acima na seção 3.1.6.

### 3.1.8 Estudos de dinâmica molecular

Os estudos de dinâmica molecular (DM) foram conduzidos utilizando o *High-Performance Computing* (HPC), do Centro de Computação de Alto Desempenho da Universidade de São Paulo (USP). Para as simulações de DM, utilizamos o *software* **GROMACS** 2019.3 (CHÁVEZ THIELEMANN et al., 2019) com o campo de força *Charmm36* (HUANG et al., 2016), e os ligantes foram parametrizados pelo *CGenFF*

(VANOMMESLAEGHE; MACKERELL, 2012). As coordenadas iniciais para conduzir as simulações DM foram obtidas a partir das poses de *docking* dos ligantes, obtidas respectivamente pelo **Glide** e pelo **GOLD** (conforme descrito na [seção 3.1.5](#)).

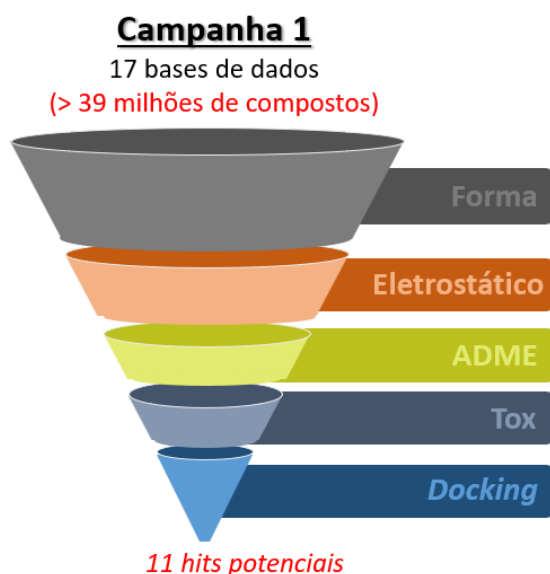
A estrutura preparada da proteína GSK-3 $\beta$  (PDB ID 1PYX) foi colocada em uma caixa cúbica com vetores 9,36x9,36x9,36 nm e solvatada com 24.715 moléculas de água *TIP3P* (JORGENSEN et al., 1983). Para neutralizar as cargas neste sistema, adicionamos 6 íons de Cl<sup>-</sup>. O tempo total de simulação foi de 100 ns, com tempo de integração de 2 fs. O sistema foi minimizado utilizando-se o método *Steepest Descent*, para evitar contatos desfavoráveis entre os átomos, e a convergência foi arquivada em energia potencial abaixo de 500 kJ/mol.nm. Em seguida, utilizamos ambas metodologias de NVT e NPT para atingir o equilíbrio, mantendo-se constante a pressão em 1 bar (por meio do barostato de Berendsen) e a temperatura em 300 K, com acoplamentos de tempo de 2,0 e 0,1 ps, respectivamente. Por fim, realizamos a análise estrutural e da produção da trajetória seguindo o mesmo protocolo publicado recentemente (DE OLIVEIRA et al., 2020; FEDERICO et al., 2020).

### 3.2 **PARTE II**: Triagens virtuais para a identificação de inibidores alostéricos de GSK-3 $\beta$

Nesta parte do trabalho executamos duas campanhas de triagem virtual, usando fontes distintas de bases de dados e variando a ênfase das metodologias empregadas para a triagem de compostos, conforme mostrado pelas **Figura 3.1** e **Figura 3.2**. As metodologias empregadas estão também descritas em detalhes nos respectivos tópicos, assim como as bases de dados utilizadas em cada campanha.

Os *hits* obtidos em cada campanha de triagem foram então adquiridos e testados por ensaios biológicos frente a inibição de GSK-3 $\beta$ . Estudos adicionais foram realizados para analisar as relações de estrutura-atividade e diversidade estrutural, bem como para a validação *in silico* do mecanismo de ação dos *hits*.

#### 3.2.1 **Campanha 1**: Triagem baseada em forma e em estrutura



**Figura 3.1** - Campanha 1 de triagem virtual exibindo as metodologias aplicadas para a obtenção de potenciais *hits* como moduladores alostéricos GSK-3 $\beta$ .

Esta Campanha 1 de triagem virtual teve como metodologia inicial e mais enfática a triagem baseada em forma (*shape-based*), conduzida considerando a similaridade 3D de forma do *software* **ROCS** (HAWKINS; SKILLMAN; NICHOLLS, 2007; OPENEYE, 2020d). Foi ainda considerada a similaridade 3D de potencial eletrostático do *software* **EON** (OPENEYE, 2020e). Ambos os *softwares* são detalhados no tópico 3.2.1.2.



Após filtragem dos compostos por similaridade 3D, os remanescentes foram submetidos a predições farmacocinéticas e toxicológicas (ADME/Tox) usando os *softwares* **Qikprop** e **DEREK**, respectivamente, conforme descrito no tópico 3.2.3.

A metodologia baseada em estrutura e por fim aplicada nesta campanha consistiu de *docking*. Aplicamos nosso protocolo robusto de *docking* usando o **Glide**, desenvolvido anteriormente (vide seção 3.1.5), a fim de avaliar a capacidade dos compostos até aqui filtrados em estabelecer interações intermoleculares consistentes no sítio alostérico da GSK-3 $\beta$ . Demais detalhes sobre como foram conduzidas as simulações de *docking* estão descritas no tópico 3.2.4.

Por fim, inspeção visual dos compostos até aqui filtrados foi conduzida minuciosamente para selecionar os compostos *hits* mais promissores e com ampla diversidade estrutural.

### **3.2.1.1 Bases de dados (Campanha 1)**

As dezessete bases de dados apresentadas na **Tabela 3.2** foram usadas na Campanha 1 de triagem virtual. Somadas, totalizam mais de 39 milhões de moléculas.

A preparação destas bases de dados foi realizada, individualmente, pela aplicação do **FILTER** (OPENEYE, 2020c), usando parâmetros padrão e as seguintes configurações adicionais: máximo de 2 centros quirais, 5 sistemas de anéis, 20 átomos por sistema de anéis, 16 ligações rotacionáveis, e 55 ligações rígidas. O uso de **FILTER** é altamente recomendado (OPENEYE, 2020b) para se eliminar compostos indesejados/inúteis/inviáveis em um *pipeline* de modelagem molecular - antes da geração e processamento de conformadores pela **OMEGA** -, principalmente com relação a moléculas contendo muitas ligações rotacionáveis (alta flexibilidade) e/ou muitos anéis (especialmente os flexíveis).

Na sequência, o *software* **OMEGA** (HAWKINS et al., 2010; OPENEYE, 2020c) foi usado para gerar 300 conformeros por molécula, considerando os seus parâmetros padrão, exceto para: janela de energia ajustada para 9,0 kcal/mol e RMSD ajustado para 0,6 Å (adaptações de nosso grupo de pesquisa (DA SILVA; TAFT, 2017)). Como os *softwares* **ROCS** e **EON** exigem diversidade conformacional entre as moléculas das bases de dados e a estrutura do *query* - para realizar sobreposição e analisar similaridade por forma e eletrostática, respectivamente - esta etapa de processamento

para gerar os confôrmeros usando o **OMEGA** é essencial durante a preparação das bases de dados.

**Tabela 3.2** - Dezesete bases de dados utilizadas na Campanha 1 de triagem virtual.

Nome	Versão/Ano	Referência	Número de moléculas
Chembridge CNS	2015	(CHEMBRIDGE, 2020)	53.306
Chembridge CORE Library Stock	2020	(CHEMBRIDGE, 2020)	747.804
Chembridge Diverset CL	2015	(CHEMBRIDGE, 2020)	50.047
Chembridge Diverset EXP	2015	(CHEMBRIDGE, 2020)	49.998
Chembridge EXPRESS-Pick Collection Stock	2020	(CHEMBRIDGE, 2020)	502.530
eMolecules	2017	(OPENEYE, 2020b)	12.607.334
IBS Bioactive Compounds	2020	(IBS, 2020)	789
IBS Natural Compounds	2020	(IBS, 2020)	69.034
IBS Synthetic Compounds	2020	(IBS, 2020)	486.367
Kishida	2016	(OPENEYE, 2020b)	2.431.082
MayBridge (Screening Collection)	2020	(MAYBRIDGE, 2014)	53.352
Molport Natural	2020	(MOLPORT, 2020)	113.687
NPASS	2020	(ZENG et al., 2018)	30.926
Princeton	2015	(PRINCETON, 2015)	511.340
ZINC CNS	2015	(IRWIN; SHOICHET, 2005)	377.735
ZINC Druglike	2019	(STERLING; IRWIN, 2015)	20.834.801
ZINC Naturals (Natural Products Collection)	2015	(IRWIN; SHOICHET, 2005)	113.841
			<b>&gt; 39 milhões</b>

### 3.2.1.2 Similaridade 3D pelo ROCS e EON

No **ROCS** (HAWKINS; SKILLMAN; NICHOLLS, 2007; OPENEYE, 2020d), usamos como *query* a suposta pose bioativa predita por nós para o **composto 1** e a aplicamos às dezessete bases de dados preparadas. Vale ressaltar que o uso desta pose como *query* foi validada em antemão (vide seção 3.1.7).

Assim, filtramos essas bases de dados com o **ROCS**, mantendo as 5.000 moléculas mais bem ranqueadas, de cada base, de acordo com os respectivos índices de *ROCSTC* (forma + cor). Na sequência, a triagem foi realizada por similaridade 3D de potencial eletrostático usando o *software* **EON** (OPENEYE, 2020e) – utilizando-se a mesma pose de **1** como *query*. Assim, selecionamos as 1.000 moléculas mais bem ranqueadas (de cada base) de acordo com os respectivos índices *EONTC*.

Aproveitamos aqui para apresentar, resumidamente, alguns detalhes sobre os *softwares* **ROCS** e **EON**.

## **ROCS**

A forma (*shape*) que os ligantes apresentam quando ligados ao sítio de interação da proteína-alvo (conformação bioativa) é um aspecto determinante para a atividade biológica de uma molécula, visto que o sítio de interação da proteína-alvo apresenta uma cavidade que impõe ao ligante uma determinada forma para efetuar a interação (GRANT; GALLARDO; PICKUP, 1996; NICHOLLS et al., 2010).

Neste sentido, o programa **ROCS** (HAWKINS; SKILLMAN; NICHOLLS, 2007) compara eficientemente a forma de um dado composto referência (*query*) com a forma de compostos presentes em bases de dados; por isso amplamente utilizado em *LBVS*.

Este fundamenta-se na ideia de que a comparação entre os compostos pode ser feita pela sobreposição de seus volumes. Para então se estabelecer uma correlação entre os termos volume e forma, inicia-se uma lógica pela definição de volume ( $V$ ): qualquer campo escalar, i.e., *função característica* com um único valor numérico em cada ponto do espaço:

$$V = \int f(x, y, z) dv$$

Equação 3.6

Sendo  $f$  a *função característica* para uma dada molécula, em termos de seu  $V$ , e  $x$ ,  $y$  e  $z$  são coordenadas no espaço. Considerando que o volume de um campo escalar é uma contração da informação representada por essa *função característica*, levanta-se a questão de que duas moléculas (representadas por suas diferentes *funções características*:  $f$  e  $g$ ) não podem ter a mesma forma se seus volumes não forem os mesmos. Define-se, deste modo, a similaridade por forma ( $S_1$ ):

$$S_1 = \int |f(x, y, z) - g(x, y, z)| dV$$

Equação 3.7

Se esta integral for zero, então  $f$  e  $g$  correspondem a mesma função e, portanto, a mesma forma. Quanto maior a integral, mais diferentes são as formas definidas por  $f$  e  $g$ .  $S_1$  é então sinônimo de métrica (i.e., mensura a forma). Como obedece a desigualdade triangular, permite a elaboração da métrica  $S_2$ , que elevada ao quadrado corresponde a equação fundamental para comparação por forma (Equação 3.8), e pode ser reescrita pela Equação 3.9.

$$S_2^2 = \int f(x, y, z)^2 dV + \int g(x, y, z)^2 dV - 2 \int f(x, y, z)g(x, y, z) dV$$

Equação 3.8

$$S_{f,g} = I_f + I_g - 2O_{f,g}$$

Equação 3.9

Os termos  $I$  correspondem as próprias sobreposições de volume por cada molécula, o termo  $O$  é a sobreposição entre as duas funções e, juntos, constituem os três termos necessários para comparar as formas de dois campos escalares (volumes). Contudo, nota-se que é possível derivar o coeficiente de Tanimoto (TC) por meio da recombinação dos termos:

$$TC_{f,g} = \frac{O_{f,g}}{I_f + I_g - O_{f,g}}$$

Equação 3.10

Vale mencionar que as sobreposições de moléculas, bem como as respectivas comparações por forma, foram otimizadas considerando a abordagem analítica de Gaussian. O conceito de *função característica* de modo binário foi substituído pela soma de funções contínuas, ou seja, uma gaussiana.

Após tal otimização, os valores de TC correspondem então a sobreposição exclusivamente por forma: *ShapeTanimoto*. Todavia, desenvolveu-se, em complemento, a métrica comparativa *ColorTanimoto* que lida com o alinhamento de características químicas (*color features*), visto que tal acréscimo trouxe um impacto bastante positivo para refinar a medida global de similaridade baseada em forma.

Os *color features* permitem a caracterização química dos compostos considerando: grupos aceptores ou doadores de ligação de hidrogênio, cátions ou anions, grupos alifáticos hidrofóbicos e tamanhos de anéis. A similaridade entre duas moléculas em termos dos tais *color features* são calculados por um ‘campo de força de cor’ (*Implicit Mills Dean*) que possibilita a descrição de: tipos de ‘átomos de cor’; a quais grupos funcionais os ‘átomos de cor’ devem ser aplicados; e se há interação atrativa ou repulsiva entre os ‘átomos de cores’.

Por fim, a soma dos coeficientes de similaridade *ShapeTanimoto* e *ColorTanimoto* constitui a métrica *ROCS\_TanimotoCombo* (*ROCSTC*). Esta varia de 0 a 2, sendo que valores mais próximos de 2 retratam uma alta sobreposição, de modo global (*shape + color features*), entre as conformações de uma determinada molécula e o *query* utilizado.

## **EON**

Em sequência ao **ROCS**, outro programa que pode ser empregado como filtro *LBVS* é o programa **EON**, que faz comparações entre compostos levando em conta os seus mapas de potencial eletrostático (TRESADERN; BEMPORAD; HOWE, 2009; HAWKINS; STAHL, 2018).

O seu funcionamento também requer o uso de um *query* e, em seguida, os demais compostos da base de dados são comparados e classificados com base nos respectivos mapas de potencial eletrostático. O **EON** usa possui um desenvolvimento lógico e de equações análogo ao do **ROCS**. Aplica, ainda, uma medida de Tanimoto baseada em campo de força para comparar o potencial eletrostático de duas pequenas moléculas (OPENEYE, 2023b).

O potencial eletrostático das moléculas é calculado por meio do *toolkit* OpenEye **Zap**, que resolve a equação de Poisson-Boltzmann (PB). Para tanto, calculam-se as cargas parciais utilizando o campo de força *MMFF94* e considera-se o pH neutro para as moléculas em análise.

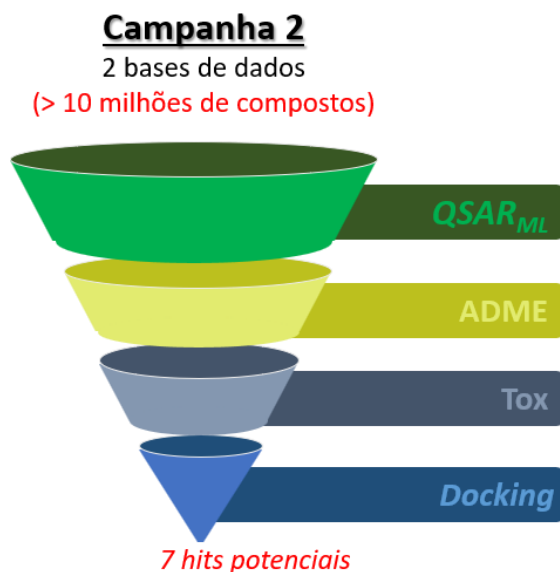
No **EON**, a principal medida eletrostática de Tanimoto (ET) consiste da *ET\_pb* que aplica um potencial dielétrico externo a partir dos cálculos de PB. Tal medida é usada para classificar/ranquear as moléculas, variando de -1/3 e 1. O valor de -1/3 corresponde a moléculas com potenciais eletrostáticos similares, porém opostos (em termos de cargas/polaridade), enquanto que o valor de 1 corresponde a moléculas com o mesmo potencial eletrostático. Por fim, também é reportado um *score* similar ao *ROCSTC* denominado *ET\_combo* (*EONTC*), que consiste da soma do *ShapeTanimoto* calculado pelo **ROCS** com o *ET\_pb* calculado pelo **EON**.

### **3.2.2 Campanha 2: Triagem baseada em QSAR via aprendizagem de máquina e em estrutura**

Nesta segunda campanha de triagem, ênfase foi dada a aplicação de modelos de QSAR via aprendizagem de máquina. Após o desenvolvimento e avaliação dos tais modelos, o mais robusto e preditivo foi selecionado com base em suas características estatísticas.

Em seguida, aplicamos o modelo preditivo de QSAR selecionado, para filtrar apenas os compostos das bases de dados que foram preditos como ativos e dentro do domínio de aplicabilidade. Assim, os compostos resultantes foram classificados em

ordem decrescente de acordo com os respectivos índices de predições de confiança/pontuação (QSAR score), priorizando os 40.000 melhores.



**Figura 3.2** - Campanha 2 de triagem virtual exibindo as metodologias aplicadas para a obtenção de potenciais *hits* como moduladores alostéricos GSK-3 $\beta$ .

Na sequência, estes foram submetidos às predições ADME/Tox utilizando **QikProp** e **DEREK** e filtrados considerando os critérios citados no tópico 3.2.3.

Por fim, os compostos sobreviventes foram filtrados por meio do mesmo protocolo de *docking* usando o **Glide** (conforme descrito no tópico 3.2.4), seguido de inspeção visual das correspondentes interações intermoleculares no sítio alostérico 7 da GSK-3 $\beta$ .

### 3.2.2.1 Bases de dados (Campanha 2)

**Tabela 3.3** - Duas bases de dados utilizadas na Campanha 2 de triagem virtual.

Nome	Versão/Ano	Referência	Número de moléculas
Molport Screening	2020	(MOLPORT, 2020)	7.500.903
Enamine Screening Collection	2020	(ENAMINE, 2021)	2.906.530
			<b>&gt; 10 milhões</b>

Com vista a aplicação dos modelos de QSAR em triagem virtual, duas bases de dados totalizando mais de 10 milhões de compostos foram usadas: *Molport Screening* (MOLPORT, 2020) e *Enamine Screening Collection* (ENAMINE, 2021) – vide **Tabela 3.3**. Para prepará-las, tais bases foram submetidas exatamente ao mesmo procedimento de curagem de dados (descrito como etapa prévia à geração dos modelos de QSAR).

### 3.2.2.2 Modelagem de QSAR

Os modelos de QSAR foram desenvolvidos e rigorosamente validados seguindo as melhores práticas da área (FOURCHES; MURATOV; TROPSHA, 2010; CHERKASOV et al., 2014; MURATOV et al., 2020). Para tanto, utilizou-se, inicialmente, o *software* **KNIME** 4.6.0 (BERTHOLD et al., 2008; KNIME AG, 2021). Posteriormente os modelos foram também gerados usando linguagem de programação no **Python** 3.6 (<https://www.python.org>) - em especial para aplicá-los em triagem virtual. A seguir, subdividimos as etapas sequenciais envolvidas na modelagem de QSAR.

#### **Compilação e curagem da série de dados/compostos**

Para desenvolver os modelos de QSAR utilizamos o mesmo *dataset* de 88 compostos (vide **Tabela 3.1**), os quais apresentam modulação alostérica de GSK-3 $\beta$  conhecida experimentalmente e expressas por valores de IC<sub>50</sub> (ZHANG et al., 2013, 2014; BROGI et al., 2017; PALOMO et al., 2017). Foi considerado um valor de corte de IC<sub>50</sub> de 20  $\mu$ M para classificá-los em ativos/inativos, obtendo assim 37 ativos e 51 inativos. Ressalta-se, portanto, que os modelos gerados foram classificatórios (ou binários).

Conforme protocolo estabelecido pelo grupo do Prof. Dr. Alexander Tropsha (*UNC - Chapel Hill*) (FOURCHES; MURATOV; TROPSHA, 2010, 2016), realizamos a curagem dos dados, padronizando as estruturas químicas dos compostos e as correspondentes informações biológicas usando um *workflow* padrão de **KNIME**, integrado com ChemAxon **Standardizer** v.20.8.0 (ChemAxon, Budapeste, Hungria, <http://www.chemaxon.com>). Neste, executa-se a normalização de grupos químicos específicos como anéis aromáticos e grupos nitro, assim como a remoção de contraíons, metais, sais, compostos inorgânicos, organometálicos, polímeros e misturas.

Em adição, compostos idênticos reportados mais de uma vez na série de dados, i.e., duplicatas, são identificadas e analisadas cuidadosamente, visando, em geral, a remoção das mesmas (com manutenção do dado/replicata mais representativa). Isso, é feito para evitar a superestimação da preditividade dos modelos de QSAR. Além disso, também possibilita estimar a qualidade dos dados, ou seja, se os valores de atividade para um mesmo composto são convergentes, estima-

se uma alta qualidade dos dados; enquanto que, se os valores são divergentes, estima-se uma baixa qualidade dos dados.

### **Descritores moleculares**

Os modelos de QSAR foram desenvolvidos empregando dois conjuntos de descritores moleculares: *Morgan fingerprints* (MORGAN, 1965; FIGUERAS, 1993; ROGERS; HAHN, 2010) e propriedades *RDKit* (<http://www.rdkit.org>) (RINIKER; LANDRUM, 2013; RDKIT: OPEN-SOURCE CHEMINFORMATICS, 2021).

Os descritores *Morgan fingerprints* foram gerados com 2048 *bits* e um raio atômico de 3, usando o *node* do *RDKit* implementado no **KNIME** e também no **Python**. Os descritores/propriedades *RDKit* foram calculados usando o **KNIME** e **Python**, sendo que estes correspondem a 117 propriedades físico-químicas e estruturais, e.g.: SlogP, SMR, LabuteASA, TPSA, contagens relevantes, dentre outras. Enfatiza-se que as definições detalhadas para os descritores *RDKit* constam na página <https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>.

É importante notar que, em antemão a geração dos modelos, as propriedades *RDKit* foram selecionadas usando filtros de baixa variância e correlação; além disso, a importância (peso) de cada descritor foi calculada após a geração dos modelos. Avaliamos sistematicamente o desempenho de cada modelo construído, utilizando cada conjunto de descritores, bem como as suas combinações.

### **Algoritmo para construção dos modelos de QSAR**

Utilizou-se o algoritmo de aprendizagem de máquina *Random Forest* (BREIMAN, 2001) implementado na biblioteca *scikit-learn* do **Python** (<http://scikit-learn.org>) (PEDREGOSA FABIANPEDREGOSA et al., 2011). Também, no **KNIME**, empregando-se os *nodes* 'Tree Ensemble Learner' e 'Tree Ensemble Predictor', para os quais foram utilizadas as seguintes configurações: Índice de Gini como critério de divisão e 750 árvores de decisão para construir cada floresta.

Em suma, o *Random Forest (RF)* é um algoritmo baseado na construção de um conjunto de diferentes árvores de decisão, cada uma delas baseada em um subconjunto aleatório do conjunto de dados original. As predições de cada árvore de decisão são combinadas para formar a predição final. Vários estudos de sucesso envolvendo modelagem de QSAR empregando *RF* foram realizados anteriormente na literatura (SVETNIK et al., 2003; ALVES et al., 2015; MELO-FILHO et al., 2016).



### **Validação dos modelos de QSAR**

A validação dos modelos de QSAR foi feita empregando-se o procedimento de validação cruzada externa de *5-fold* (*5-fold external cross-validation*). Tal procedimento foi realizado usando o seguinte protocolo: a série completa de compostos com atividade experimental conhecida é dividida aleatoriamente em cinco subconjuntos de tamanhos iguais. Um desses subconjuntos (20 % de todos os compostos) é separado como a série de validação externa (série de teste), enquanto os quatro subconjuntos restantes (80 % de todos os compostos) formam a série de modelagem (série de treinamento). Este procedimento é repetido cinco vezes, permitindo que cada um dos cinco subconjuntos seja usado como série de teste. Cada modelo é construído usando apenas a série de treinamento e é fundamental enfatizar que os compostos nunca fazem parte simultaneamente de ambas as séries. Desta forma, as seguintes características estatísticas são atribuídas aos modelos gerados para avaliar as respectivas preditividades:

$$TCC = \frac{(\textit{sensibilidade} - \textit{especificidade})}{2} \quad \text{Equação 3.11}$$

$$\textit{Sensibilidade} = \frac{VP}{VP + FN} \quad \text{Equação 3.12}$$

$$\textit{Especificidade} = \frac{VN}{VN + FP} \quad \text{Equação 3.13}$$

$$VPP = \frac{VP}{VP + FP} \quad \text{Equação 3.14}$$

$$VPN = \frac{VN}{VN + FN} \quad \text{Equação 3.15}$$

Em que, TCC é a taxa de classificação correta (*CCR, correct classification rate*), VPP é o valor de preditividade positiva (*PPV, positive predictive values*) e VPN é o valor de preditividade negativa (*NPV, negative predictive values*). VP é o número de verdadeiros positivos, VN é o número de verdadeiros negativos, FP é o número de falsos positivos e FN é o número de falsos negativos.

Em adição, vinte rodadas de aleatorização de Y (*Y-randomization*) foram realizadas para a série de dados, garantindo assim que o desempenho de cada modelo não fosse devido a correlações ao acaso. Esta validação adicional consiste

de gerar os modelos de QSAR após aleatorizar a variável Y (no caso a variável dependente / a atividade biológica / valores de IC<sub>50</sub>). Caso este procedimento resulte em modelos com características estatísticas melhores do que a do modelo em questão, o mesmo pode ser desprezado por possível correlação ao acaso ou por incompatibilidade dos descritores moleculares para descrever a atividade (KUZ'MIN; ARTEMENKO; MURATOV, 2008).

O Domínio de Aplicabilidade (*APD*, *applicability domain*) de cada modelo foi definido, com vista à sinalização de compostos (da série de teste) para os quais as predições podem ser não confiáveis. Para isso, o *APD* foi estimado pelo método *z-cutoff* (ZHANG et al., 2006; TROPSHA; GOLBRAIKH, 2007), que considera medidas de similaridade com base nas distâncias euclidianas entre todos os compostos de treinamento e os compostos de teste. A distância de um composto de teste para seu vizinho mais próximo, na série de treinamento, é comparada ao limite predefinido de *APD*. Se a similaridade estiver além desse limite, a predição é considerada não confiável. Neste método, o *APD* é expresso conforme a Equação 3.16:

$$APD = \langle D \rangle + Z\sigma$$

Equação 3.16

para a qual, calcula-se os valores de  $\langle D \rangle$  e de  $\sigma$  da seguinte maneira: 1º) é calculada a média das distâncias euclidianas entre todos os pares de compostos de treinamento; 2º) formula-se o conjunto de distâncias que ficaram abaixo da média; 3º)  $\langle D \rangle$  e  $\sigma$  são finalmente calculados como a média e o desvio padrão de todas as distâncias incluídas neste conjunto. Z é um valor de corte empírico e o valor padrão é de 0,5 (AFANTITIS et al., 2011).

### 3.2.3 Predições ADME/Tox

Para prever propriedades físico-químicas/farmacocinéticas e toxicológicas (ADME/Tox) dos compostos usamos os *softwares* **QikProp** e **DEREK**, respectivamente, em ambas Campanhas 1 e 2 de triagem virtual.

Os critérios adotados no **QikProp** para filtrar os compostos mais promissores foram estabelecidos considerando propriedades fármaco similar (*drug-like*), com ênfase àquelas que se assemelham a drogas atuantes no SNC (WAGER et al., 2010; GHOSE et al., 2012; RANKOVIC, 2017). Fizemos, entretanto, sutis adaptações, ficando os critérios delineados pelas seguintes propriedades: MW ≤ 360, PSA ≤ 90 Å<sup>2</sup>,

$^{QP}\log P_{ow} = -2.0 - 6.5$ ;  $^{QP}\log BB > -0,5$ ; número de heteroátomos  $\leq 21$ ; absorção oral humana  $> 80\%$ ;  $^{QP}P_{Caco} > 500$  nm/s;  $^{QP}P_{MDCK} > 500$  nm/s.

Para a predição de *endpoints* (ou parâmetros, ou tipos) de toxicidade usando o **DEREK**, atenção especial foi dada aos seguintes: carcinogenicidade, genotoxicidade, neurotoxicidade, cardiotoxicidade e hepatotoxicidade. Estes foram avaliados em ambas espécies de mamíferos e bactérias, e foram refeitos quando os respectivos alertas foram disparados como plausíveis, prováveis ou certos.

### 3.2.4 Simulações de *docking*

Em ambas as Campanhas 1 e 2 de triagem virtual, conduzimos simulações de *docking* ao final, para avaliar as poses/interações dos compostos e, assim, selecioná-los como *hits*. Tais simulações correspondem ao protocolo de *docking* desenvolvido em duas etapas, conforme descrito na *PARTE I* - sendo uma preliminar (seção 4.1.2) e outra de refinamento (seção 4.1.4). Vale notar que a validação deste protocolo também foi rigorosamente conduzida, conforme seção 4.1.5. Os preparos dos ligantes e da proteína, em anemão as execuções de *docking*, estão também descritos anteriormente (seção 3.1.2).

Recapitulando, brevemente, as simulações de *docking* aqui conduzidas se deram por meio do uso do *software* **Glide**, usando a função de *score* *Glide XP*, não permitindo inversão de N dos ligantes, mantendo o resíduo Ser236 flexível e usando um *grid* de 10x10x10 Å centralizado nas coordenadas ( $x = 11,78$ ,  $y = 13,15$ ,  $z = 38,33$ ) do sítio alostérico 7 da GSK-3 $\beta$  (código PDB 1PYX).

Por fim, todas as figuras de *docking* foram preparadas usando o *software* **Maestro** (SCHRÖDINGER, 2021) e/ou **Pymol** (SCHRÖDINGER LLC., 2010).

### 3.2.5 Ensaios de atividade biológica com a GSK-3 $\beta$

Os ensaios de atividade biológica foram realizados no *International Centre for Kinase Profiling, University of Dundee*. Tais ensaios *in vitro* foram realizados, em colaboração comercial com o referido centro de pesquisa, seguindo-se os métodos descritos por Davies et al. (2000) (DAVIES et al., 2000).

Em resumo, eles consistiram de ensaio de ligação de filtro radiométrico usando [ $\gamma$ -<sup>33</sup>P]ATP (concentração de aproximadamente  $K_m$ , 5  $\mu$ M), substrato Phospho GS2 YRRAAVPPSPSLSRHSSPHQS\*EDEEE (20  $\mu$ M). Além disso, foram realizados

ensaios com a enzima GSK-3 $\beta$  em Mops 8 mM, pH 7,0, contendo EDTA 0,2 mM e concentração dos compostos de 100  $\mu$ M. Todos os ensaios foram realizados em duplicata, por 40 min à temperatura ambiente em incubações de 25  $\mu$ L.

### 3.2.6 Outras análises computacionais

Os gráficos de radar de similaridade foram construídos considerando os respectivos cálculos dos *fingerprints* moleculares *MACCS* e *Morgan*, computados pelo *node* do *RDKit* implementado no **KNIME**.

O estudo de *blind docking* foi feito usando o *software* **AutoDock** (MORRIS et al., 2009; THE SCRIPPS RESEARCH INSTITUTE, 2020) implementado no **PyRx** (DALLAKYAN; OLSON, 2015). A fim de reproduzir um exemplo de estudo de *blind docking* previamente reportado na literatura (PALOMO et al., 2011) – que é corroborado pelas nossas investigações envolvendo detecção de cavidade e mapeamento de superfície da GSK-3 $\beta$  (conforme apresentado na *PARTE I*) – os seguintes parâmetros foram adotados: centróide determinado pelas coordenadas  $x = 26,7110$ ,  $y = 5,9133$ ,  $z = 38,2832$ , *grid* de  $193 \times 166 \times 167$  Å, com espaçamento de  $0,375$  Å e algoritmo genético (GA) Lamarckiano (número de corridas de GA 200, tamanho da população 200, número máximo de *evals* 2.500.000). As poses geradas para cada composto foram inspecionadas manualmente e agrupadas de acordo com sua aparente sobreposição e/ou interações dentro dos sítios ortostéricos e alostéricos. A pose de melhor *score* foi então selecionada a partir de cada grupo de poses, em cada sítio.

A pose de melhor *score* obtida para cada composto, no sítio ortostérico e alostérico, foi então submetida à avaliação da cobertura/ocupação de sítio (*pocket coverage*, *PC*) usando o *software* **DoGSiteScorer** (VOLKAMER et al., 2010, 2012).

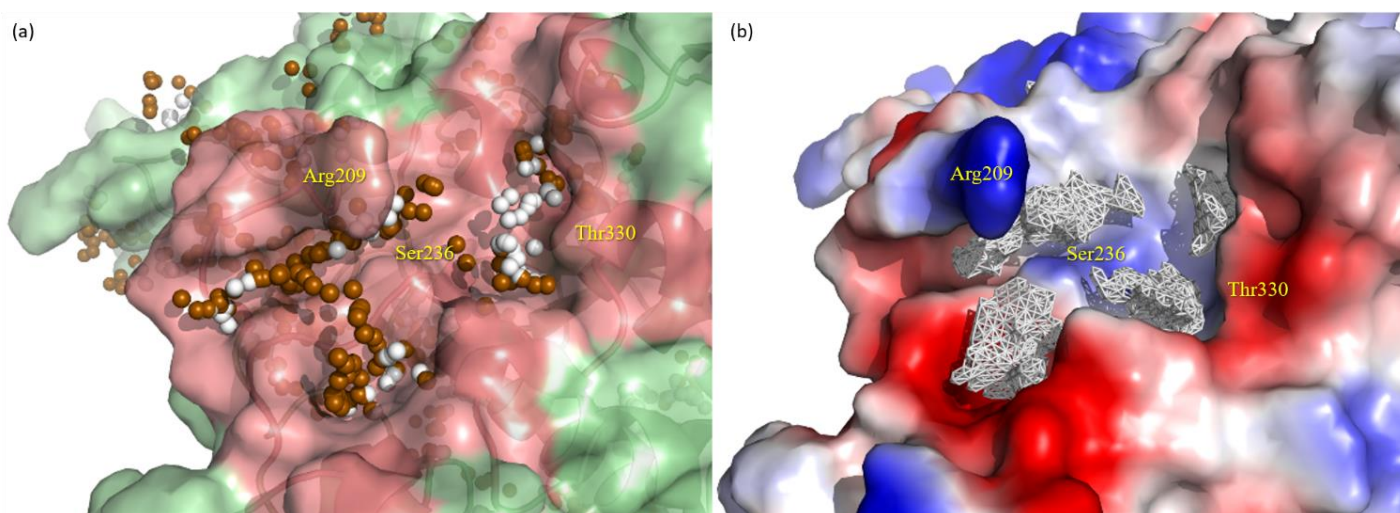
## 4 RESULTADOS E DISCUSSÃO

### 4.1 ***PARTE I***: Reforçando a proposta de cavidades e de poses bioativas para os inibidores alostéricos de GSK-3 $\beta$

#### 4.1.1 Detecção de cavidades e predição do potencial dos sítios alostéricos

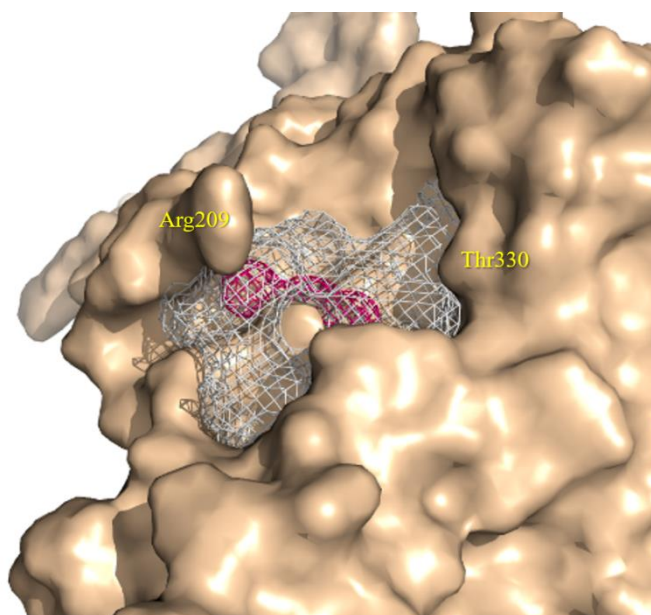
Para explorar e detectar os sítios presentes na superfície da GSK-3 $\beta$ , cinco diferentes *softwares/webserver*s foram empregados, conforme detalhado nos Materiais e Métodos - *PARTE I*. A análise das cavidades detectadas na estrutura da GSK-3 $\beta$  (1PYX.pdb) *apo* (ou seja, sem ligante) foi feita de modo visual, com vista as características estruturais do **composto 1** e com base na pose previamente proposta na literatura para este composto no sítio de interesse (PALOMO et al., 2017).

Os *softwares* **fpocket** e **Superstar** permitiram a visualização do formato das cavidades da GSK-3 $\beta$ , aparentemente indicando uma forma mais compatível da cavidade alostérica 7 com o **composto 1** (**Figura 4.1**). O **fpocket**, além de definir adequadamente o sítio compatível com **1**, forneceu esferas de atributos que caracterizam a ocupação de regiões na cavidade por grupos polares (em marrom) e por grupos apolares (em branco). Assim, foi parcialmente indicada uma preferência pela cadeia apolar/hidrofóbica de **1** quando acomodada próxima ao resíduo Thr330 e o correspondente bolsão hidrofóbico.



**Figura 4.1** - (a) Cavidade detectada pelo **fpocket** e representada por esferas marrons (grupos polares) e brancas (grupos apolares), com a enzima representada nas formas de Superfície e *Cartoon*; (b) Forma da cavidade predita pelo **Superstar** em *mesh* branco, com o mapa de potencial eletrostático da enzima mapeado em sua superfície; imagens produzidas no **Pymol**, representando a cavidade alostérica 7 da GSK-3 $\beta$  (1PYX.pdb).

O **metaPocket** mapeou a cavidade 7 da GSK-3 $\beta$  em consenso por meio de três metodologias diferentes: **LigsiteCS**, **fpocket** e **Ghecom**. Na **Figura 4.2**, em malha branca observa-se o sítio agrupado por estas três metodologias, considerando as respectivas similaridades espaciais previstas. Em malha rosa, observa-se o centro de massa deste sítio agrupado, indicando os resíduos próximos com maior potencial de interação. Além disso, o **metaPocket** computou um valor de *z-score* de 3,25 para a cavidade 7, apenas menor do que o valor de *z-score* de 22,94 computado para o sítio catalítico; indicando, assim, a alta propensão de a cavidade 7 atuar como um sítio alostérico apto para interagir e acomodar ligantes, de modo geral.

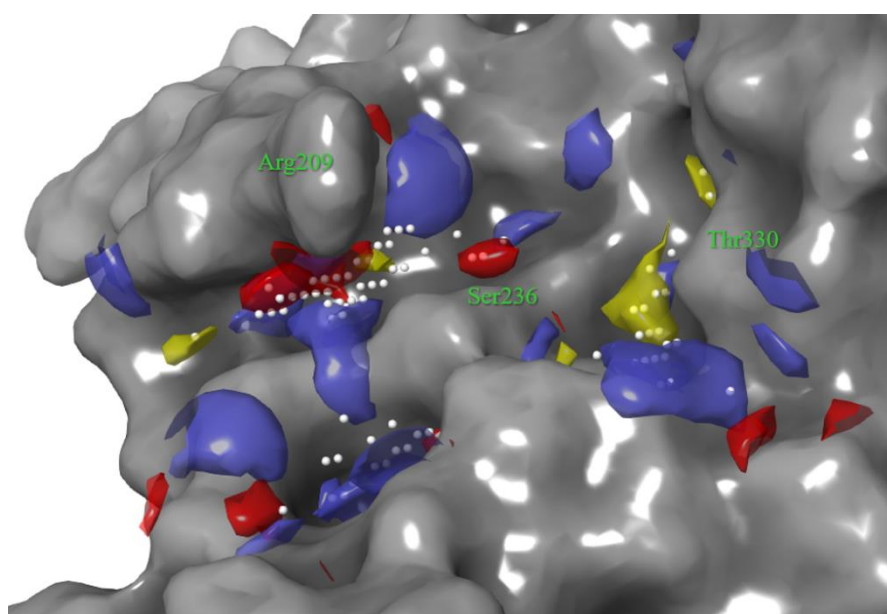


**Figura 4.2** - Cavidade melhor pontuada pelo **metaPocket** e representada em *mesh* branco (*cluster* de sítios indicados em consenso por 3 metodologias) e esferas rosas (centro de massa do respectivo sítio global), com a enzima representada na forma de Superfície.

O **Sitemap** indicou regiões mais representativas na cavidade 7 da GSK-3 $\beta$ , ilustradas pelas esferas brancas (**Figura 4.3**). Além disso, foi capaz de atribuir a segunda maior pontuação de *druggability* de 0,898 para tal cavidade (também apenas inferior ao valor de 1,021 atribuído ao sítio catalítico). As demais cores mapeadas por este *software* indicam padrões farmacofóricos da cavidade que serão discutidos posteriormente (vide a frente **Figura 4.8**).

O **PARS** avalia o volume, a flexibilidade e a conservação estrutural dos resíduos de aminoácidos que constituem a cavidade e, diferente dos demais, é capaz de indicar o potencial de esta ser alostérica. Este *webserver* indicou três esferas independentes que compõem a cavidade 7 da GSK-3 $\beta$  (**Figura 4.4**). A primeira esfera

amarela detectada abaixo do resíduo de Arg209 indica a região de maior volume e foi ranqueada em primeiro lugar como sítio alostérico putativo, por ter apresentado *p-value* de 0,71 e conservação estrutural de 12,60%. A segunda esfera amarela, segunda ranqueada, apresentou *p-value* de 0,79 e conservação estrutural de 25,80%, enquanto a terceira esfera azul clara apresentou *p-value* de 0,47 e conservação estrutural de 56,90%. As três esferas juntas (especialmente a primeira e terceira) devem indicar razoavelmente que a cavidade 7 representa o sítio alostérico mais promissor da GSK-3 $\beta$ , em vista da análise global de sua estrutura bem como em comparação com demais cavidades detectadas em sua superfície.

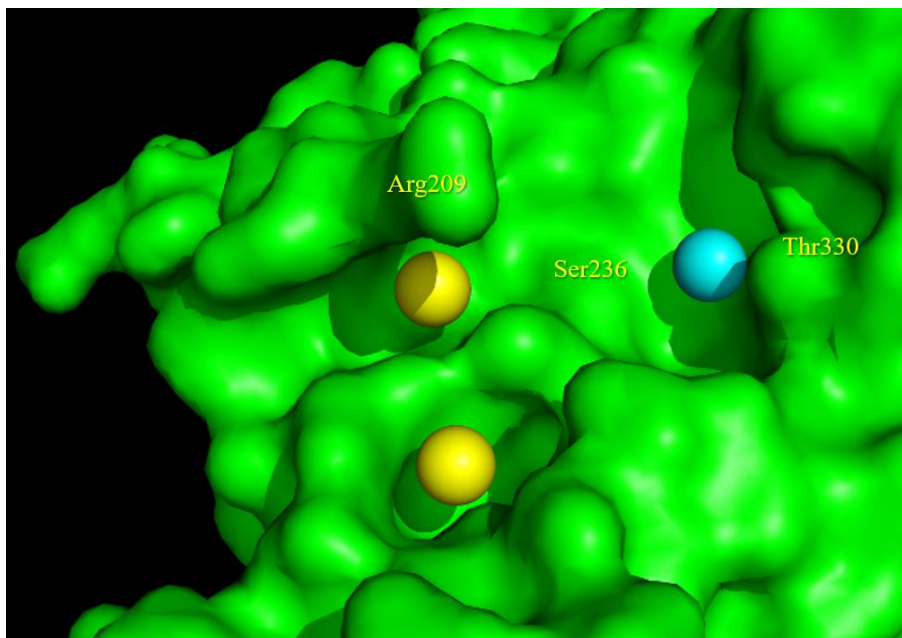


**Figura 4.3** - Cavidade detectada pelo **Sitemap** representada por esferas brancas (demais mapas contornados são discutidos na [seção 4.1.3](#)), com a enzima representada na forma de Superfície.

Vale ressaltar que Palomo et al. (PALOMO et al., 2011) realizaram uma comparação entre algumas estruturas da GSK-3 $\beta$  encontradas no PDB, utilizando apenas o *software* **fpocket** para detectar a presença consensual das 7 cavidades que foram encontradas nestas diferentes estruturas – das quais as cavidades 4, 5, 6 e 7 mostraram ser alostéricas em potencial. Neste trabalho, de forma complementar, avaliamos dentre as cavidades anteriormente apontadas como alostéricas, qual teria o maior potencial e probabilidade em interagir com inibidores/ligantes (principalmente com base na estrutura química do **composto 1**).

Nossa abordagem, começou por analisar todas as possíveis cavidades ('rasas' ou 'enterradas') detectadas na superfície do GSK-3 $\beta$ . Usando cinco *softwares/webservers* diferentes, todos indicaram em consenso que a cavidade 7

apresenta uma maior probabilidade de representar um sítio alostérico, bem como valores de *score* razoáveis de *druggability* - apenas menores do que os valores constatados para o sítio ativo da GSK-3 $\beta$ .



**Figura 4.4** - Cavidade gerada pelo *webserver* PARS indicando a cavidade 7 com maior potencial de representar um sítio alostérico na GSK-3 $\beta$  (1PYX.pdb), com a enzima representada na forma de Superfície.

Desta forma, considerando estes resultados obtidos pelos *softwares* de detecção de cavidades e de predição de seus potenciais, entre as 4 cavidades alostéricas descritas anteriormente, a cavidade alostérica 7 é, de fato, potencialmente indicada como uma cavidade mais susceptível a interagir com uma determinada molécula 'órfã', e.g. com um inibidor não competitivo de ATP desta enzima.

#### 4.1.2 Estudo preliminar de *docking*: perspectiva da cavidade

Um estudo preliminar de *docking* foi realizado para detectar qual cavidade alostérica da GSK-3 $\beta$  poderia apresentar maior afinidade com o **composto 1**. Para isso, foram usados quatro *softwares* de *docking* (**GOLD**, **Glide**, **AutoDock** e **FRED**) para analisar a afinidade com **1**, em termos de interações e os respectivos valores de *score*, individualmente, em cada uma das 4 cavidades alostéricas (**Figura 4.5**).

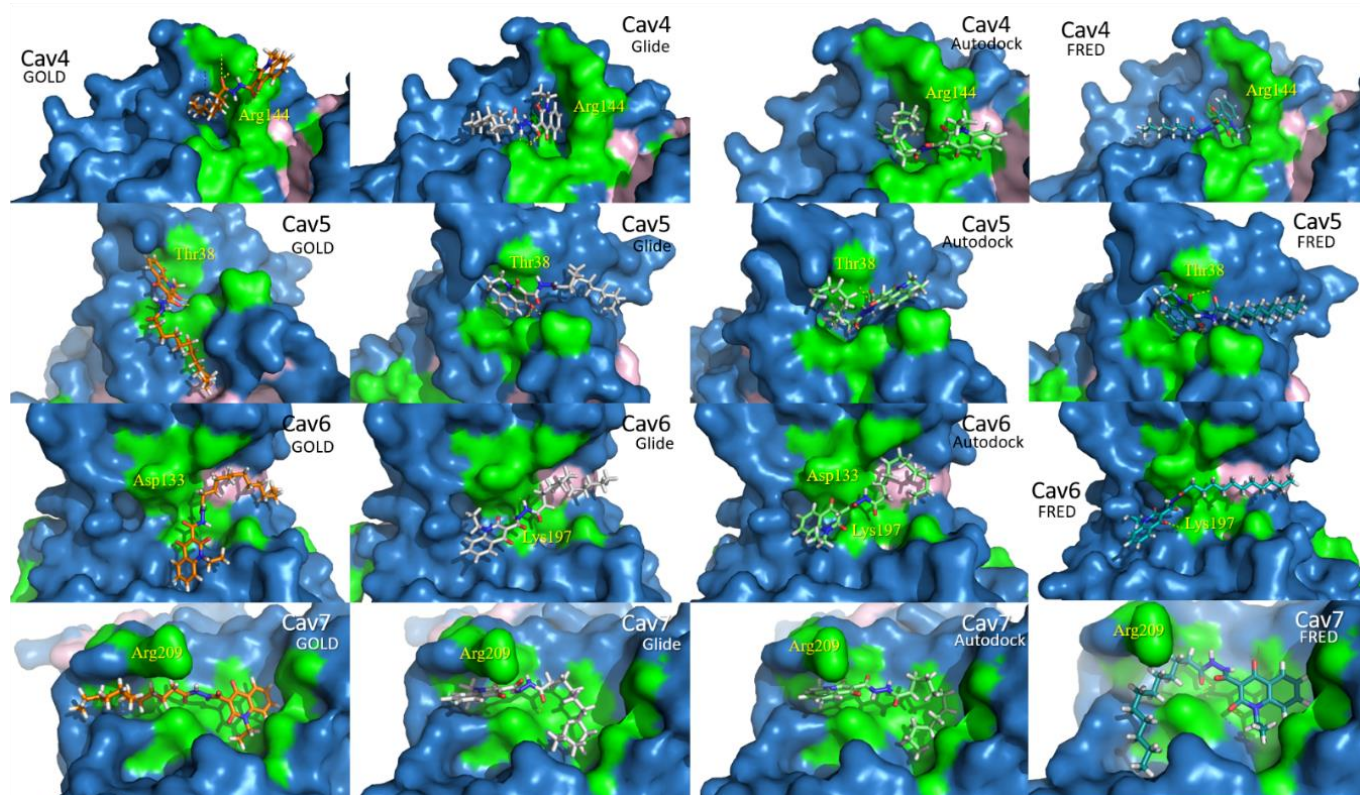
Em consenso, os quatro *softwares* de *docking* indicaram um valor de *score* mais favorável para **1** quando interagindo com a cavidade 7 da GSK-3 $\beta$ , ao invés das outras 3 cavidades alostéricas putativas, como pode ser visto na **Tabela 4.1**.



No entanto, na **Figura 4.5** pode-se observar que as poses mais bem pontuadas, obtidas por cada *software*, não apresentam homogeneidade em relação às orientações/conformações dos demais *softwares* utilizados. Em outras palavras, os *softwares* **GOLD** e **FRED** forneceram poses de **1** na cavidade 7 acomodando a respectiva cadeia hidrofóbica sob o resíduo de Arg209, enquanto que os *softwares* **Glide** e **Autodock** acomodaram o anel quinolônico nesta mesma região.

**Tabela 4.1** - Valores de *score* obtidos por *docking* pelos programas **GOLD**, **Glide**, **Autodock** e **FRED**, para o **composto 1** nas quatro cavidades alostéricas (4, 5, 6 e 7) da GSK-3 $\beta$  (1PYX.pdb).

GOLD - CHEMPLP				Glide XP (kcal/mol)				Autodock score (kcal/mol)				FRED - Chemgauss4			
Cavidade				Cavidade				Cavidade				Cavidade			
4	5	6	7	4	5	6	7	4	5	6	7	4	5	6	7
54,88	71,20	56,07	88,36	-4,47	-5,85	-2,92	-6,13	-2,45	-4,12	-3,32	-6,98	-3,99	-3,46	-4,36	-5,53



**Figura 4.5** - Representação das poses de *docking* obtidas para o **composto 1**, nas cavidades alostéricas 4, 5, 6 e 7 da proteína GSK-3 $\beta$  (1PYX.pdb), usando os programas **GOLD**, **Glide**, **Autodock** e **FRED**, com a enzima representada na forma de Superfície e a cavidade destacada na cor verde.

Vale ressaltar que o **composto 1** apresenta uma alta flexibilidade devido à presença de 14 ligações rotacionáveis, em especial por sua cadeia alifática C<sub>11</sub>H<sub>23</sub> que permite que a estrutura adote muitas conformações. Isso foi verificado pela baixa homogeneidade de poses obtidas em cada *software*, independentemente do algoritmo

de busca conformacional empregado por cada *software* de *docking* (**Glide** e **FRED** = busca sistemática, enquanto **GOLD** e **Autodock** = pesquisa estocástica).

O fator diferencial para as predições das poses, portanto, poderia ser apontado em direção as diferentes funções de *score* implementadas nos *softwares*. A função de *score* do **GOLD CHEMPLP** pode corresponder em alguns aspectos e características com a função de *score* do **FRED Chemgauss4**.

Podemos especular que existem múltiplos potenciais lineares para modelar parâmetros de forças de van der Waals e de forças repulsivas incluídos no **CHEMPLP** (além dos parâmetros relacionados as interações via ligações de hidrogênio (VERDONK et al., 2003)) que podem funcionar de forma semelhante para modelar os parâmetros de interações implementados no **Chemgauss4**, que são baseados em distâncias medidas por raios de van der Waals entre modelos atômicos de átomos pesados (OPENEYE, 2020b). Isso, mas não apenas isso, poderia indicar porque ambos os programas forneceram a pose do **composto 1** invertida, em relação ao **Glide** e **Autodock**, cujas funções de *score* são mais conhecidas por suas características de base empírica (FERREIRA et al., 2015).

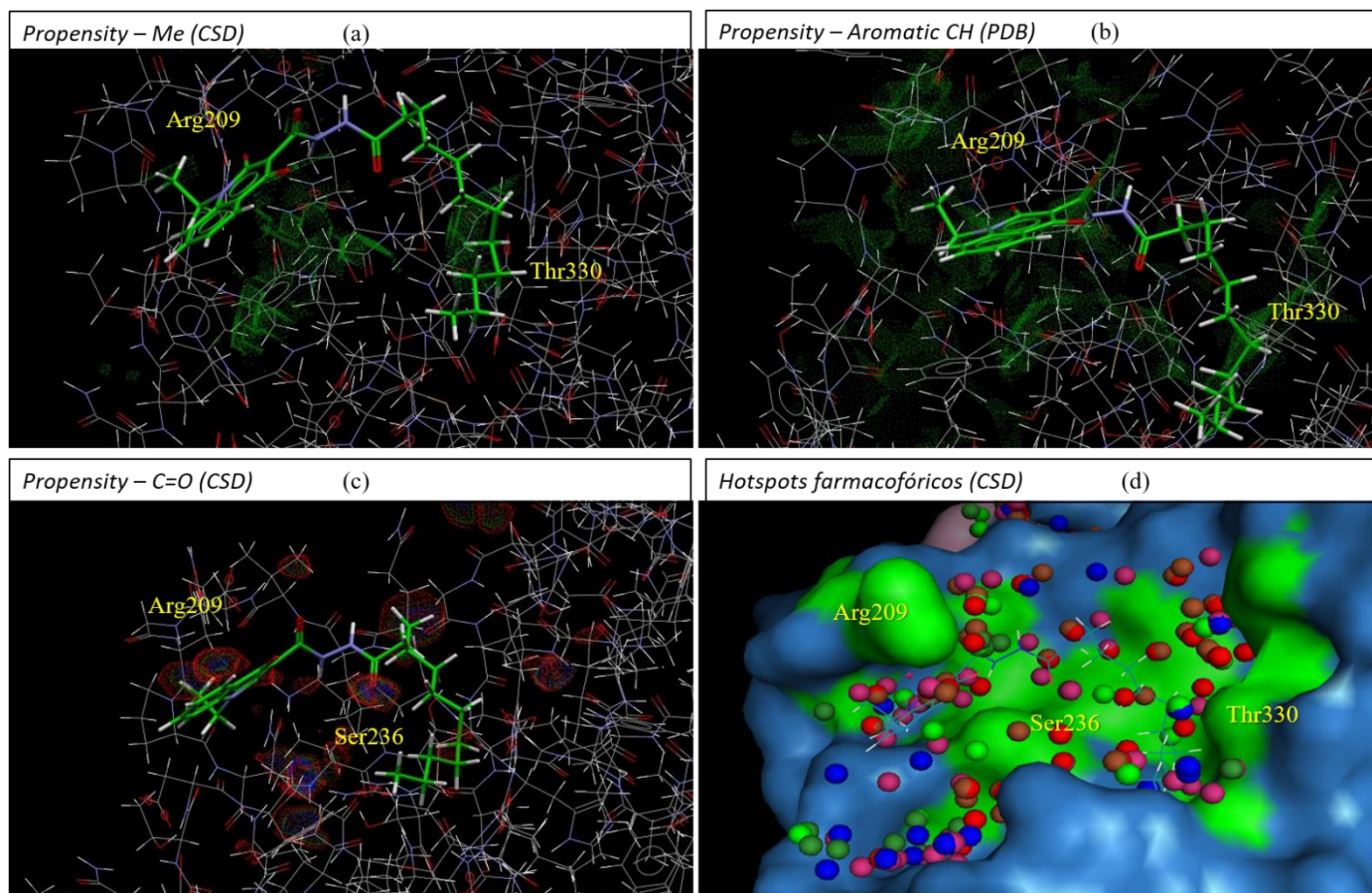
De modo geral, as funções de *score* dos programas não conseguem modelar adequadamente determinadas interações que poderiam ocorrer, e.g., transferência de carga entre a Arg209 (protonada em pH fisiológico) e o anel quinolônico do(s) composto(s), uma vez que são baseadas em modelos de Mecânica Molecular. No entanto, reconhecemos que rastrear quais fatores envolvidos nas funções de *score* dos *softwares* de *docking* que poderiam ser responsáveis por estes resultados diferentes de predição de pose, não deve ser algo tão óbvio, e demais testes sistemáticos poderiam ser necessários para maiores esclarecimentos.

Contudo, com base nos resultados observados neste estudo preliminar, não foi possível concluir qual é a pose/orientação mais representativa (suposta pose bioativa) de **1** na cavidade 7 da GSK-3 $\beta$ . Por isso, na próxima seção 4.1.3 descrevemos a aplicação de outra metodologia computacional complementar (mapas de superfície/contorno) para obter maiores esclarecimentos.

#### 4.1.3 Avaliando a pose de **1** dentro da cavidade alostérica 7 da GSK-3 $\beta$

Três *softwares* foram usados para mapear áreas de contorno e superfície, bem como para gerar *hotspots* farmacofóricos, com o objetivo de se esclarecer a pose e

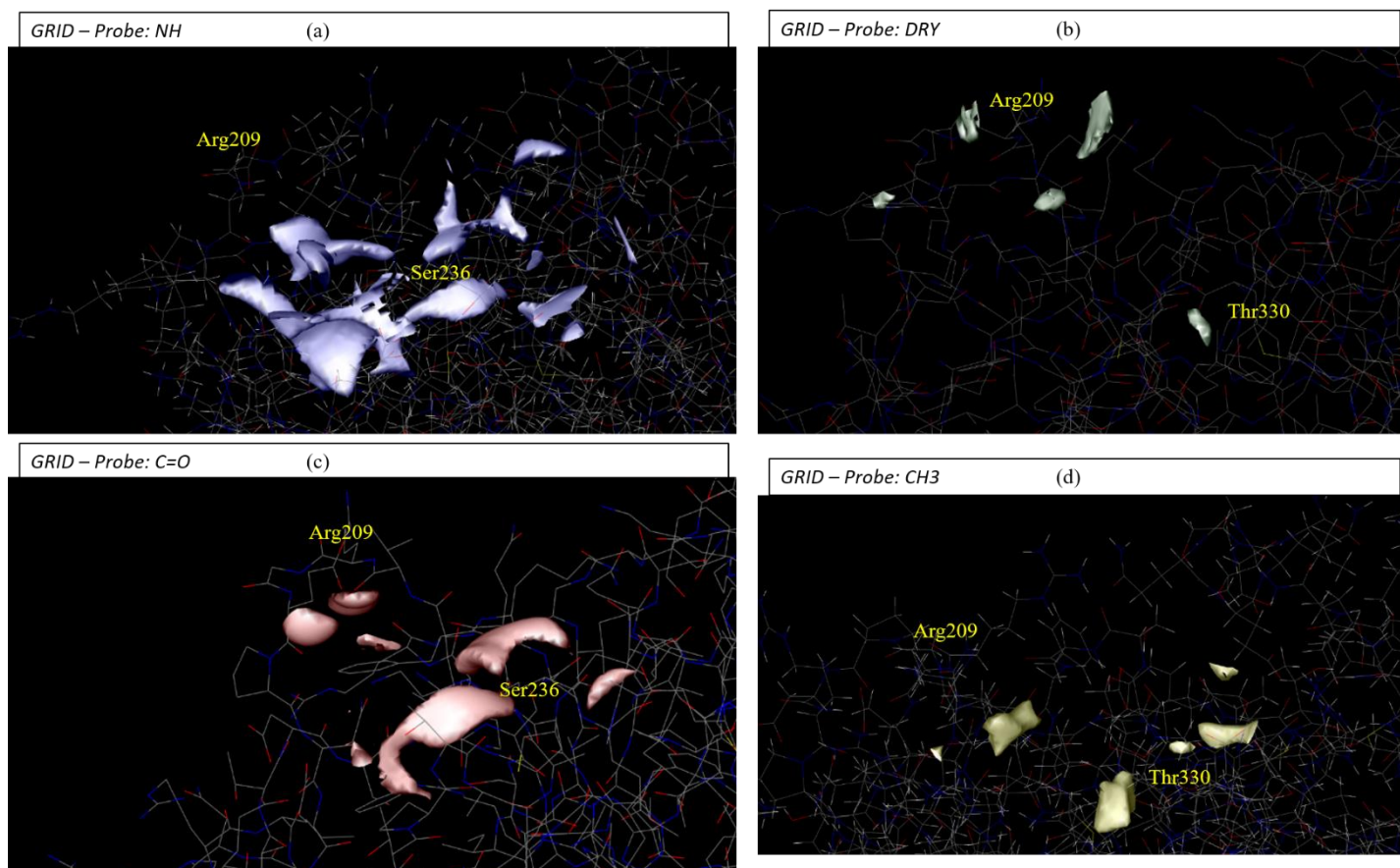
orientação do **composto 1** de forma mais coerente quando interagindo na cavidade alostérica 7 da GSK-3 $\beta$ . Os *softwares* usados foram **Superstar**, **GRID** e **Sitemap** (descritos em detalhes em Materiais e Métodos). Esta etapa complementar foi necessária tendo em vista que o estudo preliminar de *docking* (seção 4.1.2), indicou diferentes poses em termos de conformação e orientação para **1** na cavidade 7.



**Figura 4.6** - (a, b, c) Mapas de contorno gerados pelo *software* **Superstar**, usando a função *propensity*, com base nos bancos de dados CSD/PDB, e usando as respectivas *probes*/grupos funcionais; o pontilhado de vermelho indica uma chance 2x maior do que o aleatório de se encontrar o grupo da respectiva *probe* naquela posição, verde 4x, azul 8x. (d) *Hotspots* farmacofóricos gerados pelo **Superstar** com base nos bancos de dados CSD; esfera verde (C-H), esfera azul (N), esfera vermelha (O). Imagens são da estrutura da GSK-3 $\beta$  (1PYX.pdb); quando possível, a pose do **composto 1** gerada pelo programa de *docking* Glide foi inserida em sobreposição para análise.

O primeiro *software* aqui utilizado foi o **Superstar**, que foi capaz de gerar mapas de contorno usando a função de tendência (*propensity*), baseada em dados das bases do CSD, PDB e *IsoStar*. Este *software* emprega *probes* com diferentes características químicas, “pintando” regiões próximas aos resíduos da cavidade. As regiões pintadas por pontos vermelhos indicam que existe uma chance 2x maior de tal *probe* (ou “grupo de prova”) ser encontrada naquela posição do que aleatoriamente; em verde 4x; e em azul 8x.

Assim foi possível inferir que: a cadeia alifática de hidrocarbonetos do **composto 1** deve ser mais favorável quando próxima aos resíduos Arg328, Thr330 e Pro331 (**Figura 4.6**, a e d); a presença de um grupo carbonílico C=O logo acima do N-H da amida de cadeia principal da Ser236 é bastante provável (**Figura 4.6c**), e presença de grupos com carbonos aromáticos é mais provável sob Arg209 (**Figura 4.6b**). Em adição, o **Superstar** também gerou *hotspots* farmacofóricos, representados por esferas coloridas, que corroborou tais indicativos supracitados, conforme pode ser observado na **Figura 4.6d**.

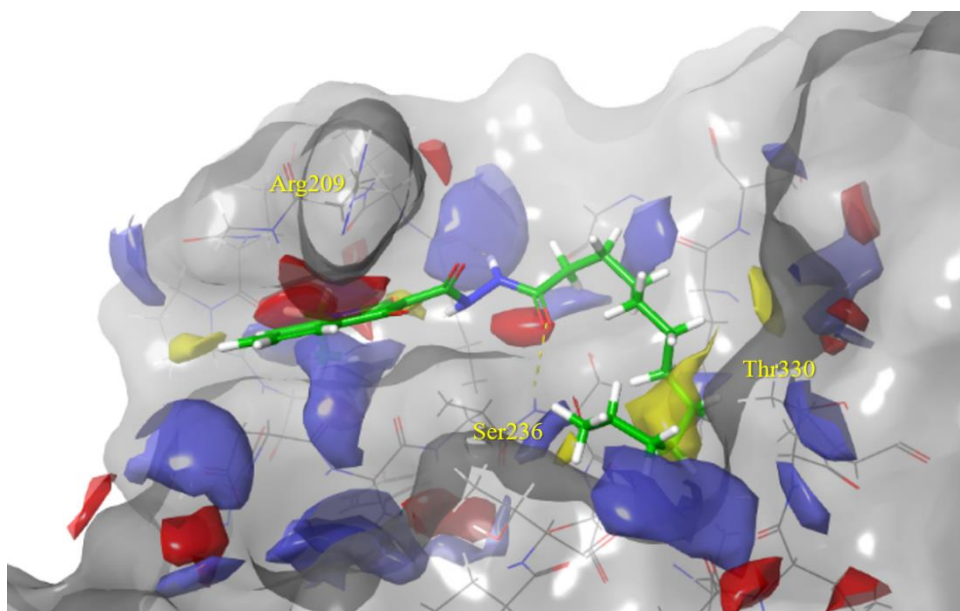


**Figura 4.7** - Mapas de superfície gerados pelo *software* GRID, usando as respectivas *probes*, em diferentes faixas de energia de interação.

De modo análogo ao **Superstar**, o *software* **GRID** gerou mapas de contorno por meio de *Molecular Interaction Fields*, *MIFs*, que mostraram possíveis interações com as sondas aplicadas. Foram usadas sondas de NH de amina, DRY hidrofóbica, C=O de carbonila, e CH<sub>3</sub> de metila, e analisadas as mais baixas energias necessárias para ocorrência de interação entre as mesmas e as regiões da cavidade (assim “pintando” as respectivas regiões). Desta forma, foi possível deduzir fortemente que a cadeia alifática de hidrocarbonetos de **1** tende a ficar próxima ao resíduo Thr330 (**Figura 4.7**, b e d); grupos de amina N-H são preferenciais abaixo do resíduo de

Arg209 (**Figura 4.7a**), e grupos de carbonila C=O são preferenciais logo acima do N-H de amida do resíduo de Ser236 (**Figura 4.7c**).

Por fim, o terceiro *software* aqui empregado, o **Sitemap** gerou mapas de contorno/superfície que corroboraram os mapas obtidos pelos dois *softwares* anteriores, em especial: C=O acima do N-H de Ser236 e anel de quinolona sob Arg209. Adicionalmente, permitiu reforçar que, de fato, a cadeia alifática de hidrocarbonetos do **composto 1** deve estar melhor acomodada no bolsão hidrofóbico da cavidade 7 da GSK-3 $\beta$ , que é constituído principalmente pelos resíduos de Arg328, Thr330 e Pro331. Na **Figura 4.8**, é possível observar os mapas “pintados” nas proximidades dos resíduos da cavidade, sendo que a cor amarela indica regiões preferenciais para grupos hidrofóbicos, vermelho grupos aceptores de ligações de hidrogênio, e azul grupos doadores de ligação de hidrogênio.



**Figura 4.8** - Mapas de superfície gerados pelo **Sitemap**; amarelo indica região hidrofóbica, vermelho região aceptora de ligação de hidrogênio, azul região doadora de ligação de hidrogênio.

Assim, os resultados dos 3 *softwares* de geração de mapas de superfície, rodados de forma complementar e independente, culminaram em indícios consensuais de grupos químicos mais favoráveis em regiões que são compatíveis com a pose do **composto 1** predita pelos *softwares* de *docking* **Glide** e **Autodock**. A sobreposição destas poses de **1** com os mapas gerados na cavidade alostérica 7 da GSK-3 $\beta$ , demonstra uma acomodação compatível, e.g. vide **Figura 4.8** com a pose de **1** obtida pelo **Glide** em acordo com os mapas gerados pelo **Sitemap**.

Com relação as possíveis interações, notamos que três principais interações foram observadas para essa pose de **1** (obtida pelo **Glide**) dentro da cavidade alostérica 7 da GSK-3 $\beta$ : ligação de hidrogênio entre C=O de seu grupo carboidrazida e N-H de amida da Ser236; interação cátion- $\pi$  (ou, ainda, por transferência de carga entre orbitais de fronteira) entre sua porção de quinolona e Arg209; e interações hidrofóbicas entre sua cadeia alifática e a região hidrofóbica Thr330, Pro331 e Arg328. Interessantemente, essas interações também foram previamente descritas por estudos de *blind docking* (PALOMO et al., 2017) e, agora, de fato, corroborados por nossa avaliação utilizando os *softwares* de geração de mapas de contorno/superfície.

#### 4.1.4 Refinamento e validação dos protocolos de *docking*: perspectiva da pose

Os resultados dos estudos preliminares de *docking*, sob a perspectiva das cavidades, indicaram a cavidade 7 da GSK-3 $\beta$  como sítio alostérico mais apto a acomodar o **composto 1**. No entanto, levantou-se o problema da orientação das poses, uma vez que **AutoDock** e **Glide** geraram poses invertidas com relação as geradas pelo **GOLD** e **FRED**. Na seção 4.1.3 anterior, observou-se que as poses de **1** previstas pelos *softwares* **AutoDock** e **Glide** corroboraram as avaliações de mapa de contorno/superfície. Desta forma, vimos a necessidade de aprofundar os estudos de *docking*, sob a perspectiva da pose, para se obter uma pose de **1** ainda mais confiável (tendo em vista diferentes metodologias de *docking*) para esclarecer o problema da pose, em duas etapas.

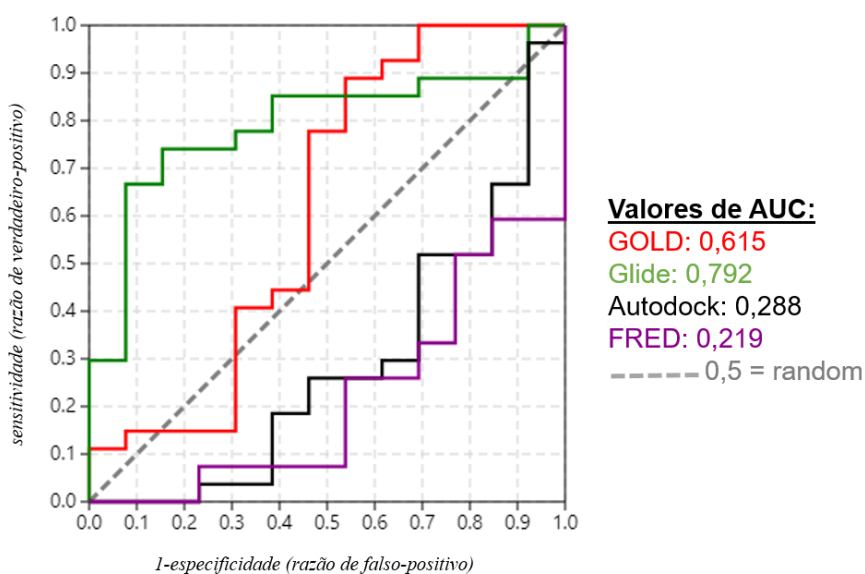
Numa primeira etapa, diferentes protocolos de *docking* foram avaliados para cada *software*, por meio de refinamento e variações de parâmetros/configurações para executar as respectivas simulações de *docking* (vide seção 3.1.5).

Em segunda etapa, foi realizada a validação dos melhores protocolos (para cada *software*) via análise de curva ROC, conforme descrito na seção 3.1.6. Para tanto, em princípio, utilizou-se a série de 40 compostos, que corresponde a **1** e seus análogos (composto 1 a 40 da **Tabela 3.1**) extraídos da literatura (PALOMO et al., 2017). Nesta referência, esta série foi sintetizada e testada em ensaios biológicos de inibição alostérica de GSK-3 $\beta$  para obtenção dos correspondentes valores de IC<sub>50</sub>. Neste trabalho (e também de acordo com tal referência), consideramos a nota de corte de 20  $\mu$ M para valores de IC<sub>50</sub>, separando-os em 27 ativos e 13 inativos.

Cada protocolo de *docking* foi inicialmente testado para análise da obtenção da pose do **composto 1** na cavidade 7 de GSK-3 $\beta$ . Em seguida, foram expandidos para avaliar as poses obtidas para o *dataset* de 40 compostos. Dessa forma, os protocolos refinados e mais eficientes de cada *software* (**GOLD, Glide, Autodock e FRED**) foram selecionados. Na segunda etapa, estes protocolos foram usados para plotar as curvas ROC exibidas na **Figura 4.9**, para avaliar o protocolo do *software* com maior capacidade de discriminar compostos ativos dos inativos, dentre a série de 40 compostos (**Tabela 4.2**).

**Tabela 4.2** - Valores de *AUC*, *TG* e *BEDROC* obtidos a partir das correspondentes curvas ROC plotadas para as validações de *docking*, usando a série de dados de 40 compostos.

Validação de <i>docking</i>	40 compostos		
	<i>AUC</i>	<i>TG</i>	<i>BEDROC</i>
<b>GOLD</b>	0,615	0,169	0,848
<b>Glide</b>	0,792	0,381	0,993
<b>Autodock</b>	0,288	0,318	0,133
<b>FRED</b>	0,219	0,449	0,150



**Figura 4.9** - Curvas ROC obtidas usando a série de 40 compostos para validação dos protocolos/programas de *docking*.

Observa-se que, dentre os quatro *softwares* avaliados, o **Glide** apresentou maior valor de *AUC*, bem como valores satisfatórios de *TG* e *BEDROC* (conforme critérios detalhados na [seção 3.1.6](#)); assim, indicando sua maior capacidade de distinguir e classificar os compostos ativos e inativos contidos na série de 40

compostos. Assim, nesta validação preliminar, o **Glide** foi considerado o *software* com melhor desempenho para atribuir os respectivos valores de pontuação de *docking*

Todavia, reconhecemos que nesta validação, uma série pequena de compostos foi usada. Em adição, estes compostos são provenientes da mesma série de 40 análogos, ou seja, com alta similaridade estrutural. Portanto, é possível e provável que os resultados obtidos possam apresentar viés diante destas limitações dos dados. Então, buscamos na próxima seção 4.1.5, ampliar as séries e usar compostos com maior diversidade estrutural, para, desta forma, otimizar as validações dos protocolos de *docking*, garantindo uma maior confiabilidade nas mesmas.

#### 4.1.5 Otimização das validações de *docking*

A fim de otimizar as validações de *docking*, especialmente no que diz respeito a ampliação da série de inibidores alostéricos de GSK-3 $\beta$ , expandimos tal série para 88 compostos (sendo 37 ativos e 51 inativos), considerando o valor de corte de IC<sub>50</sub> de 20  $\mu$ M para classificá-los. A **Tabela 3.1** mostra estes 88 inibidores alostéricos GSK-3 $\beta$  compilados da literatura e seus respectivos valores de IC<sub>50</sub>.

Vale mencionar que os compostos adicionados foram oriundos de outros trabalhos da literatura (ZHANG et al., 2013, 2014; BROGI et al., 2017), nos quais revelaram-se os inibidores alostéricos de GSK-3 $\beta$ . Estes, comprovadamente inibidores não competitivos de ATP no sítio catalítico, tiveram os seus modos de ação inferidos via ligação à cavidade alostérica 7 por meio de experimentos (ensaios *in vitro* com gráficos duplo recíproco) e/ou simulações de *docking* (consulte a seção 3.1.5).

Também, realizamos as validações com adição de *decoys* gerados a partir da estrutura dos 37 compostos ativos, usando o *webserver* **DUD-E**, o que resultou em aproximadamente 50 *decoys* por composto-ativo; totalizando o *dataset* de 1938 compostos.

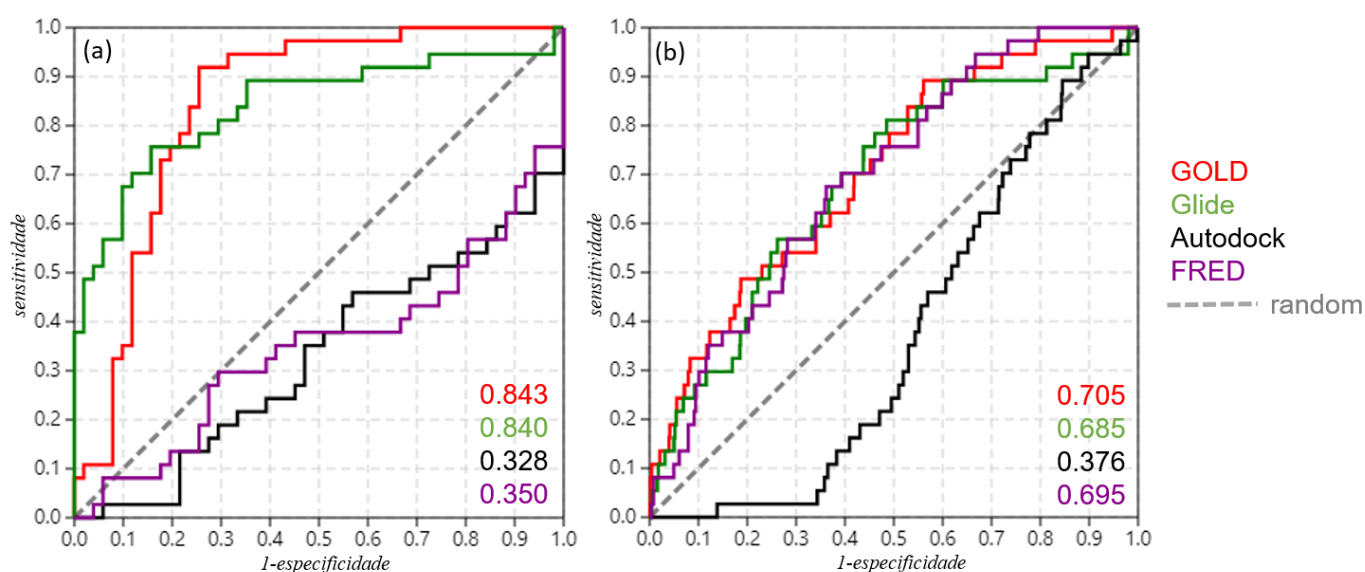
Os mesmos procedimentos de validação dos melhores protocolos de *docking* para cada *software* e usando a série menor de 40 compostos - conforme apresentado anteriormente na seção 4.1.4 - foram aqui aplicados para esta nova série de 88 compostos, com e sem os *decoys* gerados pelo **DUD-E**. Dessa forma, as curvas ROC foram plotadas e os correspondentes valores de *AUC* foram obtidos, conforme mostrado na **Tabela 4.3** e na **Figura 4.10**.



**Tabela 4.3** - Valores de *AUC* obtidos a partir das correspondentes curvas *ROC* plotadas para as validações de *docking* (**GOLD**, **Glide**, **Autodock** e **FRED**), usando série de 88 compostos sem e com *decoys*.

Validação de <i>docking</i>	88 compostos			1938 compostos (88 + <i>decoys</i> )		
	<i>AUC</i>	<i>TG</i>	<i>BEDROC</i>	<i>AUC</i>	<i>TG</i>	<i>BEDROC</i>
<b>GOLD</b>	<b>0.843</b>	0.474	0.729	<b>0.705</b>	0.310	0.223
<b>Glide</b>	<b>0.840</b>	0.455	0.985	<b>0.685</b>	0.240	0.177
<b>Autodock</b>	0.328	0.272	0.148	0.376	0.157	0.002
<b>FRED</b>	0.350	0.207	0.309	0.695	0.267	0.149

\*Em verde programas que corroboram os estudos de predição de pose (*cavity detection* e mapeamento de superfície), em vermelho não corroboram; em negrito maiores valores de *AUC* observados em cada situação.



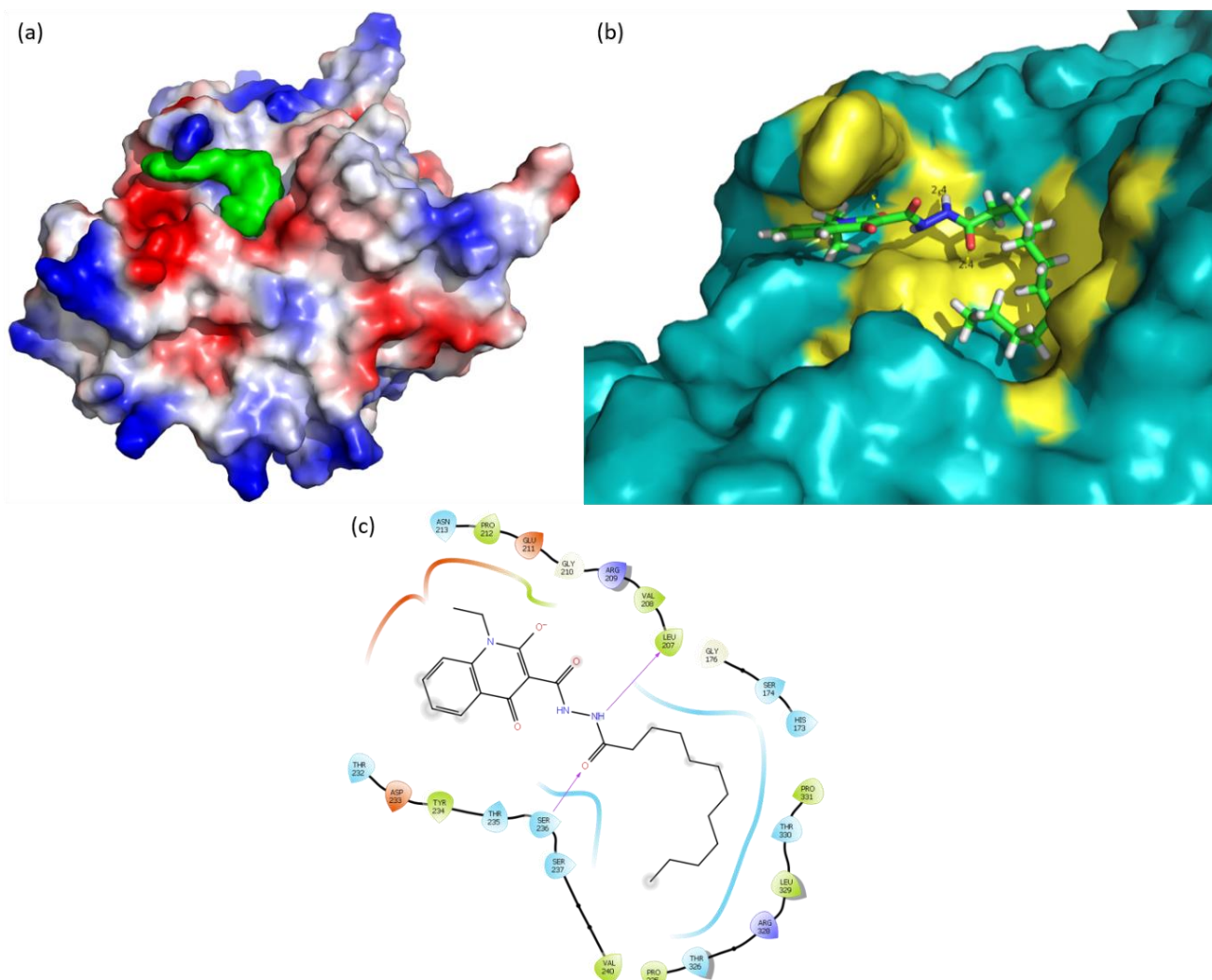
**Figura 4.10** - Validação de *docking* mostrando as curvas *ROC* obtidas para cada programa usando (a) série de 88 compostos, e (b) série de 1938 compostos (88 + *decoys*).

A **Tabela 4.3** mostra que, ao considerar apenas a série de 88 compostos, o **GOLD** e **Glide** apresentaram valores de *AUC* muito próximos, ou equivalentes, de 0,843 e 0,840, respectivamente. O **Autodock** e **FRED** também foram equivalentes, mas com valores de *AUC* em torno de 0,3. Pode-se também fazer uma observação semelhante com relação aos valores de *AUC* obtidos para a série de 88 compostos adicionados dos *decoys* gerados. Isso demonstra que **GOLD** e **Glide** apresentaram melhor desempenho na classificação dos ativos, mesmo com a diversidade estrutural dos compostos tendo sido expandida, agora para a série de 88 compostos, e inclusive com os *decoys*. Em particular, o **FRED** apresentou melhora em seu desempenho (de acordo com seus valores de *AUC*, que passou de 0,350 para 0,695) quando aplicado ao conjunto de 88 compostos + *decoys*.

Usando a série de 88 compostos, o **GOLD** e o **Glide** apresentaram valores de *TG* acima de 0,4; bem como os mais altos valores de *BEDROC*. Já usando a série de 1938 compostos, o **GOLD**, o **Glide** e o **FRED** apresentaram valores de *AUC* acima de 0,68, entretanto valores baixos para as métricas *TG* e *BEDROC*. Isso indica que, com o aumento do *dataset* para a validação de *docking*, tais programas tiveram comprometimento de seus desempenhos e uma probabilidade incerta com relação a classificação de um dado composto ativo ser corretamente classificado (em antemão a um composto selecionado aleatoriamente).

Assim, foi possível inferir que as validações para o **GOLD** e o **Glide** foram mais bem sucedidas quando se usou a série de 88 compostos. Contudo, relembramos que os *softwares* **Glide** e **Autodock** (em verde) foram os que corroboraram os estudos de mapeamento de superfície, enquanto o **GOLD** e **FRED** (em vermelho) não; desta forma, tivemos uma prévia de que o protocolo de *docking* do **Glide** fosse o mais válido, dentre as análises feitas aqui, para descrever as interações dos inibidores alostéricos de GSK-3 $\beta$  frente a respectiva cavidade 7.

Vale ainda ressaltar que, conforme citado anteriormente, três principais interações foram observadas entre a pose de **1** obtida pelo protocolo refinado do **Glide** e a cavidade alostérica 7 da GSK-3 $\beta$ : ligação de hidrogênio entre o C=O do composto e o N-H da Ser236 da enzima; interação eletrostática entre o O<sup>-</sup> (ou OH) de quinolona do composto e a Arg209 da enzima; e interações hidrofóbicas entre cadeia alifática do composto e os resíduos Thr330, Pro331 e Arg328 da enzima. Tais interações são apresentadas na **Figura 4.11**. Ademais, essas interações foram previamente verificadas (PALOMO et al., 2017) e corroborados pela avaliação de mapas de contorno/superfície na seção 4.1.3.



**Figura 4.11** - Pose mais bem classificada de **1** na cavidade alostérica 7 da GSK-3 $\beta$  (1PYX.pdb), obtida usando o protocolo refinado de *docking* pelo **Glide**. (a) representação de superfície da pose (em verde) acomodada na cavidade (mapa de potencial eletrostático mapeado na superfície e produzido pelo **Pymol**); (b) representação em *stick* da pose, em potenciais interações com os principais resíduos (coloridos em amarelo) da cavidade (imagens do **Pymol**); (c) diagrama 2D de interações mostrando as interações inibidor-enzima preditas pelo **Maestro** (flecha rosa: ligação de hidrogênio; fita verde: interações eletrostáticas; fita azul: interações hidrofóbicas).

#### 4.1.6 Validação das poses como *query* para *shape similarity*

As poses de **1**, obtidas por cada programa de *docking*, foram avaliadas com respeito ao seu uso como *query* (molécula de referência) para rodar o programa **ROCS**, que é baseado em similaridade 3D de forma (*shape similarity*), e que é a principal metodologia aplicada na Campanha 1 de triagem virtual. Este programa requer uma pose (em termos de conformação) confiável para efetuar busca comparativa em bases de dados de compostos e, assim, os ranquear. Uma versão alternativa ao **ROCS**, que geralmente é preferencial para estudos de validação, é o

programa **vROCS**, que opera de forma idêntica, porém contém interface gráfica e possibilita a geração de curvas ROC.

Assim, selecionou-se a pose mais bem classificada de **1**, gerada por cada programa (**GOLD**, **Glide**, **Autodock** e **FRED**), e estas foram submetidas como *query* no **vROCS**, utilizando a base de dados consistindo da série de 88 compostos sem e com *decoys*. Previamente ao uso destas bases, estas foram submetidas a geração de até 300 conformações por molécula utilizando-se o **OMEGA** (OPENEYE, 2020c). Para classificar/ranquear as moléculas foram considerados os respectivos valores de *ROCSTC* resultantes do **vROCS**. A partir dos resultados gerados foram plotadas curvas ROC e obtidos os respectivos valores de *AUC*.

Vale mencionar que, em adição às poses dos programas de *docking*, foi usado como pose a conformação de mínima energia para o **composto 1** gerada pelo programa **OMEGA** - que é uma estratégia também recomendada pelos desenvolvedores do **ROCS** para seleção e uso de *query* (HAWKINS; NICHOLLS, 2012; OPENEYE, 2020b). Na **Figura 4.12** são mostradas as curvas ROC e na **Tabela 4.4** os respectivos valores de *AUC* obtidos pelo uso de cada *query*.

Pode-se observar que para todas as *queries* utilizados foram obtidos valores de *AUC* acima do considerado aleatório ( $AUC = 0,5$ ). Destaca-se os maiores valores de *AUC* observados de 0,760 e 0,744 para o **GOLD** e **Glide**, respectivamente, quando se usou a série de 88 compostos. Considerando, entretanto, a série de 1938 compostos, os maiores valores de *AUC* de 0,758 e 0,756 foram observados para o **FRED** e **Glide**, respectivamente.

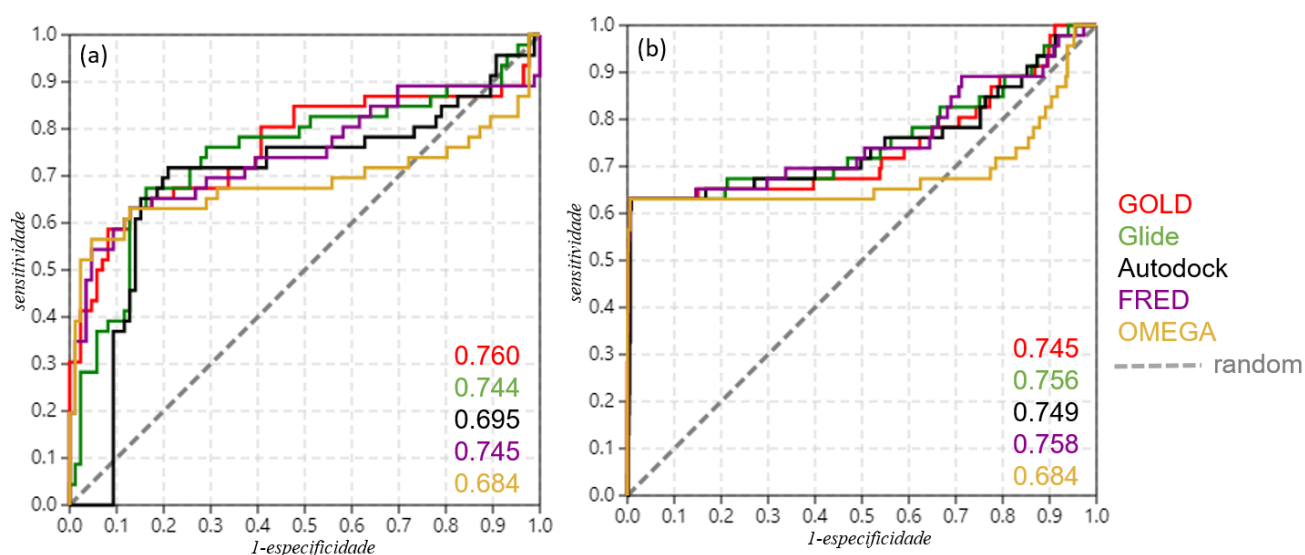
Diante das variâncias pequenas e estatisticamente insignificantes de tais valores de *AUC*, em geral, os 5 *queries* avaliados em termos do *shape* do **composto 1**, apresentaram-se razoavelmente válidos e aparentemente equivalentes para classificar ativos e 'inativos' contidos em cada série de compostos.

Também, com base nas demais métricas *TG* e *BEDROC*, todos os *queries* apresentaram bons resultados nas respectivas avaliações por curvas ROC (conforme os critérios descritos na [seção 3.1.6](#)). Para ambas as séries de 88 e 1938 compostos, todos valores de *TG* ficaram acima de 0,4 e todos os valores de *BEDROC* ficaram acima de 0,6 (com exceção do *query* correspondente a pose de *docking* predita pelo **Autodock**).

**Tabela 4.4** - Valores de *AUC* obtidos nas correspondentes curvas ROC plotadas para as validações de *shape similarity* (ROCS), usando os *queries* obtidos para o **composto 1** pelos respectivos programas de *docking*, usando série de 88 compostos sem e com *decoys*.

Validação de <i>shape similarity</i> (ROCS)	88 compostos			1938 compostos (88 + <i>decoys</i> )		
	<i>AUC</i>	TG	BEDROC	<i>AUC</i>	TG	BEDROC
<b>GOLD query</b>	<b>0,760</b>	0,496	0,947	0,745	0,583	0,666
<b>Glide query</b>	<b>0,744</b>	0,475	0,772	<b>0,756</b>	0,578	0,648
<b>Autodock query</b>	0,695	0,400	0,287	0,749	0,540	0,623
<b>FRED query</b>	0,745	0,500	0,932	<b>0,758</b>	0,580	0,668
<b>OMEGA query</b>	0,684	0,496	0,943	0,684	0,570	0,670

\*Em verde programas que corroboram os estudos de predição de pose (*cavity detection* e mapeamento de superfície), em vermelho não corroboram; em negro maiores valores de *AUC* observados em cada situação.



**Figura 4.12** - Validação ROCS mostrando as curvas ROC obtidas pelo uso de cada *query* dos respectivos programas usando (c) série de 88 compostos, e (d) série de 88 compostos junto aos *decoys*.

Em todo caso, consideramos que a pose do **Glide** apresentou maior confiabilidade para ser usada como *query* em *shape similarity*, pelo fato de ter apresentado valores de *AUC* mais elevados consensualmente nesta validação. Além disso, também mostrou valores de *AUC* elevados de modo consensual para ambas as séries de 88 e 1938 compostos ao longo da validação de *docking* (vide **Tabela 4.3**, seção 4.1.5). Acrescenta-se, ainda, que corroborou com os estudos de mapeamento de superfície da cavidade 7 (realizados na seção 4.1.3).

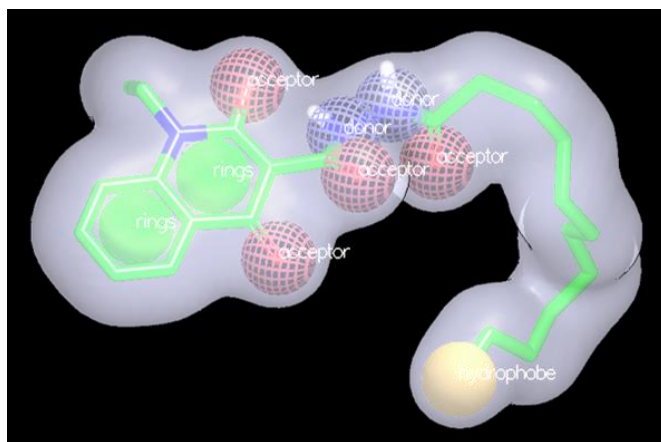
Vale notar que o uso da série expandida de 88 compostos aumenta a diversidade estrutural dos ativos/inativos, com relação a série de 40 compostos usada anteriormente, aumentando, também, o espaço amostral das moléculas e os respectivos domínios de aplicabilidade para os protocolos/programas de *docking*, bem

como para o uso dos *queries* no **ROCS**. Assim, dificulta-se as suas validações, em que será exigido uma performance mais robusta de cada programa avaliado, para que de fato apresentem confiabilidade.

Em alguns casos, em que se observou uma diminuição dos valores de *AUC* quando foram considerados os *decoys* – o que amplia ainda mais a diversidade estrutural da série –, como por exemplo para o **GOLD** e **Glide** na validação de *docking* (vide **Tabela 4.3**), isso pode ter ocorrido pelo fato de haver diferenças triviais entre os ativos e os *decoys* gerados, ou ainda, pelo fato de estes possuírem propriedades físico-químicas não muito dissimilares.

Além disso, considerando que a molécula **1** possui 14 ligações rotacionáveis (ou seja, possui alta flexibilidade), é esperado que dentre os confôrmeros para moléculas inativas e análogas de **1** (gerados pelo **OMEGA** em etapa prévia as sobreposições do **ROCS**) serão geradas muitas opções de conformações, em que possivelmente haverá uma conformação com alta sobreposição com a do *query* de **1** utilizado, conseqüentemente resultando em um alto valor de *ROCSTC*.

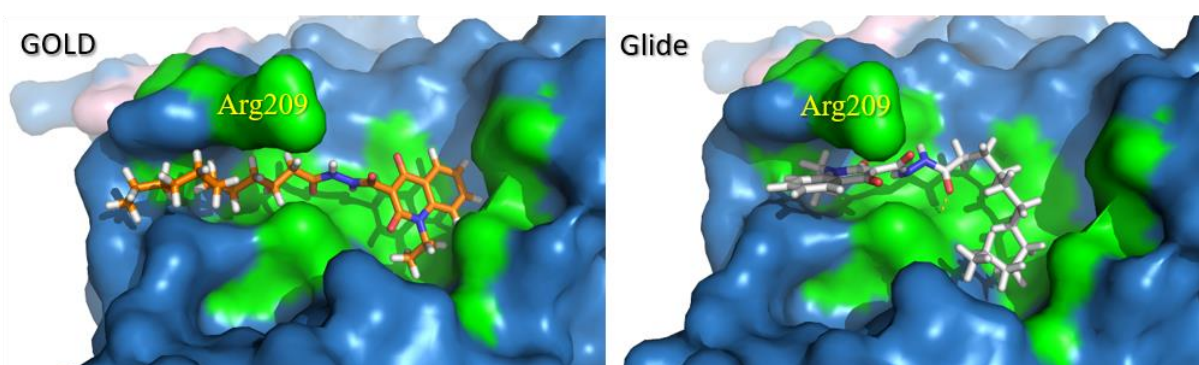
Finalmente, observa-se que os *softwares* **GOLD** e **Glide** mostraram desempenho/performance equivalente nas correspondentes validações de *docking* e de *query* para **ROCS**. No entanto, levando em consideração sobretudo que o **Glide** corroborou os estudos prévios aqui realizados (detecção de cavidade e mapeamento de superfície), bem como da literatura (PALOMO et al., 2011), enquanto o **GOLD** não o fez, nos leva a inferir que os resultados do **Glide** foram capazes de ser mais bem validados e, portanto, ligeiramente mais confiável entre os *softwares* aqui avaliados.



**Figura 4.13** - Conformação da pose **1** (predita pelo Glide) e validada como *query* para o **ROCS**, mostrando seu *shape* (contorno esbranquiçado) e suas características químicas (*color*). Em verde, os anéis aromáticos; em vermelho, grupos aceptores de ligação de hidrogênio; em azul, grupos doadores de ligação de hidrogênio; em amarelo, grupos hidrofóbicos).

#### 4.1.7 Estudos de dinâmica molecular

Tendo em vista os resultados avaliados até aqui e visando confirmar ainda mais a pose de *docking* capaz de representar uma pose bioativa mais confiável do **composto 1** na cavidade alostérica 7 da GSK-3 $\beta$ , foram conduzidos estudos de Dinâmica Molecular (DM). Para isso, utilizamos as poses de *docking* obtidas pelos softwares **Glide** e **GOLD** como arquivos de *input* nas simulações de DM, considerando que uma está em conformação/orientação invertida em relação à outra (**Figura 4.14**).

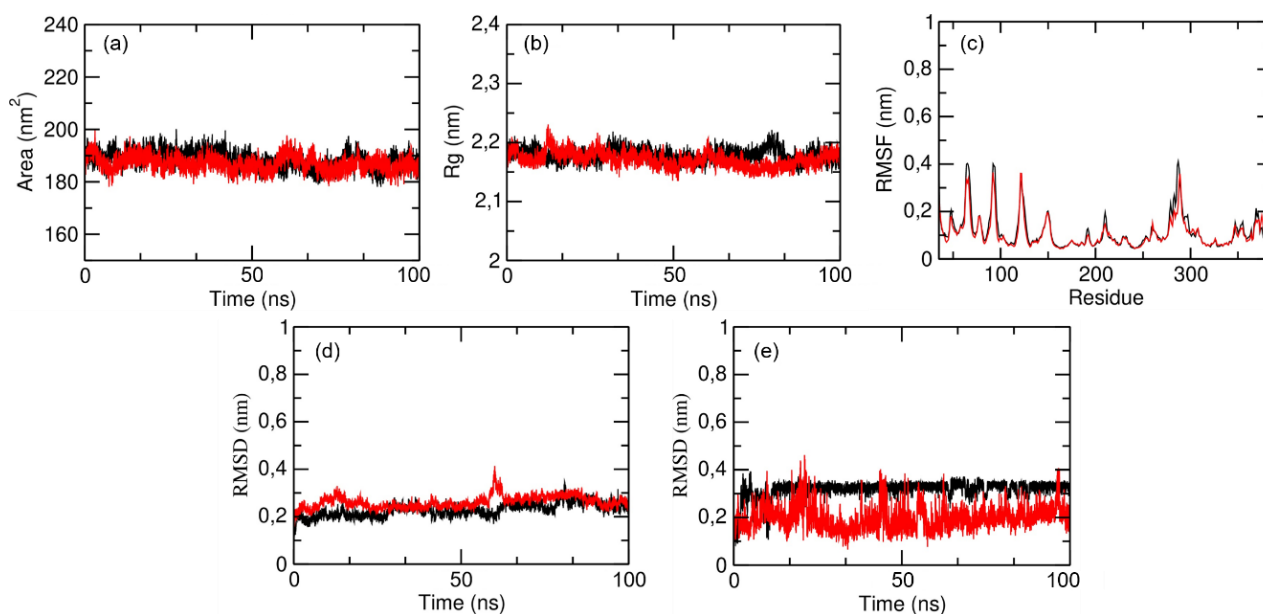


**Figura 4.14** - Conformação da pose de *docking* obtida para o **composto 1** (em representação *stick*) pelo **GOLD** e pelo **Glide**, mostrando que uma está em conformação/orientação invertida em relação à outra. A enzima é aqui representada na forma de superfície (em cor azul), com destaque para a cavidade alstérica (em cor verde).

Desta forma, pudemos obter informações importantes sobre o comportamento dinâmico das poses do **composto 1** em função de trajetórias de 100 ns e em meio solvatado. Tais simulações permitiram analisar a estabilidade dos complexos formados, com cada pose, dentro do sítio alostérico, bem como a flexibilidade dos respectivos resíduos da enzima. Esses dados foram alcançados por meio da obtenção de resultados relativos às métricas: Área de Superfície Acessível ao Solvente (SASA), Raio total de giro da proteína no espaço ( $R_g$ ), *Root-Mean-Square Fluctuation* (RMSF) and *Root-Mean-Square Deviation* (RMSD) para os complexos proteína-ligante, e RMSD apenas para cada ligante (pose) durante a trajetória, conforme apresentado na **Figura 4.15**.

Para determinar a estabilidade e integridade dos complexos durante cada trajetória, avaliamos sua compactação usando as medidas de SASA e  $R_g$ , que foram extraídas a cada 2 fs e plotadas em gráficos (**Figura 4.15**, a e b, respectivamente). Com relação a SASA, ambos os complexos proteína-ligante permaneceram compactos durante a trajetória, ou seja, as estruturas secundárias e terciárias da GSK-

3 $\beta$  foram mantidas durante toda a simulação, sem desdobramento. Esta informação pode ser observada principalmente através dos valores correspondentes  $\sim 190$  nm<sup>2</sup>, que não apresentaram variações significativas. Tal informação do SASA foi confirmada pela análise dos gráficos das medidas de  $R_g$ . Pode-se observar que também não houve grandes variações nos valores de  $R_g$  em ambas as simulações, mostrando que a dimensão estrutural dos complexos permaneceu constante ao longo da trajetória e, portanto, confirmando que os complexos permaneceram compactos e estáveis.



**Figura 4.15** - Gráficos de DM obtidos a partir dos complexos formados entre as poses de *docking* de **1** **Glide** (preto) e **GOLD** (vermelho) no sítio alostérico 7 da GSK-3 $\beta$ . (a) SASA; (b)  $R_g$ ; (c)  $RMSF$  para os complexos; (d)  $RMSD$  para os complexos; (e)  $RMSD$  apenas para os ligantes (poses).

Em relação aos gráficos de  $RMSF$  (**Figura 4.15c**) pode-se notar que os complexos proteína-ligante com ambas as poses apresentaram quatro picos de  $\sim 0,4$  nm, indicando uma grande flutuação nos resíduos/aminoácidos que compõem estas regiões. Destes, o primeiro pico indica proximidade ao resíduo de Phe67 e o segundo próximo do Phe93; de fato, ambos estes resíduos fazem parte da cavidade 2 da GSK-3 $\beta$  que corresponde a cavidade do substrato (vide **Figura 1.7**). Especulamos que suas flutuações consideráveis poderiam indicar modulações correspondentes como consequência da ligação alostérica do **composto 1** na cavidade 7. No entanto, maiores investigações seriam necessárias para confirmar isso. Em adição, a partir da literatura (PALOMO et al., 2017), sugestões para a modulação alostérica de **1** ligado ao sítio 7 foram feitas inferindo que o resíduo Glu211 se aproxima de Lys205. Aqui,



também verificamos um pico de *RMSF* considerável para este primeiro resíduo, o que pode corroborar esta informação.

Além disso, o equilíbrio e a estabilidade da formação dos complexos também foram analisados avaliando as medidas de *RMSD* ao longo da trajetória. A partir da **Figura 4.15d**, podemos então observar que não houve grandes variações nos valores de *RMSD* em ambas as simulações. Além disso, os valores médios de *RMSD* para a pose do **Glide** e do **GOLD** foram de  $0,238 \pm 0,031$  e  $0,265 \pm 0,025$  nm, respectivamente, o que poderia significar uma ligeira indicação de melhor estabilidade para o complexo com **1** (pose de **Glide**) na cavidade alostérica 7 da GSK-3 $\beta$ .

Por fim, analisamos os valores de *RMSD* apenas para os ligantes (poses) de cada complexo (**Figura 4.15e**). A partir desta análise foi possível avaliar se os ligantes permaneceram em suas poses iniciais (ou seja, nas suas poses de *docking*, em termos de conformação + orientação) da simulação, ao longo da trajetória, ou se a dinâmica do sistema (proteína, ligante, solvente) favoreceu seu movimento, levando a reajustes em suas poses como consequência de outras interações com a proteína. Observamos que a pose do **Glide** (em preto) apresentou *RMSD* constante  $\sim 0,3$  Å, enquanto a pose do **GOLD** (vermelha) apresentou algumas variações em *RMSD*  $\sim 0,2$  Å. Esses valores de *RMSD* diferem apenas em 0,1 unidade e não devem, em geral, indicar diferença significativa entre eles. Por outro lado, variações de maior amplitude da pose de **GOLD** poderiam sugerir que durante a trajetória seu posicionamento pode adotar diferentes conformações, e.g. devido ao movimento de sua longa e flexível cadeia alifática.

Em geral, os resultados foram equivalentes, com leves sugestões de que a pose do **Glide** poderia formar um complexo mais estável do que a do **GOLD**, conforme indicado pelos valores de *RMSD* obtidos ao longo das trajetórias de DM. A análise dos resultados de DM foi consistente e também nos permitiu verificar o potencial do **composto 1** inibir a GSK-3 $\beta$  alostericamente. Estudos nesta direção podem ser explorados no futuro, por exemplo, agregando análises comparativas do comportamento dos *hits* obtidos por triagem virtual, frente à inibição alostérica da GSK-3 $\beta$ .

#### 4.1.8 Considerações finais – PARTE I

As investigações conduzidas usando diferentes metodologias *in silico* - e considerando que o **composto 1** é um inibidor não competitivo de ATP que atua em (alg)um sítio alostérico da GSK-3 $\beta$  (assim confirmado por ensaios *in vitro* (PALOMO et al., 2017) -, apontaram de forma consensual que, o mais provável, é que **1** atue, de fato, na cavidade alostérica 7 da enzima. Enfatiza-se que isso fora indicado por todos os estudos realizados (nas seções 4.1.1 a 4.1.7) e, além disso, corrobora achados anteriores (PALOMO et al., 2017).

Sobretudo, dentre as análises comparativas de *docking* aqui realizadas, a suposta pose bioativa de **1** obtida pelo protocolo de *docking* desenvolvido usando o **Glide**, foi a mais robusta e condizente para correspondente aplicação em estudos de prospecção baseados em ligantes. Inclusive, a condução de tais estudos de prospecção e triagem virtual pode reforçar ainda mais a validade destes achados.

Ressaltamos que com o termo 'pose supostamente bioativa' nos referimos à conformação e orientação que tal ligante deve assumir no espaço tridimensional, em uma acomodação energeticamente mais favorável dentro do sítio da enzima, em adição a interações intermoleculares consistentes com os respectivos resíduos de aminoácidos. Desta forma, supostamente desencadeando uma modulação na atividade da enzima, bem como de outros posteriores processos bioquímicos que levem a constatação de um efeito/resposta biológica.

É claro que, objetivamente, a obtenção do respectivo complexo cristalográfico seria, em última instância, a prova de conceito para os estudos em questão. Entretanto, dentro do nosso escopo, a aplicação dos conhecimentos e dados aqui obtidos em campanhas de triagem virtual, para selecionar compostos químicos existentes e avaliar os correspondentes potenciais de inibição frente a GSK-3 $\beta$ , também se apresenta como uma abordagem coerente e significativa, do nosso ponto de vista.

De fato, a literatura (VERÍSSIMO et al., 2022) vem destacando a importância de se usar conjuntos de estratégias metodológicas (conforme conduzido aqui na *PARTE I*) como uma forma de amplificar a quantia de dados disponíveis (para casos que, como aqui, carecem de informações) e, assim, melhorar a eficácia de campanhas de planejamento de compostos bioativos.

## 4.2 **PARTE II**: Triagens virtuais para identificação de inibidores alostéricos de GSK-3 $\beta$

### 4.2.1 Triagem virtual usando a **Campanha 1**

A **Campanha 1** consistiu principalmente em aplicação de triagem virtual baseada em forma, considerando a similaridade 3D de forma avaliada pelo **ROCS** entre o *query* – suposta pose bioativa validada para o **composto 1** desenvolvida na **PARTE I** deste trabalho – e milhões de moléculas contidas nas bases de dados usadas. Especificamente, filtramos 5.000 moléculas de cada uma das 17 bases de dados, ou seja, 85.000 moléculas, a partir de mais de 39 milhões de moléculas.

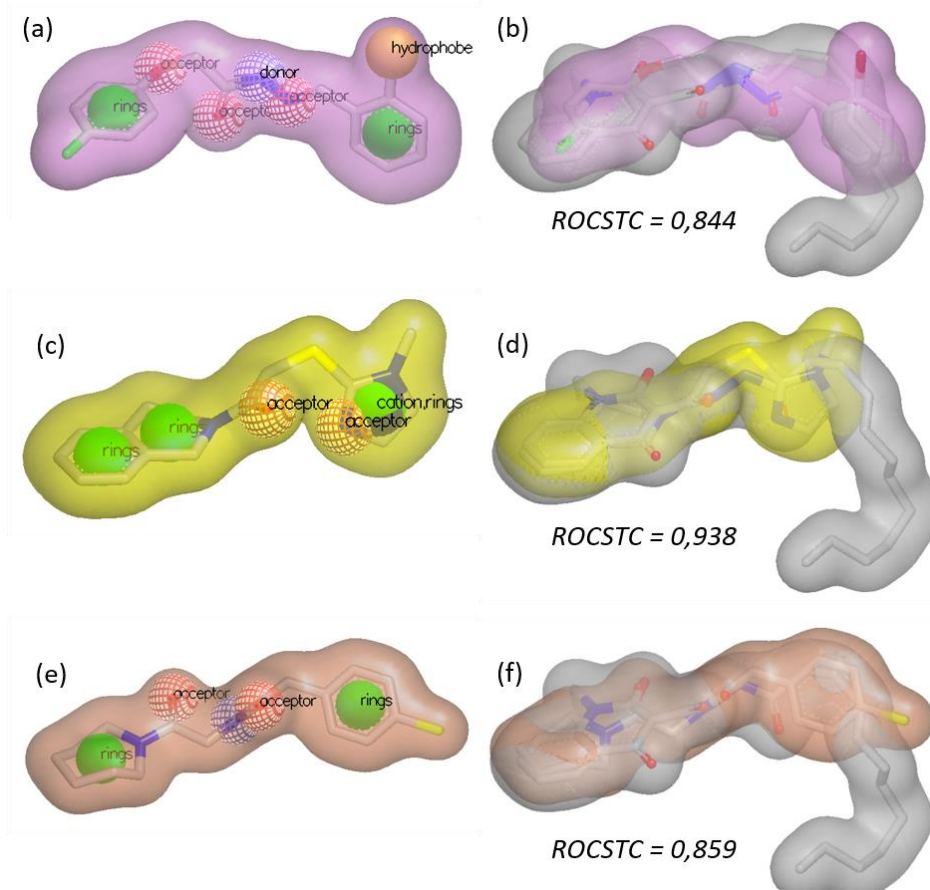
As 17 bases de dados (vide **Tabela 3.2**) foram selecionadas considerando a diversidade química das moléculas, que inclui compostos disponíveis comercialmente, produtos naturais, bem como coleções pré-filtradas de moléculas fármaco-similares (*drug-like*) ou SNC-similares, por exemplo: *Chembridge Diverset CL*, *NPASS* e *ZINC CNS*, respectivamente. Vale ressaltar que algumas dessas bases de dados correspondem a versões anteriores e/ou bases de dados pré-processadas em nosso laboratório – que se mostraram fontes eficientes de *hits* em estudos prévios de triagem virtual (RODRIGUES; SILVA, 2017; DE ALMEIDA et al., 2018; FEDERICO et al., 2020).

A título de ilustração, mostramos na **Figura 4.16** as estruturas químicas de 3 compostos (posteriormente selecionados como *hits* pela triagem da **Campanha 1**), representadas pelos seus *shapes* individualmente e, em sobreposição com o *query* do **composto 1**.

Após o uso do **ROCS**, os 85.000 compostos foram filtrados considerando a similaridade 3D de potencial eletrostático do *software* **EON**, e usando o mesmo *query* de **1**, resultando em 17.000 compostos. Esta sequência combinada de metodologias em triagem (ou seja, filtragem por forma seguida por eletrostática) também é reconhecida por fornecer resultados mais promissores em nosso grupo de pesquisa (DE ALMEIDA et al., 2018; FEDERICO et al., 2020, 2021b).

Para predições de propriedades ADME usando o **Qikprop**, ênfase foi dada às propriedades físico-químicas e farmacocinéticas que são conhecidas na literatura por se correlacionarem com fármacos que atuam no SNC, i.e., SNC-similares (WAGER et al., 2010; GHOSE et al., 2012; RANKOVIC, 2017). Com base nisso, adaptamos e

derivamos algumas regras/critérios para filtrar os compostos que possivelmente poderiam se tornar candidatos a fármacos neste contexto; restando, portanto, 1.100 moléculas após a filtragem do **Qikprop**. Estes foram submetidos à predição de toxicidade usando o **DEREK**, removendo-se aqueles que foram preditos, no mínimo, como "plausíveis" para os *endpoints* considerados.



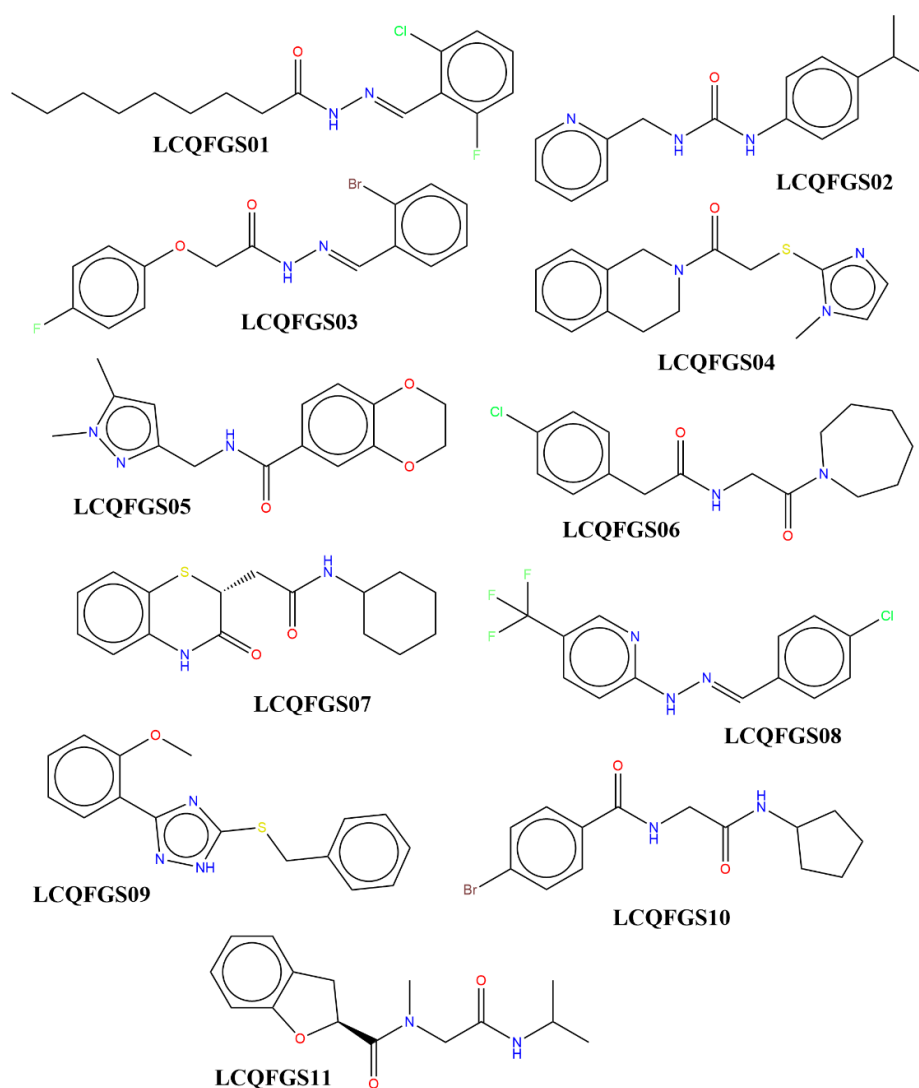
**Figura 4.16** - Representação dos *shapes* obtidos pelo **ROCS** para três compostos, bem como as respectivas sobreposições com o *shape* do *query* (**composto 1**, em representação *stick*, com forma/contorno colorida em cinza) e seus valores de *ROCSTC*. (a,b) **LCQFGS03**, em cor magenta; (c,d) **LCQFGS04**, em cor amarela; (e,f) **LCQFGS06**, em cor salmão. Os três compostos foram selecionados como *hits* pela Campanha 1- vide a frente seção 4.2.3.

Nesta altura, os 656 compostos restantes foram submetidos à triagem baseada em estrutura usando nosso protocolo refinado de *docking* do **Glide** (previamente descrito na *PARTE I*). Enfatiza-se que tal protocolo foi minuciosamente desenvolvido e validado considerando que não há nenhum complexo proteína-ligante obtido experimentalmente na cavidade alostérica 7 da GSK-3 $\beta$ , reportado até o momento.

Aplicou-se este protocolo de *docking* com ênfase na inspeção visual cuidadosa das interações entre os compostos restantes, dentro da cavidade alostérica 7 da GSK-

3 $\beta$ . Mais especificamente, procuramos manter compostos que estabelecessem ligação de hidrogênio com Ser236, grupos aromáticos/halogênios sob Arg209 e grupos hidrofóbicos próximos à Thr330 - analisando-se conjuntamente com os resultados obtidos anteriormente de valores de *ROCSTC* (**ROCS**), *EONTC* (**EON**) e ADME/Tox. Deste modo, selecionamos 11 compostos para serem comprados.

A título de ilustração, mostramos na **Figura 4.17** as estruturas químicas destes 11 compostos selecionados pela Campanha 1.



**Figura 4.17** - Representação das estruturas químicas 2D dos 11 *hits*, **LCQFGS01** – **11**, obtidos pela Campanha 1 de triagem virtual.

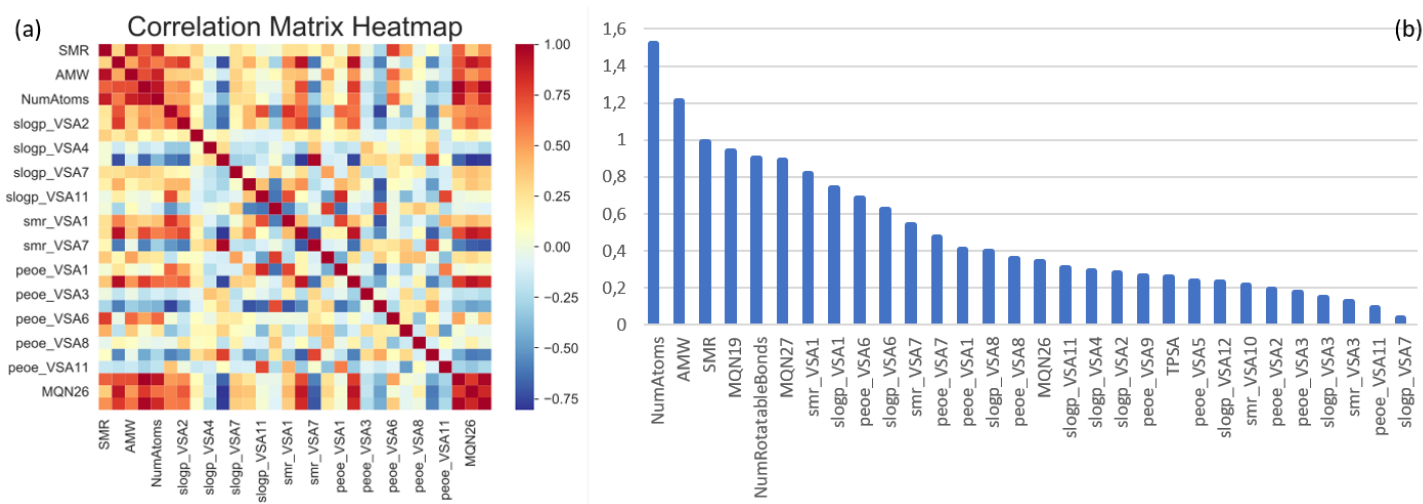
#### 4.2.2 Triagem virtual usando o Campanha 2

Ao se pesquisar por inibidores de GSK-3 $\beta$  na literatura, pode-se verificar que existe uma quantidade considerável de inibidores competitivos de ATP (ortostéricos)

bem como de não competitivos de ATP. Para este último, no entanto, raramente são encontrados dados sobre qual outro sítio deve acomodar os respectivos inibidores (podendo, ou não, serem classificados como alostéricos). Portanto, executamos uma busca exaustiva em artigos da literatura para compilar a série de dados de 88 moduladores de GSK-3 $\beta$  que potencialmente atuam no sítio alostérico 7.

Esta série, apresentada na **Tabela 3.1**, foi usada na *PARTE I* para os estudos de validação e, aqui, para gerar os modelos de QSAR. Uma vez que esta série de compostos consiste de 88 moduladores alostéricos de GSK-3 $\beta$  'selecionados a dedo', a etapa de curagem da mesma, não culminou na remoção de duplicatas e/ou na verificação de problemas adicionais de estrutura química.

Conforme mencionado nos métodos, desenvolvemos três modelos usando três conjuntos de descritores, ou seja, Modelo 1 usando *RDKit*, Modelo 2 usando *Morgan* e Modelo 3 usando *RDKit+Morgan*. Vale lembrar que cada modelo na verdade consiste de 5 modelos consensuais, cada um gerado usando *Random Forest* e usando uma combinação diferente de séries de treinamento/teste na etapa de validação cruzada externa *5-fold*.



**Figura 4.18** - (a) Mapa de calor da matriz de correlação considerando os 28 descritores (propriedades *RDKit*) para a construção do Modelo 1 de QSAR via aprendizagem de máquina. (b) Importância média (adimensional) de cada descritor considerado na construção do Modelo 1.

No Modelo 1, as propriedades *RDKit* mais relevantes foram selecionadas após a filtragem daquelas com baixa variância (< 10%) e análise de correlação entre elas. Para este último, foi gerado um mapa de calor da matriz de correlação (**Figura 4.18a**) considerando as 28 propriedades selecionadas para gerar o Modelo 1. Pode-se notar que as propriedades selecionadas não apresentaram correlação superior a 90%. Além

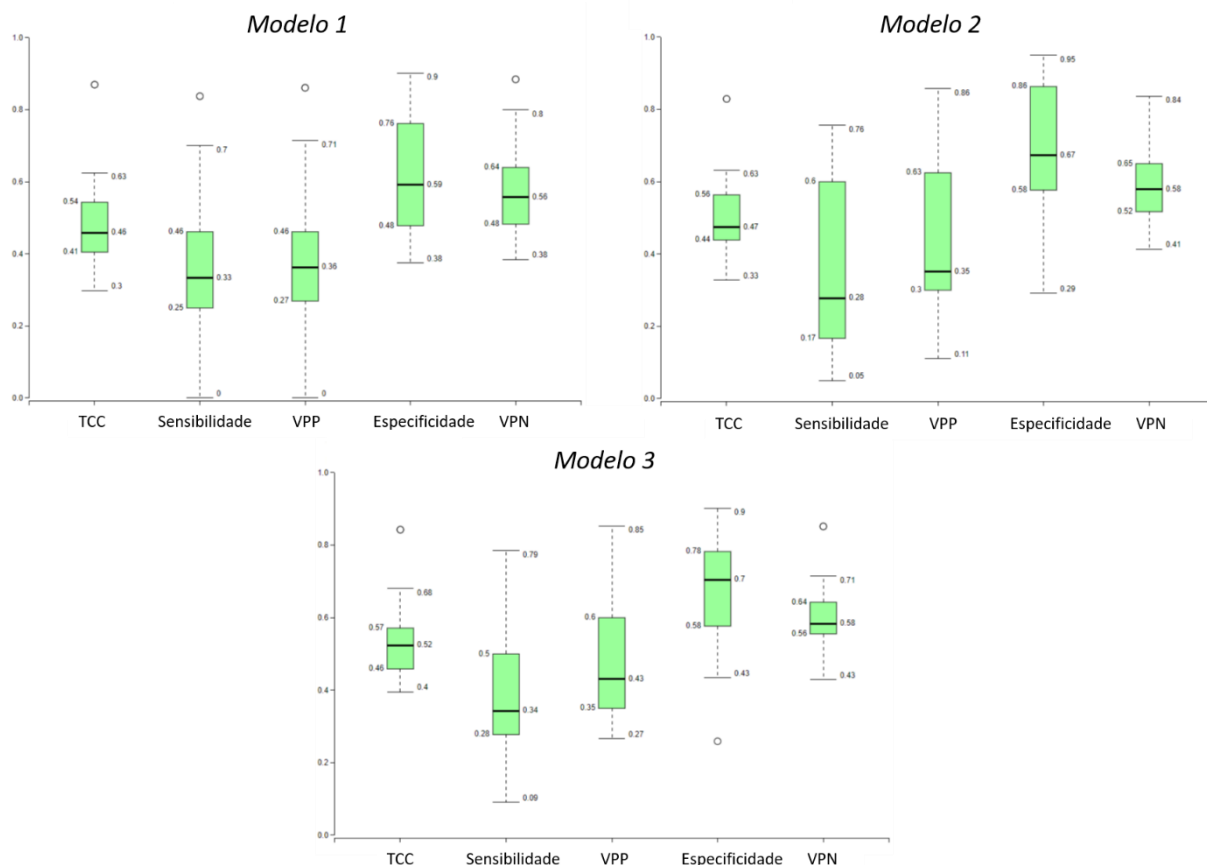
disso, a **Figura 4.18b** também mostra a importância (ou peso) de cada descritor calculada após a geração do Modelo 1.

Dentre os descritores (propriedades físico-químicas estruturais do *RDKit*) com maior peso para geração dos modelos de QSAR, observa-se que boa parte deles têm relação com o tamanho das moléculas, como por exemplo, número de átomos e *AMW* (*average molecular weight*). Enfatizamos que as definições detalhadas para os descritores *RDKit* mostrados na **Figura 4.18b** constam na página <https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>.

Em especial, notamos que os descritores número de ligações rotacionáveis, *SMR* e *slogp\_VSA1* correspondem (ou estão correlacionados) a parâmetros hidrofóbicos que expressam a lipofilicidade dos compostos. Tal constatação, é bastante coerente em vista da série de 40 compostos análogos de **1** que fora sintetizada e testada frente a inibição de GSK-3 $\beta$  (PALOMO et al., 2017); verificando-se que a cadeia lipofílica de hidrocarbonetos tem uma relação direta com os valores de IC<sub>50</sub> (i.e., quanto maior a cadeia, maior a lipofilicidade, menor o IC<sub>50</sub> e portanto mais ativo).

É importante ressaltar que o Modelo 1 mostrou as melhores métricas estatísticas para a validação de aleatorização de Y quando comparada aos Modelos 2 e 3 (**Figura 4.19**). Os gráficos de *box plot* permitem visualizar que a maior concentração/população dos valores, obtidos para cada métrica, situam-se nas caixinhas verdes e, também, que a linha preta e grossa que corta essas na horizontal, indica a mediana dos valores.

Assim, por exemplo, para o Modelo 1, a mediana dos valores de TCC obtidos pelas 20 rodadas de *Y-randomization* foi de 0,46, com uma maior população destes valores variando entre 0,54 e 0,41. O ponto/círculo branco acima indica que o valor real de TCC do Modelo 1 (de 0,87) corresponde a um *outlier* e, portanto, mostra que o modelo não foi gerado ao acaso. De forma geral, portanto, pode-se analisar os gráficos de *box plot* pela quantidade de *outliers* (círculos brancos) observados para cada métrica, de cada modelo. Em outras palavras, quanto maior a quantidade de *outliers*, melhor. Para o Modelo 1 houveram quatro *outliers*, e para o Modelo 2 e 3 houveram um e dois *outliers*, respectivamente.



**Figura 4.19** - Validação de aleatorização de Y: *box plot's* para cada métrica obtida para os correspondentes modelos 1, 2 e 3 de QSAR.

O APD foi também estimado para a geração dos modelos conforme descrito nos materiais e métodos (seção 3.2.2.2). Para os 3 modelos, foram obtidos valores de APD coverage de 1, ou seja, uma cobertura de 100% do espaço químico definido pela relação de similaridade entre os compostos contidos nas séries de treinamento e de teste, bem como pelos domínios de aplicabilidade considerados na geração dos respectivos modelos de QSAR.

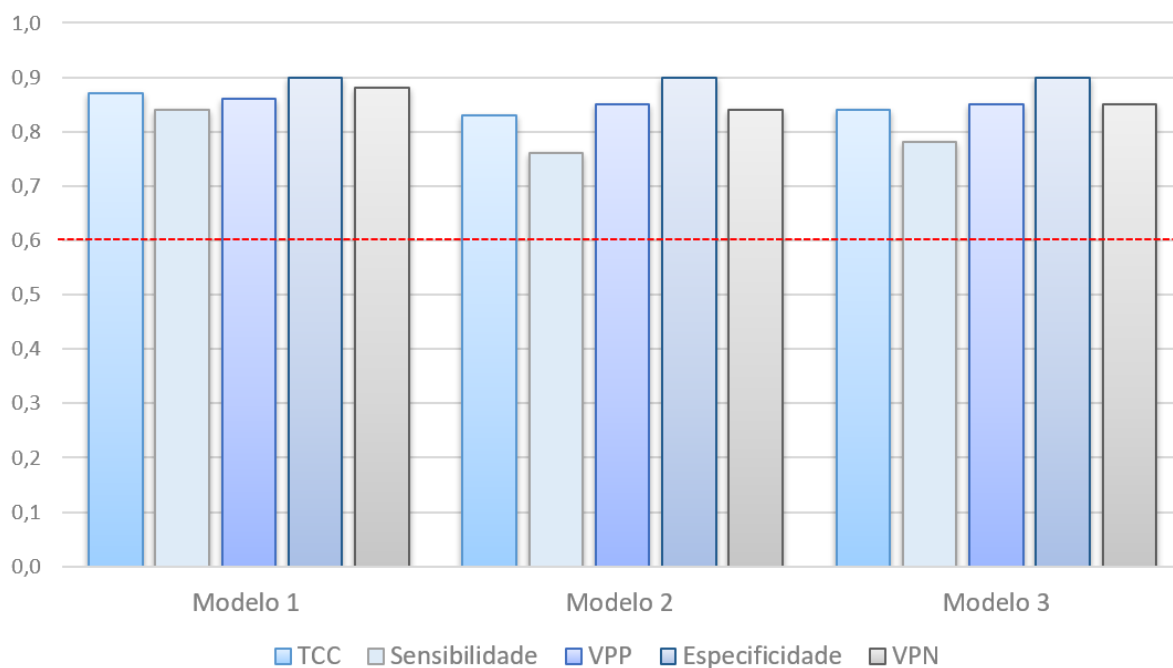
**Tabela 4.5** - Características estatísticas (métricas) dos modelos de QSAR gerados por aprendizagem de máquina.

Modelo	Algoritmo	Descritores	TCC	Sensibilidade	VPP	Especificidade	VPN
1	Random Forest	<i>propriedades RDKit</i>	0,87	0,84	0,86	0,90	0,88
2	Random Forest	<i>fingreprints Morgan</i>	0,83	0,76	0,85	0,90	0,84
3	Random Forest	<i>RDKit+Morgan</i>	0,84	0,78	0,85	0,90	0,85

Em geral, todos os modelos apresentaram características estatísticas (métricas), sobretudo com relação ao valor de TCC, superiores ao limite usual e aceito de 0,6 (TROP SHA, 2010). No entanto, o modelo de QSAR construído apenas com



descritores *RDKit* (Modelo 1) apresentou valores ligeiramente maiores de TCC de 0,87, sensibilidade de 0,84, especificidade de 0,90 e valores de preditividade positivo e negativo (VPP e VPN) de 0,86 e 0,88, respectivamente (vide **Tabela 4.5** e **Figura 4.20**).



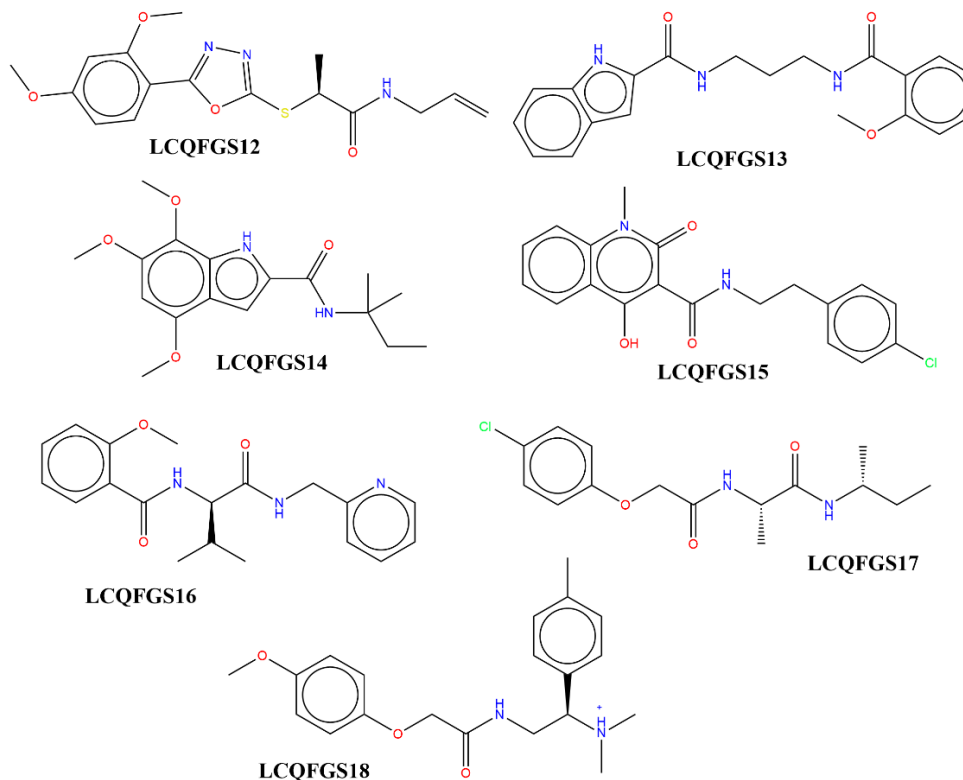
**Figura 4.20** - Representação gráfica dos valores das métricas (características estatísticas) obtidas para os modelos de QSAR gerados por aprendizagem de máquina. Linha pontilhada em vermelho indica a nota de corte de 0,6.

O Modelo 1 foi então empregado na Campanha 2 de triagem virtual usando as bases de dados *Molport Screening Compounds* e *Enamine* (mais de 10 milhões de compostos). Para filtrar os compostos, considerou-se aqueles que foram preditos como ativos (valor binário 1), dentro do APD, bem como os respectivos valores de predição de confiança/pontuação de QSAR (*QSAR score*). Este último, é um valor de 0,00 a 1,00 (ou de 0 a 100%) que indica a taxa de concordância entre todos os submodelos constituintes do modelo, capazes de predizer determinado composto como ativo (valor binário 1) ou inativo (valor binário 0).

Os 40.000 compostos preditos como ativos, mais bem ranqueados de acordo com seus valores de *QSAR score*, foram então separados e encaminhados para filtragem de propriedades ADME/Tox usando o **Qikprop** e **DEREK**. Isso resultou em 76 compostos que foram submetidos a simulações de *docking* usando o protocolo do

**Glide** para inspecionar as interações preferenciais dentro da cavidade alostérica 7 da GSK-3 $\beta$ . Por fim, 7 compostos foram selecionados para compra.

A título de ilustração, mostramos na **Figura 4.21** as estruturas químicas destes 7 compostos selecionados pela Campanha 2.



**Figura 4.21** - Representação das estruturas químicas dos 7 *hits*, **LCQFGS12 – 18**, obtidos pela Campanha 2 de triagem virtual.

#### 4.2.3 Avaliação biológica e análise dos *hits* obtidos pelas triagens virtuais

A partir das Campanhas 1 e 2 de triagem virtual foi possível obter um total de 18 *hits* virtuais. Estes foram adquiridos e testados conforme os ensaios biológicos descritos na seção 3.2.5 de Materiais e Métodos. Por meio de tais ensaios de atividade biológica, pudemos obter para os *hits*, apenas a % de inibição de GSK-3 $\beta$  na concentração de 100  $\mu$ M. É importante lembrar, no entanto, que a maioria dos compostos ativos usados como referência neste trabalho apresentam valores de IC<sub>50</sub> próximos a 3  $\mu$ M, o que poderia indicar que os correspondentes valores de inibição em %, não estariam muito distantes dos aqui obtidos, em condições semelhantes.

Além disso, segundo Palomo et al. (2017) (PALOMO et al., 2017), é esperado que o mecanismo alostérico forneça uma modulação sutil da GSK-3 $\beta$  e, assim, tais fenômenos possam ser vantajosos em vista da menor chance de produzir efeitos

adversos. Em adição, considerando a proposta desafiadora deste trabalho, ou seja, em vista da baixa quantidade de dados disponíveis (poucos inibidores conhecidos na literatura, bem como da inexistência de complexo cristalográfico ligante-receptor), os *hits* iniciais aqui obtidos por triagem virtual podem ser considerados ativos, num momento preliminar, frente a inibição da GSK-3 $\beta$ . Desta forma, estão apresentados na **Tabela 4.6** os valores de % de inibição de GSK-3 $\beta$  para os 18 *hits* obtidos, os quais variaram de 0,01 a 18,41%.

**Tabela 4.6** - Compostos *hits* obtidos por triagem virtual e correspondentes valores % de inibição de GSK-3 $\beta$  medidos pelos ensaios *in vitro*. Acrescidos de seus valores de ROCSTC, QSAR score, **Glide XP** score e propriedades físico-químicas e farmacocinéticas previstas.

Composto	GSK-3 $\beta$ % inibição	Campanha VS	ROCS TC	QSAR score	Glide XP	MW	PSA	<sup>QP</sup> logP <sub>ow</sub>	<sup>QP</sup> logBB	%HOA
<b>1</b>	-	-	-	-	-6.130	429.25	118.58	5.84	-2.04	94.20
LCQFGS01	0.10	1	0.884	-	-5.919	312.81	48.80	5.14	-0.425	100.00
LCQFGS02	0.01	1	0.917	-	-5.622	269.35	57.19	3.02	-0.357	100.00
<b>LCQFGS03</b>	<b>12.44</b>	1	0.844	-	-6.065	351.17	56.01	4.26	-0.094	100.00
<b>LCQFGS04</b>	<b>18.41</b>	1	0.938	-	-5.407	287.38	41.11	2.83	0.056	100.00
LCQFGS05	9.98	1	0.947	-	-5.045	287.32	71.22	2.82	-0.238	100.00
<b>LCQFGS06</b>	<b>14.02</b>	1	0.859	-	-5.679	308.81	62.67	2.52	-0.100	96.44
LCQFGS07	0.63	1	0.907	-	-4.760	304.40	75.36	2.10	-0.504	90.79
LCQFGS08	8.20	1	0.866	-	-5.355	299.68	35.22	4.4	0.389	100.00
LCQFGS09	4.32	1	0.813	-	-5.171	297.37	41.66	4.19	0.018	100.00
LCQFGS10	4.04	1	0.852	-	-5.005	325.20	73.61	2.67	-0.333	96.01
LCQFGS11	6.65	1	0.858	-	-5.225	276.34	71.36	1.20	-0.248	87.87
<b>LCQFGS12</b>	<b>11.59</b>	2	-	0.894	-5.131	349.40	75.98	2.84	-0.317	100.00
LCQFGS13	4.33	2	-	0.839	-6.711	351.40	92.98	3.56	-1.064	100.00
LCQFGS14	1.88	2	-	0.838	-4.865	320.39	69.97	3.69	-0.279	100.00
<b>LCQFGS15</b>	<b>16.77</b>	2	-	0.833	-5.041	356.81	81.25	4.26	-0.711	100.00
LCQFGS16	5.76	2	-	0.831	-5.683	341.41	79.89	3.12	-0.522	100.00
LCQFGS17	4.26	2	-	0.820	-5.518	312.79	78.40	2.63	-0.413	95.06
LCQFGS18	7.13	2	-	0.819	-5.716	342.44	60.07	3.06	-0.051	93.46

\* ROCSTC: ROCS TanimotoCombo índices; QSAR score: valores de confiança para as predições do modelo; **Glide XP**: valores de score da função de pontuação *extra precision* (em kcal/mol); MW: peso molecular em g/mol; PSA: área de superfície polar; <sup>QP</sup>logP<sub>ow</sub>: valores do logaritmo do coeficiente de partição em *n*-octanol/água preditos pelo **QikProp**; <sup>QP</sup>logBB: valores do logaritmo do coeficiente de permeabilidade da barreira hematoencefálica pelo **QikProp**; HOA: absorção oral humana em %.

Neste trabalho, consideramos que foram encontrados 5 *hits* mais representativos: os compostos **LCQFGS03**, **LCQFGS04**, **LCQFGS06**, **LCQFGS12** e **LCQFGS15**, com valores de % de inibição da GSK-3 $\beta$  de 12,44, 18,41, 14,02, 11,59 e 16,77, respectivamente (destacados em negrito na **Tabela 4.6**). Os três primeiros foram obtidos pela Campanha 1 de triagem (com ênfase na aplicação da similaridade

3D de forma do **ROCS**) e os demais pela Campanha 2 (com ênfase na aplicação do Modelo 1 de QSAR via aprendizagem de máquina).

Se assumirmos a nota de corte de 10% de inibição, foram 3/11 e 2/7 de compostos razoavelmente ativos alcançados pelas Campanhas 1 e 2, respectivamente, o que corresponde a uma taxa de sucesso de aproximadamente 28% para as ambas as campanhas de triagem virtual. Vale ressaltar que cada campanha de triagem partiu de bases de dados diferentes (contendo milhões de moléculas) e que foram realizadas com ênfase em diferentes metodologias/abordagens.

Na **Tabela 4.6**, estão incluídos os valores de *ROCSTC* para os compostos obtidos pela Campanha 1, os valores de *QSAR score* para os compostos obtidos pela Campanha 2 e também os valores de *score* de *docking* usando o **Glide** (*Glide XP*) para todos eles. Aparentemente, não há uma correlação direta entre a qualidade dessas três métricas e as correspondentes % de inibição de GSK-3 $\beta$ . Por exemplo, o composto **LCQFGS04** mostrou a maior inibição de GSK-3 $\beta$ , mas não o maior valor de *Glide XP* de *score docking*. De qualquer forma, é possível verificar que todos os *hits* apresentaram valores variando de médio a alto para essas três métricas, o que de fato explica porque os selecionamos para obtenção, compra e teste.

É válido mencionar que os compostos **LCQFGS01** e **LCQFGS02** foram previamente citados em nossa publicação (SILVA et al., 2021b), com potencial de apresentar atividade em vista dos cálculos químico-quânticos empregados em comparação ao **composto 1**. Entretanto, após as suas avaliações biológicas foi verificado que estes não apresentaram atividades razoáveis, segundo os critérios aqui adotados.

Além disso, analisou-se as propriedades ADME/Tox preditas para os 5 *hits* selecionados, em comparação com o **composto 1** - o inibidor alostérico de GSK-3 $\beta$  extraído da literatura que fora majoritariamente empregado como molécula de referência ao longo do desenvolvimento deste trabalho. Todos os 5 compostos apresentaram maiores valores preditos para absorção oral humana (%HOA), bem como valores de  $^{QP}\log BB$  maiores que -0,5 (exceto para **LCQFGS15**).

Nota-se, de fato, que os compostos **LCQFGS13**, **15** e **16**, apresentaram valores fora deste critério de  $^{QP}\log BB > -0,5$ . Vale aqui mencionar que, durante a condução da Campanha 2, foi necessário adequar tal critério do **Qikprop**, para evitar a filtragem

excessiva de compostos com potencial mostrado pelas métricas de *QSAR score*. De qualquer modo, enfatiza-se que este critério (nota de corte) de  $^{QP}logBB$  é importante para o desenvolvimento de compostos destinados a atuar no SNC. Por outro lado, seus valores de *MW* e *PSA* estão dentro das faixas de valores desejáveis para drogas planejadas para a doença de Alzheimer, conforme proposto na literatura (WAGER et al., 2010; GHOSE et al., 2012; RANKOVIC, 2017).

No geral, apesar dos baixos valores percentuais de inibição de GSK-3 $\beta$  verificado para os 5 *hits*, houve uma clara melhora de suas propriedades ADME/Tox, quando comparados ao composto de referência. Essas propriedades são adequadas para candidatos a fármacos relacionados a doenças do SNC, incluindo DA, e foram planejadas com antecedência, durante as etapas de filtragem de cada campanha de triagem. Diante disso, é esperado que estes 5 *hits* possam servir no futuro, seguramente, de ponto de partida para propor otimizações estruturais que resultarão em *leads* mais potentes, para os quais, por sua vez, pretendemos obter os correspondentes valores experimentais de IC<sub>50</sub> de inibição de GSK-3 $\beta$ .

#### **4.2.3.1 Relações estrutura-atividade e diversidade estrutural**

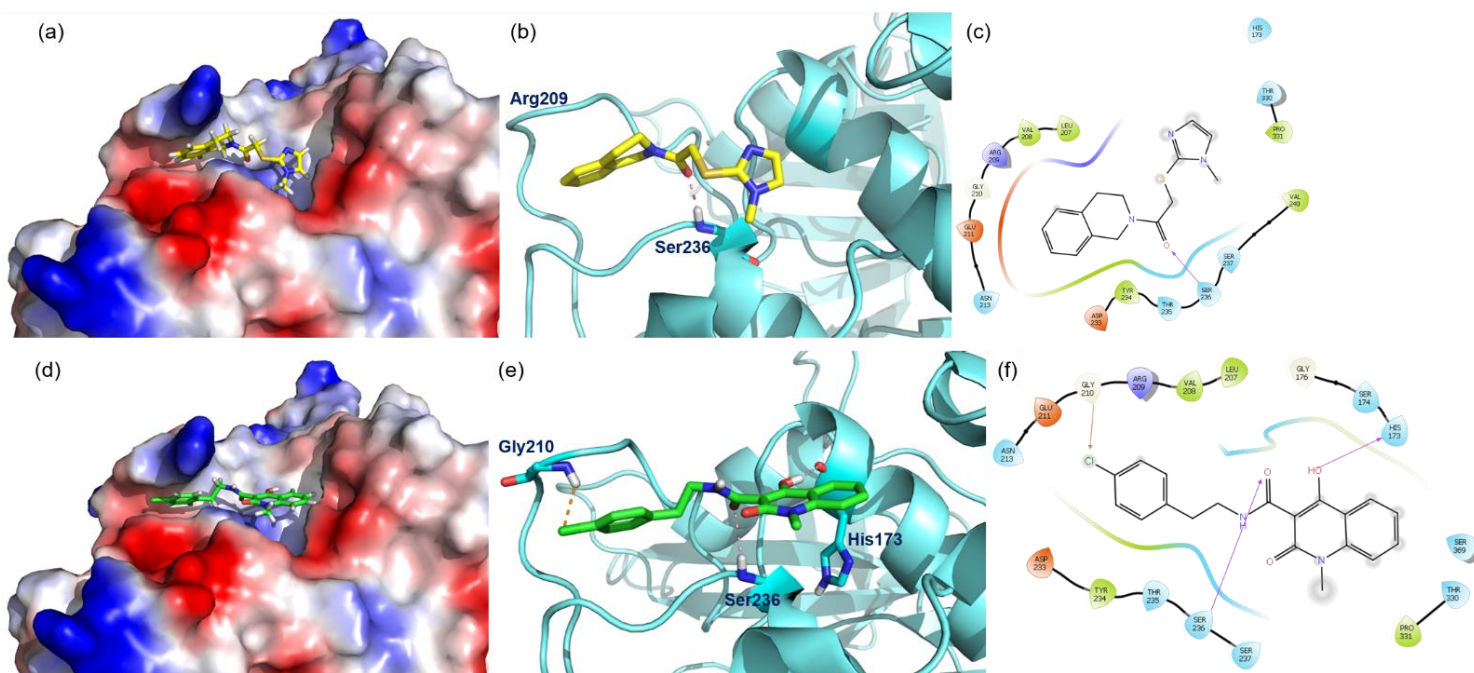
A fim de propor possíveis interações que poderiam ser responsáveis pela inibição observada para os 5 *hits* selecionados, foram propostas relações estrutura-atividade analisando-se as respectivas poses de *docking*. Como base, consideramos as 3 interações principais verificadas no sítio alostérico 7 da GSK-3 $\beta$  - mencionado aqui anteriormente na **Figura 4.11** e na literatura (PALOMO et al., 2017; SILVA et al., 2021b). A saber: ligação de hidrogênio com o N-H de Ser236; interação eletrostática ou cátion- $\pi$  de grupo aromático sob Arg209; e interações hidrofóbicas entre grupo alifático e a região hidrofóbica Thr330, Pro331 e Arg328.

Vale citar que, a partir dos 5 *hits* priorizados, considerou-se o **LCQFGS04** e **LCQFGS15** como os mais representativos *hits* obtidos pelas Campanha 1 e 2 de triagem virtual, respectivamente. Assim, na **Figura 4.22**, estão exibidas as interações preditas pelo protocolo refinado de *docking* do **Glide** para ambos estes compostos.

Para o **LCQFGS04**, observa-se uma interação de ligação de hidrogênio entre sua carbonila C=O e o N-H de amida do resíduo Ser236. Além disso, este se acomoda adequadamente, em termos de polaridade, conforme é observado no mapa de superfície do potencial eletrostático da cavidade alostérica mapeado na superfície.

Como esperado, o seu grupo de isoquinolona se situa sob o resíduo de Arg209 que é, de fato, um *hotspot* compatível com tal grupo. O grupo 1-metilimidazol não apresentou interações, entretanto, diante da flexibilidade conferida ao mesmo, via a cadeia alifática com o heteroátomo de enxofre, tal grupo metil pode interagir, possivelmente, com resíduos do bolsão hidrofóbico em proximidade.

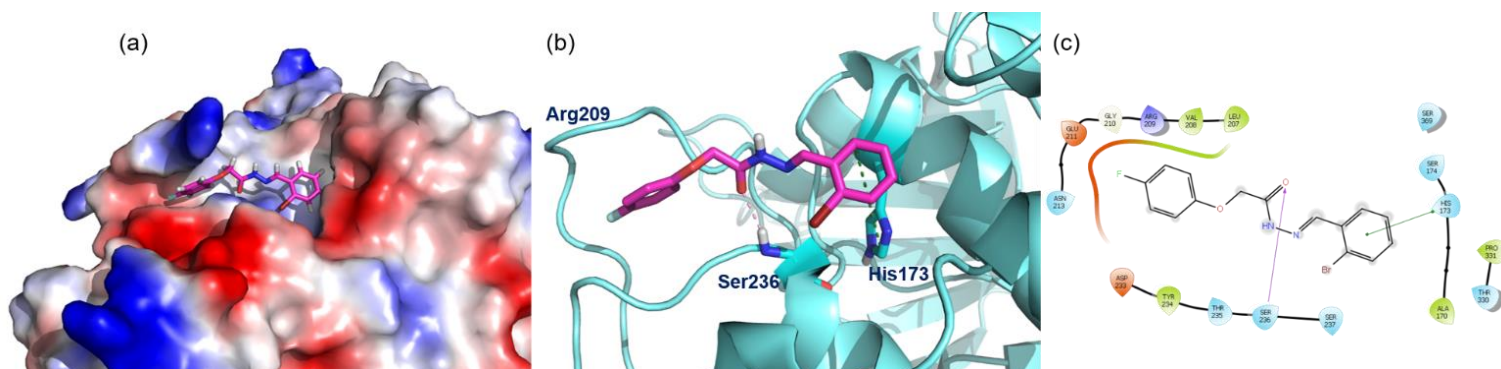
O **LCQFGS15** apresentou a mesma importante ligação de hidrogênio com a Ser236, bem como uma ligação de hidrogênio entre sua hidroxila OH e o C=O de His173. Vale notar que esta hidroxila corresponde a uma espécie mais favorável na correspondente tautomeria ceto-enólica. Também, apresentou uma terceira interação de halogênio entre seu átomo de cloro e o NH de Gly210. Esta interação é bastante interessante (e também relatado anteriormente) porque restringe o anel aromático sob o resíduo de Arg209, ainda que não modele corretamente, e assim apenas simule, a interação de transferência de carga existente.



**Figura 4.22** - Interações de *docking* observadas entre o sítio alostérico 7 da GSK-3 $\beta$  (1PYX.pdb) e os *hits* **LCQFGS04** (em representação *stick*, com carbonos em cor amarela) e **LCQFGS15** (em representação *stick*, com carbonos em cor verde). (a,d) mostrando a acomodação dos ligantes em fase (ou de acordo) com o mapa de potencial eletrostático mapeado na superfície; (b,e) principais resíduos responsáveis pelas interações-chave com os ligantes são destacados, com a enzima em representação *Cartoon*; (c,f) diagrama 2D das interações inibidor-enzima.

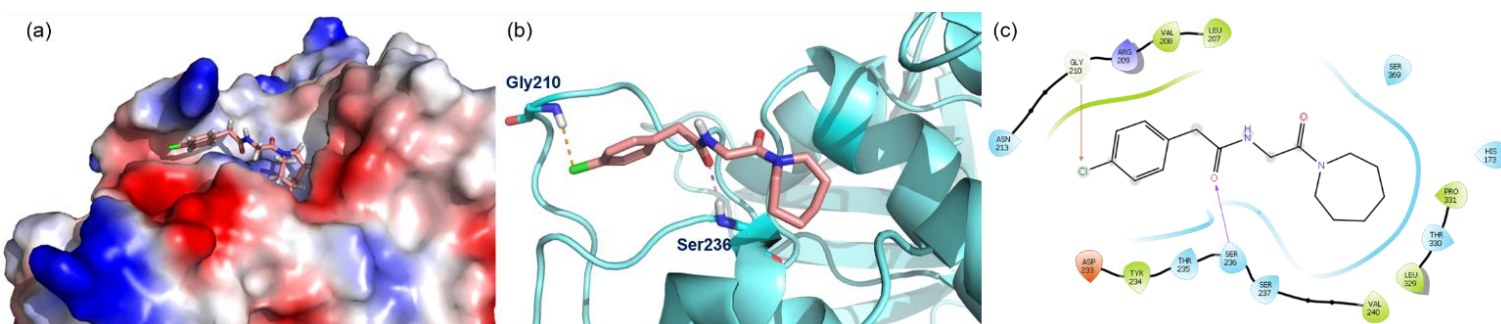
Os demais compostos *hits* obtidos neste trabalho foram também analisados com respeito as relações estrutura-atividade pelas respectivas interações e poses, conforme mostrado nas **Figura 4.23**, **Figura 4.24** e **Figura 4.25**.

Para o **LCQFGS03**, observa-se uma interação de ligação de hidrogênio entre sua carbonila C=O e o N-H de Ser236, bem como uma interação de empilhamento  $\pi$ - $\pi$  entre o bromobenzeno e His173. Em adição, apresenta um halogênio de F no anel aromático sob a Arg209 que poderia também realizar ligação de halogênio.



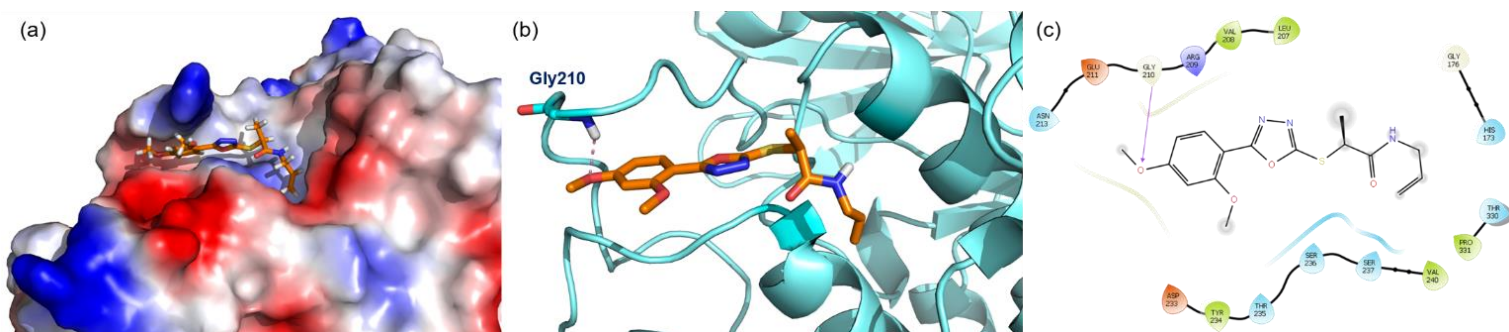
**Figura 4.23** - Interações de *docking* observadas entre o sítio alostérico 7 da GSK-3 $\beta$  (1PYX.pdb) e o *hit* **LCQFGS03**, (em representação *stick*, com carbonos em cor magenta). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em *Cartoon*; em (c), as interações inibidor-enzima são representadas em diagrama 2D.

Para o **LCQFGS06**, observa-se uma interação de ligação de hidrogênio entre sua carbonila C=O e o N-H de Ser236, bem como uma interação de halogênio entre o cloro e a Gly210, de modo similar ao observado para o **LCQFGS15**. Em adição, o anel de azepano saturado poderia cumprir o papel de ocupar o bolsão hidrofóbico próximo ao resíduo de Pro331.



**Figura 4.24** - Interações de *docking* observadas entre o sítio alostérico 7 da GSK-3 $\beta$  (1PYX.pdb) e o *hit* **LCQFGS06**, (em representação *stick*, com carbonos em cor salmão). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em *Cartoon*; em (c), as interações inibidor-enzima são representadas em diagrama 2D.

Para o **LCQFGS12**, observa-se uma interação de hidrogênio entre o grupo metoxi e o o resíduo de aminoácido Gly210, vizinho ao resíduo de Arg209 que forma a alça sobre tal anel aromático de dimetoxibenzeno. O composto ainda apresenta um grupo 1,3,4-oxadiazol, um heteroátomo de enxofre e um carbono quiral que não apresentaram interações explícitas; entretanto, sua cadeia terminal de carbonos insaturados pode interagir com o bolsão hidrofóbico.



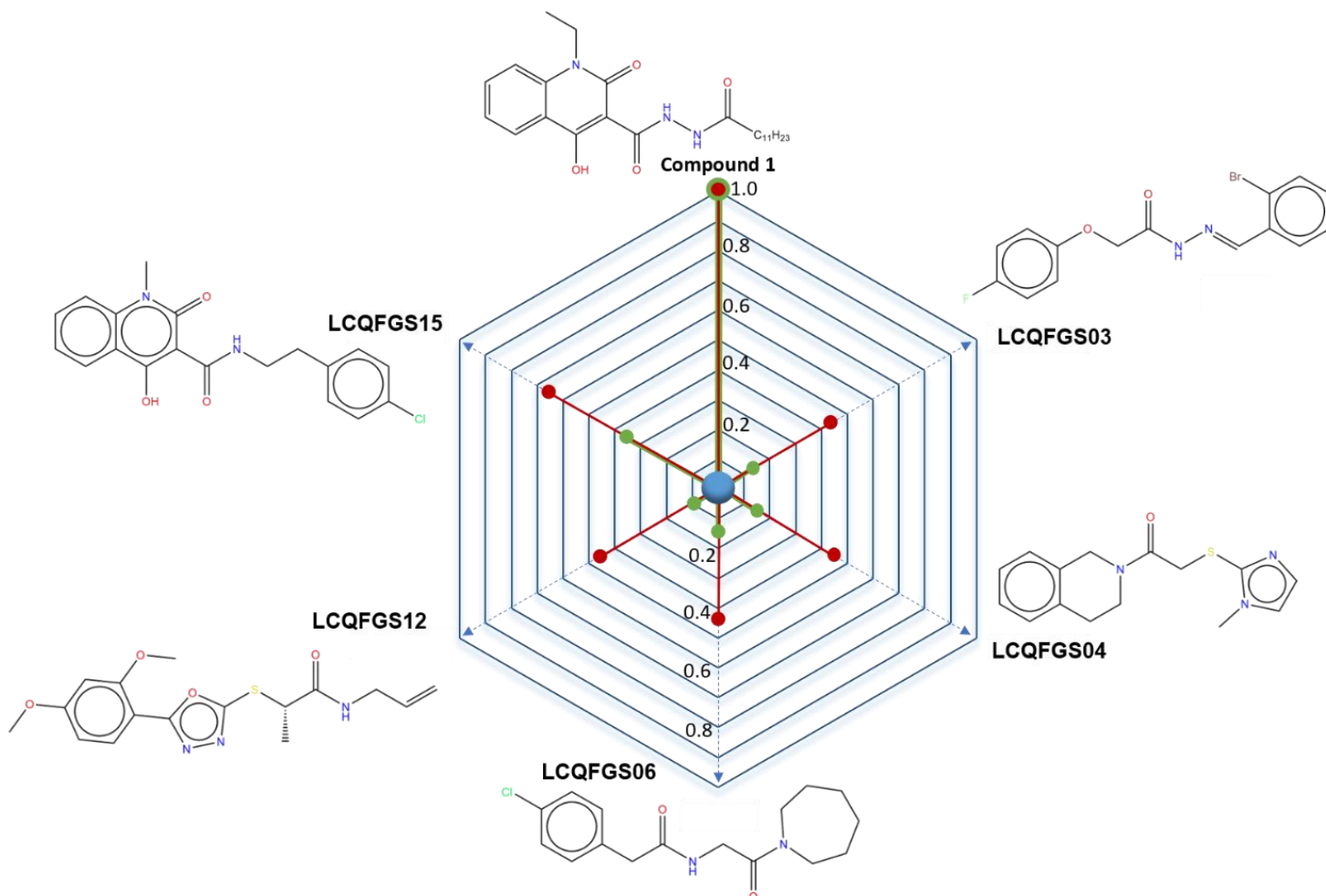
**Figura 4.25** - Interações de docking observadas entre o sítio alostérico 7 da GSK-3 $\beta$  (1PYX.pdb) e o hit **LCQFGS12**, (em representação *stick*, com carbonos em cor de laranja). Em (a), a enzima é representada por uma superfície com o potencial eletrostático nela mapeado; em (b), a enzima é representada em *Cartoon*; em (c), as interações inibidor-enzima são representadas em diagrama 2D.

Enfatiza-se que as hipóteses de relações estrutura-atividade, aqui embasadas nas interações observadas por *docking*, foram primordialmente apresentadas para os 5 *hits* mais representativos, devido a ultrapassarem a nota de corte de 10% de inibição da GSK-3 $\beta$ . Entretanto, nota-se que os demais compostos (dentre os 18 *hits* testados) não apresentaram % de inibição que distoam muito da nota de corte aqui estabelecida, então, eventualmente, também foram verificadas interações de *docking* com resíduos importantes do sítio 7 da GSK-3 $\beta$ . Estas são, portanto, apresentadas na **Figura 4.26**.





considerável na diversidade estrutural dos *hits* encontrados. O **LCQFGS15** foi o único composto com similaridade superior a 0,60 (*Morgan*) devido ao fato de possuir um anel quinolona quase idêntico ao da molécula referência **1**.



**Figura 4.27** - Gráfico de radar mostrando as similaridades entre o **composto 1** de referência e os cinco *hits*. Similaridades representadas em termos do cálculo dos respectivos índices Tanimoto usando dois *fingerprints* de similaridade (*MACCS*, em vermelho, e *Morgan*, em verde).

#### 4.2.3.2 Validação *in silico* da inibição alostérica de GSK-3 $\beta$

A fim de reforçar e validar a hipótese de que os 5 *hits* aqui encontrados podem, de fato, interagir na cavidade alostérica 7 da GSK-3 $\beta$ , realizamos uma avaliação adicional de *blind docking*. Para isso foi utilizado o software **AutoDock**, conforme descrito nas metodologias ([tópico 3.2.6](#)).

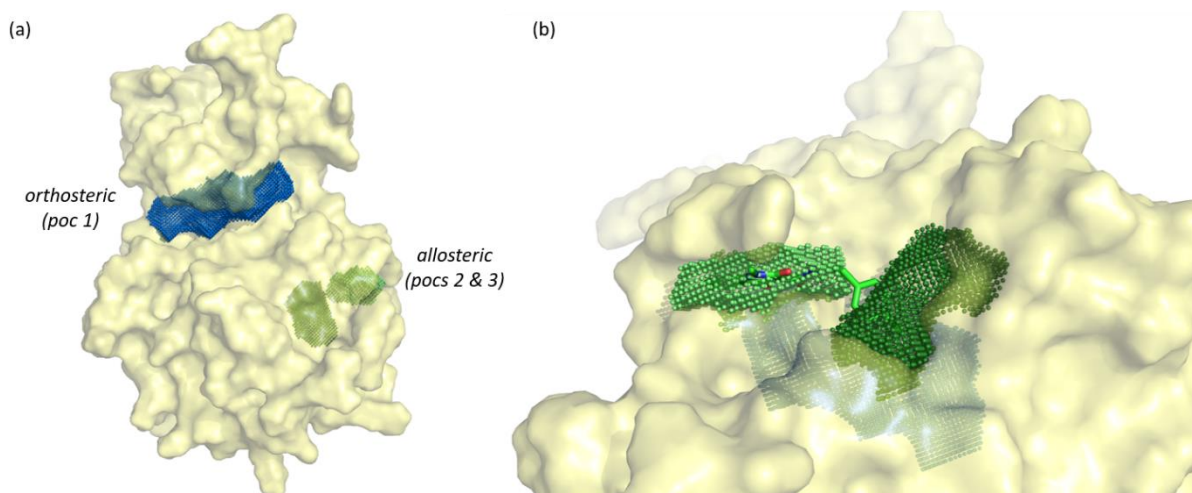
É reconhecido que o sítio ortostérico possui um volume maior quando comparada ao sítio alostérico. No entanto, considerando que o sítio alostérico foi 'visto' pelo software **DogSiteScorer** como dividido em duas subcavidades (*pocs 2 e 3*), se

compararmos o sítio ortostérico com a soma destas subcavidades (em termos de valores de *druggability score*), o sítio alostérico realmente supera o ortostérico.

**Tabela 4.7** - Melhor pose de *blind docking* para cada composto, correspondentes valores de *docking score*, bem como valores de ocupação de sítio (PC) na cavidade ortostérica (poc 1) e alostérica (pocs 2 e 3) da enzima GSK-3 $\beta$ .

Composto	Sítio/pose ortostérico		Sítio/pose alostérico	
	<i>Docking score</i>	PC (poc 1)	<i>Docking score</i>	PC (pocs 2 & 3)
1	-6.91	47.32	<b>-7.24</b>	<b>53.6 &amp; 75.99</b>
LCQFGS03	-6.98	32.36	<b>-7.17</b>	<b>42.39 &amp; 82.70</b>
LCQFGS04	<b>-6.59</b>	24.60	-6.30	<b>0.03 &amp; 85.44</b>
LCQFGS06	-7.19	34.33	<b>-7.31</b>	<b>46.35 &amp; 72.99</b>
LCQFGS12	-6.13	25.52	<b>-6.36</b>	<b>19.32 &amp; 86.81</b>
LCQFGS15	<b>-7.62</b>	31.32	-7.47	<b>26.98 &amp; 91.83</b>

*Docking score*: valores de *score* em kcal/mol preditos pelas simulações de *blind docking* do **Autodock**; PC: valores de ocupação do sítio (*pocket coverage*) preditos para cada pose, em cada *poc* (subcavidade), usando o software **DoGSiteScorer**.



**Figura 4.28** – (a) Cavidades preditas pelo **DoGSiteScorer**, na estrutura da GSK-3 $\beta$  (representada em superfície), sendo a ortostérica (*poc 1*) representada por esferas azuis e a alostérica (*pocs 2 & 3*) por esferas verdes. (b) exemplo mostrando a melhor pose obtida por *blind docking* para **LCQFGS15**, permitindo a visualização da sua ocupação do sítio (PC, *pocket coverage*).

Conforme calculado pelo **DogSiteScorer**, o sítio ortostérico (*poc 1*) tem um volume de 930,96 Å<sup>3</sup> e um *druggability score* de 0,764, enquanto o sítio alostérico (constituído pelos *pocs 2 e 3*) tem um volume de 464,29 Å<sup>3</sup> (286,58 e 177,71 Å<sup>3</sup>) e um *druggability score* de 0,772 (0,550 e 0,222). Isso pode sugerir duas coisas: (i) que a cavidade alostérica pode apresentar uma drogabilidade comparável à cavidade ortostérica (corroborando o estudo prévio de detecção de cavidade da *PARTE I* desta

Tese); (ii) que, proporcionalmente, a ‘ocupação do sítio’ (ou cobertura do sítio, em termos de volume, do inglês *pocket coverage*, PC) dos compostos pode representar uma boa medida e indicativa da preferência do composto em interagir com determinado sítio. Com isso em mente, analisamos os resultados de *blind docking* com foco no parâmetro PC - em complemento aos *docking score* -, o que deve indicar uma métrica razoável para descrever qual sítio tem maior probabilidade de acomodar os compostos *hits*.

Sabendo-se que o **composto 1** é um modulador alostérico real, que se liga ao sítio alostérico 7 da GSK-3 $\beta$ , usamos seus resultados como referência para comparar com os dados obtidos para os 5 *hits*. Os valores de *docking score* e de PC para **1** foram maiores para sua respectiva pose alostérica do que para a ortostérica (vide **Figura 4.28**). Em vista disso, os resultados resumidos na **Tabela 4.7** mostraram que todos os 5 *hits* apresentaram maiores valores de *docking score* em direção ao sítio alostérico, exceto para **LCQFGS04** e **15**. No entanto, considerando seus valores de PC, observamos que todos eles mostraram preferência em direção ao sítio alostérico (respectivamente constituído pelos *pocs* 2 & 3).

#### 4.2.4 Considerações finais – PARTE II

Os estudos realizados nesta parte da Tese demonstraram que foi possível obter *hits* razoavelmente potentes e com uma ampla diversidade estrutural, mesmo com uma quantidade limitada de dados na literatura para se usar como material de partida, o que atribuiu um caráter bastante desafiador às campanhas conduzidas.

Isso foi viável empregando-se, de forma integrada, metodologias de *LBVS* (usando similaridade 3D baseada em forma, bem como modelos preditivos de *QSAR* via aprendizado de máquina) e de *SBVS* (*docking*), e partindo-se de diferentes bases de dados. Observamos que variar a abrangência das metodologias nos estudos de triagem virtual e buscar um consenso entre elas pode ser uma estratégia promissora que deve ser cada vez mais explorada e aplicada para a identificação de novos *hits*.

Embora fosse esperado que os valores de inibição de GSK-3 $\beta$ , aqui expressos em %, fossem suficientes para proceder com a determinação dos respectivos valores de IC<sub>50</sub>, consideramos que tais resultados obtidos através dos ensaios experimentais realizados forneceram dados interessantes em termos de análise de relações estrutura-atividade, bem como de diversidade estrutural. De fato, o objetivo de ampliar

a diversidade estrutural foi alcançado com sucesso conforme demonstrado no tópico 4.2.3.1.

Além disso, os compostos foram selecionados com vista à predição de suas propriedades ADME/Tox mais favoráveis, no contexto de candidatos a fármacos para o SNC. Desta forma, os cinco *hits* aqui revelados são bastante atrativos para a continuação dos nossos estudos. Para isso, possivelmente incluiremos propostas de otimizações moleculares visando a manutenção dos parâmetros físico-químicos/farmacocinéticos mais apropriados, bem como aumentando as suas afinidades com o sítio alostérico da enzima - assim verificadas nos respectivos ensaios de atividade biológica como inibidores alostéricos de GSK-3 $\beta$ , com interesse no desenvolvimento de novos e seletivos fármacos para futuro tratamento da DA.

## 5 CONCLUSÕES E PERSPECTIVAS

Os estudos realizados nesta Tese, subdivididos em *PARTE I* e *II*, foram realizados sobretudo visando o planejamento computacional e subsequente obtenção de novos e diversos (estruturalmente) inibidores alostéricos de GSK-3 $\beta$ , direcionados ao tratamento da doença de Alzheimer (DA).

A GSK-3 $\beta$  é, sem dúvida, uma proteína-alvo extremamente relevante associada a múltiplas vias fisiopatológicas da DA. A sua modulação alostérica foi proposta anteriormente na literatura como uma estratégia vantajosa em vista de sua menor chance de causar efeitos colaterais, entretanto carecia de aprofundamentos.

Na *PARTE I*, portanto, realizamos um intenso estudo *in silico* considerando o principal e mais representativo inibidor alostérico de GSK-3 $\beta$ , a fim de reforçar em qual cavidade alostérica o mesmo deve se ligar, bem como qual a pose que o mesmo deve assumir para interagir com a enzima. Para isso, aplicou-se metodologias de detecção de cavidades, mapeamento de superfície, análise do *shape* 3D do ligante e de simulações *docking* (com respectivas validações de *queries* e protocolos) e dinâmica molecular.

Vale destacar que a aplicação destas variadas metodologias, constitui-se de uma abordagem importante para o enriquecimento de dados sobre o tema, que carecia de dados para a efetuação de campanhas de planejamento. Abordagens como esta vêm sendo cada vez mais estimuladas pela comunidade científica.

Em geral, foi possível confirmar que a cavidade 7 representa o sítio alostérico mais propenso da GSK-3 $\beta$  a acomodar ligantes, conseqüentemente modulando a sua atividade. A suposta pose bioativa do composto **1** na cavidade alostérica 7 apresenta interações significativas e coerentes - que corroboram nossos estudos prévios e também da literatura -, permitindo, assim, o seu uso em estudos de prospecção. Portanto, os resultados foram consistentes o suficiente para fornecer achados relevantes, aprofundando assim o conhecimento sobre a modulação alostérica da GSK-3 $\beta$ .

Na *PARTE II* foram realizadas duas campanhas de VS para identificar novos inibidores alostéricos de GSK-3 $\beta$  com diversidade estrutural, aproveitando e validando, desta forma, os resultados gerados na *PARTE I*. Vale mencionar que, inicialmente, tínhamos o intuito de conduzir mais e diversas campanhas de VS,

entretanto avaliamos que as Campanhas 1 e 2, aqui performadas, já englobavam variações consideráveis de bases de dados e de metodologias para prospecção.

Na primeira, portanto, foram usadas bases de dados diversas, somando quase 40 milhões de compostos, e enfatizou-se a metodologia de similaridade 3D por forma para filtrar os compostos. Na segunda, foram desenvolvidos modelos preditivos de QSAR via aprendizagem de máquina para aplicação na triagem, partindo-se de bases com mais de 10 milhões de compostos. Ambas passaram por filtros de predições ADME/Tox e avaliação de interações de *docking* para seleção final dos compostos.

Dezoito compostos foram selecionados e avaliados por ensaios biológicos. Destes, cinco foram considerados *hits* ativos frente a inibição da GSK-3 $\beta$ , nesta avaliação preliminar. Destaca-se, entretanto, que os compostos ativos usados como referência apresentam valores de IC<sub>50</sub> ~ 3  $\mu$ M. Assim, uma vez que este trabalho é um dos pioneiros na aplicação de VS dentro deste contexto, dificilmente seriam alcançados *hits* com inibição igual ou melhor a dos compostos de referência. Em adição, considera-se que a pequena quantidade de dados disponíveis (poucos inibidores conhecidos, bem como a inexistência de complexo cristalográfico ligante-receptor) atribui um caráter bastante desafiador para este trabalho.

Em todo caso, foram obtidos resultados interessantes em termos de análise de relações estrutura-atividade, bem como de diversidade estrutural para os cinco *hits*. Além disso, como tiveram as suas propriedades ADME/Tox preditas antecipadamente, os cinco *hits* se apresentam mais favoráveis no contexto de candidatos a fármacos para o SNC, quando comparado aos compostos de referência.

Importantemente, foi possível concluir que estes *hits* obtidos na *PARTE II* satisfazem as principais interações reveladas pela *PARTE I* (e corroboradas pela literatura). Assim, os dados obtidos pelas metodologias aplicadas em consenso ao longo da *PARTE I*, para elucidar tanto a cavidade alostérica 7 da GSK-3 $\beta$  assim como a pose bioativa do inibidor de referência, foram determinantes para a condução bem sucedida das campanhas de VS da *PARTE II*, que culminou na identificação destas novas estruturas.

Portanto, os cinco *hits* aqui revelados são bastante promissores em vista da continuação dos nossos estudos. Intencionamos, no futuro, desenvolver propostas de otimizações moleculares, visando a manutenção das mais apropriadas propriedades ADME/Tox, assim como melhorar as suas potências como inibidores, de fato, alostéricos de GSK-3 $\beta$ . Assim, esperamos continuar contribuindo, via estratégias

promissoras, para o desenvolvimento e descoberta de potenciais candidatos a fármacos para o tratamento da doença de Alzheimer.



## REFERÊNCIAS

- ABRAHAM, D. J. Drug Discovery in Academia – A Case Study. In: OPREA, T. I. (Ed.). **Cheminformatics in Drug Discovery**. 2005. ed. [s.l.] Wiley-VCH Verlag GmbH & Co. KGaA, 2005. p. 457–484.
- AFANTITIS, A. et al. Ligand - based virtual screening procedure for the prediction and the identification of novel  $\beta$ -amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. **European Journal of Medicinal Chemistry**, v. 46, n. 2, p. 497–508, fev. 2011.
- ALEXANDER, D. L. J.; TROPSHA, A.; WINKLER, D. A. Beware of  $R^2$ : Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. **Journal of Chemical Information and Modeling**, v. 55, n. 7, p. 1316–1322, 2015.
- ALQAHTANI, S. In silico ADME-Tox modeling: progress and prospects. **Expert Opinion on Drug Metabolism & Toxicology**, v. 13, n. 11, p. 1147–1158, 2 nov. 2017.
- ALTALIB, M. K.; SALIM, N. Hybrid-Enhanced Siamese Similarity Models in Ligand-Based Virtual Screen. **Biomolecules**, v. 12, n. 11, p. 1719, 20 nov. 2022.
- ALVES, V. et al. Quimioinformática: Uma Introdução. **Química Nova**, v. 41, n. 2, p. 202–212, 2017.
- ALVES, V. M. et al. Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. **Toxicology and Applied Pharmacology**, v. 284, n. 2, p. 262–272, 15 abr. 2015.
- AMASHITA, F. Y.; ASHIDA, M. H. Review In Silico Approaches for Predicting ADME Properties of Drugs. v. 19, n. 5, p. 327–338, 2004.
- ANDERSON, A. C. The Process of Structure-Based Drug Design. **Chemistry & Biology**, v. 10, n. 9, p. 787–797, set. 2003.
- ANDREWS, P. R.; CRAIK, D. J.; MARTIN, J. L. Functional group contributions to drug-receptor interactions. **Journal of medicinal chemistry**, v. 27, n. 12, p. 1648–57, dez. 1984.
- APLIN, C. et al. Evolving Experimental Techniques for Structure-Based Drug Design. **The Journal of Physical Chemistry B**, v. 126, n. 35, p. 6599–6607, 8 set. 2022.
- AURENHAMMER, F. Voronoi diagrams—a survey of a fundamental geometric data structure. **ACM Computing Surveys**, v. 23, n. 3, p. 345–405, 1 set. 1991.
- AUTODOCK. **User Guide - Version 4.2 - Theory: Overview of the Free Energy Scoring Function**. Disponível em: <[https://autodock.scripps.edu/wp-content/uploads/sites/56/2022/04/AutoDock4.2.6\\_UserGuide.pdf](https://autodock.scripps.edu/wp-content/uploads/sites/56/2022/04/AutoDock4.2.6_UserGuide.pdf)>. Acesso em: 15 fev. 2014.
- AVRAHAMI, L. et al. GSK-3 inhibition: Achieving moderate efficacy with high selectivity. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1834, n. 7, p. 1410–1414, 2013.
- BALLESTER, P. J.; FINN, P. W.; RICHARDS, W. G. Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. **Journal of Molecular Graphics and Modelling**, v. 27, n. 7, p. 836–845, abr. 2009.
- BARNUM, D. et al. Identification of Common Functional Configurations Among Molecules. **Journal of Chemical Information and Computer Sciences**, v. 36, n. 3, p. 563–571, 1 jan. 1996.
- BARONI, M. et al. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. **Journal of Chemical Information and Modeling**, v. 47, n. 2, p. 279–294, 1 mar. 2007.
- BARREIRO, E. J.; FRAGA, C. A. M. **Química Medicinal: as bases moleculares da ação dos fármacos**. 3a. ed. Porto Alegre: Artmed, 2015.
- BASCIU, A. et al. No dance, no partner! A tale of receptor flexibility in docking and virtual screening. In: **Annual Reports in Medicinal Chemistry**. [s.l.] Academic Press, 2022. 59p. 43–97.
- BASITH, S. et al. Exploring G protein-coupled receptors (GPCRs) ligand space via cheminformatics approaches: Impact on rational drug design. **Frontiers in Pharmacology**, v. 9, n. MAR, p. 1–26, 2018.
- BAUMANN, D.; BAUMANN, K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. **Journal of Cheminformatics**, v. 6, n. 1, p. 1–19, 2014.
- BERENGER, F. et al. Lean-Docking: Exploiting Ligands' Predicted Docking Scores to Accelerate Molecular Docking. **Journal of Chemical Information and Modeling**, v. 61, n. 5, p. 2341–2352, 24 maio 2021.
- BERO, S. A. et al. Similarity Measure for Molecular Structure: A Brief Review. **Journal of Physics: Conference Series**, v. 892, n. 1, p. 012015, set. 2017.
- BERRHAIL, F.; BELHADEF, H. Genetic Algorithm-based Feature Selection Approach for Enhancing the Effectiveness of Similarity Searching in Ligand-based Virtual Screening. **Current Bioinformatics**, v. 15, n. 5, p. 431–444, 14 out. 2020.

- BERTHOLD, M. R. et al. KNIME: The Konstanz Information Miner. In: **Studies in Classification, Data Analysis, and Knowledge Organization**. [s.l.: s.n.]p. 319–326.
- BERTRAND, J. A. et al. Structural characterization of the GSK-3 $\beta$  active site using selective and non-selective ATP-mimetic inhibitors. **Journal of Molecular Biology**, v. 333, n. 2, p. 393–407, 2003.
- BEUREL, E.; GRIECO, S. F.; JOPE, R. S. Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. **Pharmacology & Therapeutics**, v. 148, n. 303, p. 114–131, 10 abr. 2015.
- BICKERTON, G. R. et al. Quantifying the chemical beauty of drugs. **Nature Chemistry**, v. 4, n. 2, p. 90–98, 2012.
- BIDON-CHANAL, A. et al. Evidence for a new binding mode to GSK-3: Allosteric regulation by the marine compound palinurin. **European Journal of Medicinal Chemistry**, v. 60, p. 479–489, 2013.
- BLEICHER, K. H. et al. Hit and lead generation: Beyond high-throughput screening. **Nature Reviews Drug Discovery**, v. 2, n. 5, p. 369–378, 2003.
- BOEHM, M. et al. Similarity Searching and Scaffold Hopping in Synthetically Accessible Combinatorial Chemistry Spaces. **Journal of Medicinal Chemistry**, v. 51, n. 8, p. 2468–2480, abr. 2008.
- BOHACEK, R. S.; MCMARTIN, C.; GUIDA, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. **Medicinal Research Reviews**, v. 16, n. 1, p. 3–50, jan. 1996.
- BREIMAN, L. Random Forests. **Machine Learning 2001 45:1**, v. 45, n. 1, p. 5–32, out. 2001.
- BROGI, S. et al. First dual AK/GSK-3 $\beta$  inhibitors endowed with antioxidant properties as multifunctional, potential neuroprotective agents. **European Journal of Medicinal Chemistry**, v. 138, p. 438–457, 2017.
- BROWN, F. K. et al. The evolution of drug design at Merck Research Laboratories. **Journal of Computer-Aided Molecular Design**, v. 31, n. 3, p. 255–266, 2017.
- BUCKLE, D. R. et al. Glossary of Terms Used in Medicinal Chemistry Part II (IUPAC Recommendations 2013). In: **Annual Reports in Medicinal Chemistry**. [s.l.: s.n.]p. 387–418.
- BUDD HAEBERLEIN, S.; HARRIS, T. Promising Targets for the Treatment of Neurodegenerative Diseases. **Clinical Pharmacology & Therapeutics**, v. 98, n. 5, p. 492–501, 2015.
- BURA, E.; GASTWIRTH, J. L. The Binary Regression Quantile Plot: Assessing the Importance of Predictors in Binary Regression Visually. **Biometrical Journal**, v. 43, n. 1, p. 5–21, fev. 2001.
- C. HANSCH, T. F. p-sigma-pi analysis. A Method for the Correlation of Biological Activity and Chemical Structure. **Correlation of biological activity and chemical structure**, v. 86, n. 1, p. 1616–1626, 1964.
- CANDIDO, S. et al. The analysis of miRNA expression profiling datasets reveals inverse microRNA patterns in glioblastoma and Alzheimer's disease. **Oncology Reports**, v. 42, n. 3, p. 911–922, 28 jun. 2019.
- CARUSO, A. et al. Stress as risk factor for Alzheimer's disease. **Pharmacological Research**, v. 132, n. April, p. 130–134, 2018.
- CCDC. **GOLD - Protein Ligand Docking Software** (<https://www.ccdc.cam.ac.uk/solutions/csd-discovery/Components/Gold/>), 2022.
- CERETO-MASSAGUÉ, A. et al. Molecular fingerprint similarity search in virtual screening. **Methods**, v. 71, n. C, p. 58–63, jan. 2015.
- CHAN, S. L. MolAlign: an algorithm for aligning multiple small molecules. **Journal of Computer-Aided Molecular Design**, v. 31, n. 6, p. 523–546, 1 jun. 2017.
- CHANG, Y. et al. A Guide to In Silico Drug Design. **Pharmaceutics**, v. 15, n. 1, p. 49, 23 dez. 2022.
- CHARIFSON, P. S. et al. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. **Journal of Medicinal Chemistry**, v. 42, n. 25, p. 5100–5109, dez. 1999.
- CHÁVEZ THIELEMANN, H. et al. From GROMACS to LAMMPS: GRO2LAM. **Journal of Molecular Modeling**, 2019.
- CHEMBRIDGE. **The gold standard in small molecule screening libraries and building blocks** ([https://www.chembridge.com/screening\\_libraries/index.php](https://www.chembridge.com/screening_libraries/index.php)).
- CHEN, L.; JIAO, J.; ZHANG, Y. Therapeutic approaches for improving cognitive function in the aging brain. **Frontiers in Neuroscience**, v. 16, p. 1060556, 8 dez. 2022.
- CHERKASOV, A. et al. QSAR Modeling: Where Have You Been? Where Are You Going To? **Journal of Medicinal Chemistry**, v. 57, n. 12, p. 4977–5010, 26 jun. 2014.
- CLARK, D. E. Virtual Screening: Is Bigger Always Better? Or Can Small Be Beautiful? **Journal of Chemical Information and Modeling**, v. 60, n. 9, p. 4120–4123, 28 set. 2020.

- CLARK, R. D. et al. Consensus scoring for ligand/protein interactions. **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 281–295, 1 jan. 2002.
- CLEVES, A. E.; JOHNSON, S. R.; JAIN, A. N. Electrostatic-field and surface-shape similarity for virtual screening and pose prediction. **Journal of Computer-Aided Molecular Design**, v. 33, n. 10, p. 865–886, 24 out. 2019.
- CORMIER, K. W.; WOODGETT, J. R. Recent advances in understanding the cellular roles of GSK-3. **F1000Research**, v. 6, n. 0, p. 167, 20 fev. 2017.
- CRAMER, R. D. The inevitable QSAR renaissance. **Journal of Computer-Aided Molecular Design**, v. 26, n. 1, p. 35–38, 30 jan. 2012.
- CRAMER, 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**, 1988.
- DA SILVA, C. H. T. de P.; TAFT, C. A. 3D descriptors calculation and conformational search to investigate potential bioactive conformations, with application in 3D-QSAR and virtual screening in drug design. **Journal of Biomolecular Structure and Dynamics**, v. 35, n. 13, p. 2966–2974, 2017.
- DA SILVA, C. H. T. P.; CARVALHO, I.; TAFT, C. A. Virtual Screening, Molecular Interaction Field, Molecular Dynamics, Docking, Density Functional, and ADMET Properties of Novel AChE Inhibitors in Alzheimer's Disease. **Journal of Biomolecular Structure and Dynamics**, v. 24, n. 6, p. 515–523, 2007.
- DALLAKYAN, S.; OLSON, A. J. Small-Molecule Library Screening by Docking with PyRx. In: **Methods in molecular biology (Clifton, N.J.)**. [s.l.] Methods Mol Biol, 2015. 1263p. 243–250.
- DARSHIT, B. S. et al. Identification and in vitro evaluation of new leads as selective and competitive glycogen synthase kinase-3 $\beta$  inhibitors through ligand and structure based drug design. **Journal of Molecular Graphics and Modelling**, v. 53, p. 31–47, 2014.
- DAVIES, S. P. et al. Specificity and mechanism of action of some commonly used protein kinase inhibitors. **Biochemical Journal**, v. 351, n. Pt 1, p. 95, 10 out. 2000.
- DAYLIGHT. **Chemical Information Systems, Inc.** Disponível em: <<https://www.daylight.com/>>.
- DE ALMEIDA, J. R. et al. Discovery of novel dual acetylcholinesterase inhibitors with antifibrillogenic activity related to Alzheimer's disease. **Future Medicinal Chemistry**, v. 10, n. 9, p. 1037–1053, maio 2018.
- DE OLIVEIRA, O. V. et al. Repurposing approved drugs as inhibitors of SARS-CoV-2 S-protein from molecular modeling and virtual screening. **Journal of Biomolecular Structure and Dynamics**, v. 0, n. 0, p. 1–10, 2 jun. 2020.
- DE, P. et al. Prediction reliability of QSAR models: an overview of various validation tools. **Archives of Toxicology**, v. 96, n. 5, p. 1279–1295, 10 maio 2022.
- DE SIMONE, A. et al. Glycogen Synthase Kinase 3 $\beta$ : A New Gold Rush in Anti-Alzheimer's Disease Multitarget Drug Discovery? **Journal of Medicinal Chemistry**, v. 64, n. 1, p. 26–41, 14 jan. 2021.
- DEARDEN, J. C.; CRONIN, M. T. D.; KAISER, K. L. E. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). **SAR and QSAR in Environmental Research**, v. 20, n. 3–4, p. 241–266, abr. 2009.
- DEGOEY, D. A. et al. Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. **Journal of Medicinal Chemistry**, v. 61, n. 7, p. 2636–2651, 12 abr. 2018.
- DHILLON, S. Aducanumab: First Approval. **Drugs**, v. 81, n. 12, p. 1437–1443, 29 ago. 2021.
- DI MARTINO, R. M. C. et al. Versatility of the Curcumin Scaffold: Discovery of Potent and Balanced Dual BACE-1 and GSK-3 $\beta$  Inhibitors. **Journal of Medicinal Chemistry**, v. 59, n. 2, p. 531–544, 2016.
- DO AMARAL, A. T. et al. A evolução da Química Medicinal no Brasil: Avanços nos 40 anos da Sociedade Brasileira de Química. **Química Nova**, v. 40, n. 6, p. 694–700, 2017.
- DOMÍNGUEZ, J. M. et al. Evidence for irreversible inhibition of glycogen synthase kinase-3 $\beta$  by tideglusib. **Journal of Biological Chemistry**, v. 287, n. 2, p. 893–904, 2012.
- DOS SANTOS MAIA, M. et al. Consensus Analyses in Molecular Docking Studies Applied to Medicinal Chemistry. **Mini-Reviews in Medicinal Chemistry**, v. 20, n. 14, p. 1322–1340, 1 set. 2020.
- DOWEYKO, A. M. 3D-QSAR illusions. **Journal of Computer-Aided Molecular Design**, v. 18, n. 7–9, p. 587–596, 2004.
- DUELEN, R. et al. Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development. In: **Introduction to Biotech Entrepreneurship: From Idea to Business**. Cham: Springer International Publishing, 2019. p. 89–128.
- DULSAT, J. et al. Evaluation of Free Online ADMET Tools for Academic or Small Biotech Environments. **Molecules**, v. 28, n. 2, p. 776, 12 jan. 2023.

- DUMURGIER, J.; TZOURIO, C. Epidemiology of neurological diseases in older adults. **Revue Neurologique**, v. 176, n. 9, p. 642–648, 2020.
- DURANT, J. L. et al. Reoptimization of MDL Keys for Use in Drug Discovery. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 6, p. 1273–1280, 1 nov. 2002.
- DUTTA, S.; BOSE, K. Remodelling structure-based drug design using machine learning. **Emerging Topics in Life Sciences**, v. 5, n. 1, p. 13–27, 14 maio 2021.
- ECKERT, H.; BAJORATH, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. **Drug discovery today**, v. 12, n. 5–6, p. 225–33, mar. 2007.
- EMPEREUR-MOT, C. et al. Predictiveness curves in virtual screening. **Journal of Cheminformatics**, v. 7, n. 1, p. 1–17, 2015.
- EMPEREUR-MOT, C.; ZAGURY, J.-F.; MONTES, M. Screening Explorer—An Interactive Tool for the Analysis of Screening Results. **Journal of Chemical Information and Modeling**, v. 56, n. 12, p. 2281–2286, 27 dez. 2016.
- ENAMINE. **Screening Collection**. Disponível em: <<https://enamine.net/compound-collections/screening-collection>>.
- ENSLEIN, K.; CRAIG, P. N. A toxicity estimation model. **Journal of Environmental Pathology and Toxicology**, 1978.
- ERKKINEN, M. G.; KIM, M.; GESCHWIND, M. D. Major Neurodegenerative Diseases. 2018.
- FAN, N.; HIRTE, S.; KIRCHMAIR, J. Maximizing the Performance of Similarity-Based Virtual Screening Methods by Generating Synergy from the Integration of 2D and 3D Approaches. **International Journal of Molecular Sciences**, v. 23, n. 14, p. 7747, 13 jul. 2022.
- FASSIO, A. V. et al. Prioritizing Virtual Screening with Interpretable Interaction Fingerprints. **Journal of Chemical Information and Modeling**, v. 62, n. 18, p. 4300–4318, 26 set. 2022.
- FEDERICO, L. B. et al. Identification of Novel A $\beta$ -Tubulin Modulators with Antiproliferative Activity Directed to Cancer Therapy Using Ligand and Structure-Based Virtual Screening. **International Journal of Biological Macromolecules**, v. 165, p. 3040–3050, dez. 2020.
- FEDERICO, L. B. et al. Key Aspects for Achieving Hits by Virtual Screening Studies. In: [s.l.] Springer, Cham, 2021a. p. 455–487.
- FEDERICO, L. B. et al. Potential colchicine binding site inhibitors unraveled by virtual screening, molecular dynamics and MM/PBSA. **Computers in Biology and Medicine**, v. 137, p. 104817, 1 out. 2021b.
- FEHER, M. Consensus scoring for protein–ligand interactions. **Drug Discovery Today**, v. 11, n. 9–10, p. 421–428, maio 2006.
- FERREIRA, L. et al. Molecular Docking and Structure-Based Drug Design Strategies. **Molecules**, v. 20, n. 7, p. 13384–13421, 22 jul. 2015.
- FERRERO, H. et al. Adrenomedullin, a Novel Target for Neurodegenerative Diseases. **Molecular Neurobiology**, n. November 2017, p. 1–16, 2018.
- FIGUERAS, J. Morgan revisited. **Journal of Chemical Information and Computer Sciences**, v. 33, n. 5, p. 717–718, 1 set. 1993.
- FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. **Journal of Chemical Information and Modeling**, v. 50, n. 7, p. 1189–1204, 26 jul. 2010.
- FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. **Journal of Chemical Information and Modeling**, v. 56, n. 7, p. 1243–1252, 25 jul. 2016.
- FREE, S. M.; WILSON, J. W. A Mathematical Contribution to Structure-Activity Studies. **Journal of Medicinal Chemistry**, v. 7, n. 4, p. 395–399, 1 jul. 1964.
- 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, 1 mar. 2004.
- FRIESNER, R. A. et al. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. **Journal of Medicinal Chemistry**, v. 49, n. 21, p. 6177–6196, 1 out. 2006.
- GAO, Y. et al. Discovery and anti-inflammatory evaluation of benzothiazepinones (BTZs) as novel non-ATP competitive inhibitors of glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ). **Bioorganic & Medicinal Chemistry**, v. 26, n. 20, p. 5479–5493, nov. 2018.
- GARCÍA-MORALES, V. et al. Current Understanding of the Physiopathology, Diagnosis and Therapeutic Approach to Alzheimer's Disease. **Biomedicines**, v. 9, n. 12, p. 1910, 14 dez. 2021.
- GENTILE, F. et al. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. **ACS Central Science**, v. 6, n. 6, p. 939–949, 2020.
- GERTZEN, C. G. W.; GOHLKE, H. From Hansch-Fujita Analysis to AFMoC: A Road to Structure-Based QSAR. **Molecular Informatics**, v. 31, n. 10, p. 698–704, 7 out. 2012.
- GHOSE, A. K. et al. Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery. **ACS Chemical Neuroscience**, v. 3, n. 1, p. 50–68, 2012.

- GIACOBINI, E. Cholinergic foundations of Alzheimer's disease therapy. **Journal of Physiology-Paris**, v. 92, n. 3–4, p. 283–287, jun. 1998.
- GLORIAM, D. E. Bigger is better in virtual drug screens. **Nature**, v. 566, n. 7743, p. 193–194, 6 fev. 2019.
- GOLDBERG, F. W. et al. Designing novel building blocks is an overlooked strategy to improve compound quality. **Drug Discovery Today**, v. 20, n. 1, p. 11–17, jan. 2015.
- 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, jul. 1985.
- GORELIK, B.; GOLDBLUM, A. High quality binding modes in docking ligands to proteins. **Proteins: Structure, Function, and Bioinformatics**, v. 71, n. 3, p. 1373–1386, 15 maio 2008.
- GORGULLA, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. **Nature**, v. 580, n. 7805, p. 663–668, 30 abr. 2020.
- GORGULLA, C. et al. Emerging frontiers in virtual drug discovery: From quantum mechanical methods to deep learning approaches. **Current Opinion in Chemical Biology**, v. 69, p. 102156, 1 ago. 2022.
- GOULART, P. et al. Butyrylcholinesterase - BuChE: A Potential Target for Development of Drugs for Alzheimer's Disease Treatment. **Revista Virtual de Química**, v. 13, n. 1, p. 90–126, 2021.
- GRAMATICA, P. Principles of QSAR models validation: Internal and external. **QSAR and Combinatorial Science**, v. 26, n. 5, p. 694–701, maio 2007.
- GRAMATICA, P. Principles of QSAR Modeling. **International Journal of Quantitative Structure-Property Relationships**, v. 5, n. 3, p. 1–37, jul. 2020.
- GRANT, J. A.; GALLARDO, M. A.; PICKUP, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. **Journal of Computational Chemistry**, v. 17, n. 14, p. 1653–1666, 15 nov. 1996.
- GREENAMYRE, J. T. et al. Glutamate transmission and toxicity in alzheimer's disease. **Progress in Neuropsychopharmacology and Biological Psychiatry**, 1988.
- 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, 24 set. 2018.
- GUO, F. et al. Applications of Carbon Dots for the Treatment of Alzheimer's Disease. **International Journal of Nanomedicine**, v. Volume 17, p. 6621–6638, dez. 2022.
- HALGREN, T. A. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1750–1759, 1 mar. 2004.
- HALGREN, T. A. Identifying and characterizing binding sites and assessing druggability. **Journal of Chemical Information and Modeling**, 2009.
- HAMADACHE, M. et al. A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction. **Journal of Hazardous Materials**, 2016.
- HAMMETT, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. **Journal of the American Chemical Society**, v. 59, n. 1, p. 96–103, 1937.
- HANGER, D. P. et al. Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: Generation of paired helical filament epitopes and neuronal localisation of the kinase. **Neuroscience Letters**, v. 147, n. 1, p. 58–62, nov. 1992.
- HANKS, S. K.; QUINN, A. M.; HUNTER, T. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. **Science**, 1988.
- HANSCH, C. et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. **Nature**, v. 194, p. 178–180, 1962.
- HANSCH, C. The advent and evolution of QSAR at Pomona College. **Journal of Computer-Aided Molecular Design**, v. 25, n. 6, p. 495–507, 2011.
- HANSCH, C.; LEO, A. **Exploring QSAR: Fundamentals and Applications in Chemistry and Biology**. 1. ed. [s.l.] American Chemical Society - ACS, 1995.
- 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. 572–584, 26 abr. 2010.
- HAWKINS, P. C. D.; NICHOLLS, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. **Journal of Chemical Information and Modeling**, v. 52, p. 2919–2936, 2012.

- HAWKINS, P. C. D. P. C. D.; SKILLMAN, A. G. G.; NICHOLLS, A. Comparison of shape-matching and docking as virtual screening tools. **Journal of Medicinal Chemistry**, v. 50, n. 1, p. 74–82, 2007.
- HAWKINS, P. C. D.; STAHL, G. Ligand-Based Methods in GPCR Computer-Aided Drug Design. In: HEIFETZ, A. (Ed.). **Methods in Molecular Biology**. New York, NY: Humana Press, 2018. p. 365–374.
- HERS, I.; TAVARÉ, J. M.; DENTON, R. M. The protein kinase C inhibitors bisindolylmaleimide I (GF 109203x) and IX (Ro 31-8220) are potent inhibitors of glycogen synthase kinase-3 activity. **FEBS Letters**, v. 460, n. 3, p. 433–436, 5 nov. 1999.
- HIMMAT, M. et al. Adapting Document Similarity Measures for Ligand-Based Virtual Screening. **Molecules**, v. 21, n. 4, p. 476, 13 abr. 2016.
- HÖNIG, S. M. N.; LEMMEN, C.; RAREY, M. Small molecule superposition: A comprehensive overview on pose scoring of the latest methods. **WIREs Computational Molecular Science**, p. e1640, 11 out. 2022.
- HU, J. et al. LS-align: an atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. **Bioinformatics**, v. 34, n. 13, p. 2209–2218, 1 jul. 2018.
- HUANG, J. et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. **Nature Methods**, 2016.
- HUANG, S.-Y.; ZOU, X. Mean-Force Scoring Functions for Protein–Ligand Binding. In: WHEELER, R. A. (Ed.). **Annual Reports in Computational Chemistry**. New York: Elsevier, 2010a. 6p. 280–296.
- HUANG, S.-Y.; ZOU, X. Advances and Challenges in Protein-Ligand Docking. **International Journal of Molecular Sciences**, v. 11, n. 8, p. 3016–3034, 18 ago. 2010b.
- IBRAHIM, M.; GABR, M. Multitarget therapeutic strategies for Alzheimer’s disease. **Neural Regeneration Research**, v. 14, n. 3, p. 437, 2019.
- IBS. **InterBioScreen - Databases**. Disponível em: <<https://www.ibscreen.com/bases>>.
- ILOUZ, R. et al. Inhibition of glycogen synthase kinase-3 $\beta$  by bivalent zinc ions: insight into the insulin-mimetic action of zinc. **Biochemical and Biophysical Research Communications**, v. 295, n. 1, p. 102–106, jul. 2002.
- IRWIN, J. J.; SHOICHET, B. K. ZINC--a free database of commercially available compounds for virtual screening. **Journal of chemical information and modeling**, v. 45, n. 1, p. 177–82, 2005.
- JAGUST, W. Imaging the evolution and pathophysiology of Alzheimer disease. **Nature Reviews Neuroscience**, v. 19, n. 11, p. 687–700, 28 nov. 2018.
- JANKOWSKA, A. et al. Multi-Target-Directed Ligands Affecting Serotonergic Neurotransmission for Alzheimer’s Disease Therapy: Advances in Chemical and Biological Research. **Current Medicinal Chemistry**, v. 25, n. 17, p. 2045–2067, 22 maio 2018.
- JOHNSON, L. N.; NOBLE, M. E. M.; OWEN, D. J. Active and Inactive Protein Kinases: Structural Basis for Regulation. **Cell**, v. 85, n. 2, p. 149–158, abr. 1996.
- 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, 4 abr. 1997.
- 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, 1 jan. 1995.
- JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, v. 79, n. 2, p. 926–935, 1983.
- JOSHI, P. et al. (Z)-2-(3-Chlorobenzylidene)-3,4-dihydro-N-(2-methoxyethyl)-3-oxo-2H-benzo[b][1,4]oxazine-6-carboxamide as GSK-3 $\beta$  inhibitor: Identification by virtual screening and its validation in enzyme- and cell-based assay. **Chemical Biology and Drug Design**, v. 89, n. 6, p. 964–971, 2017.
- KANNT, A.; WIELAND, T. Managing risks in drug discovery: Reproducibility of published findings. **Naunyn-Schmiedeberg’s Archives of Pharmacology**, v. 389, n. 4, p. 353–360, 2016.
- KAPLAN, A. L. et al. Bespoke library docking for 5-HT<sub>2A</sub> receptor agonists with antidepressant activity. **Nature**, v. 610, n. 7932, p. 582–591, 20 out. 2022.
- KAUSAR, S.; FALCAO, A. O. An automated framework for QSAR model building. **Journal of Cheminformatics**, v. 10, n. 1, p. 1–23, 2018.
- KAWABATA, T.; NAKAMURA, H. 3D Flexible Alignment Using 2D Maximum Common Substructure: Dependence of Prediction Accuracy on Target-Reference Chemical Similarity. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1850–1863, 28 jul. 2014.
- 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, 24 ago. 2015.
- KELLICI, T. et al. Rational Drug Design Paradigms: The Odyssey for Designing Better Drugs. **Combinatorial Chemistry & High**

**Throughput Screening**, v. 18, n. 3, p. 238–256, 2015.

KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: Methods and applications. **Nature Reviews Drug Discovery**, v. 3, n. 11, p. 935–949, 2004.

KIVIPELTO, M.; MANGIALASCHE, F.; NGANDU, T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. **Nature Reviews Neurology**, v. 14, n. 11, p. 653–666, 5 nov. 2018.

KLEIN, P. S.; MELTON, D. A. A molecular mechanism for the effect of lithium on development. **Proceedings of the National Academy of Sciences**, v. 93, n. 16, p. 8455–8459, 6 ago. 1996.

KLOPMAND, G. Concepts and applications of molecular similarity, by Mark A. Johnson and Gerald M. Maggiora, eds., John Wiley & Sons, New York, 1990, 393 pp. Price: \$65.00. **Journal of Computational Chemistry**, v. 13, n. 4, p. 539–540, maio 1992.

KNIME AG. **KNIME Analytics Platform 4.3.1** Zurich, Switzerland, 2021. .

KORB, O.; STÜTZLE, T.; EXNER, T. E. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. **Journal of Chemical Information and Modeling**, 2009.

KORTAGERE, S.; KRASOWSKI, M. D.; EKINS, S. The importance of discerning shape in molecular pharmacology. **Trends in Pharmacological Sciences**, v. 30, n. 3, p. 138–147, 1 mar. 2009.

KOSHLAND, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. **Proceedings of the National Academy of Sciences**, v. 44, n. 2, p. 98–104, fev. 1958.

KRISHNAN, N. et al. Harnessing insulin-and leptin-induced oxidation of PTP1B for therapeutic development. **Nature Communications**, v. 9, n. 1, 2018.

KUBINYI, H. **QSAR: Hansch Analysis and Related Approaches**. 1. ed. Weinheim: Wiley-VCH, 1993.

KUBINYI, H. The Changing Landscape in Drug Discovery. **Discovery**, p. 24–46, 2007.

KUMAR, A.; ZHANG, K. Y. J. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. **Frontiers in Chemistry**, v. 6, n. July, p. 1–21, 25 jul. 2018.

KUZ'MIN, V. E.; ARTEMENKO, A. G.; MURATOV, E. N. Hierarchical QSAR technology based on the Simplex representation of molecular structure. **Journal of Computer-Aided Molecular Design**, v. 22, n. 6–7, p. 403–421, 6 jun. 2008.

LABUTE, P. et al. Flexible Alignment of Small Molecules. **Journal of Medicinal Chemistry**, v. 44, n. 10, p. 1483–1490, 1 maio 2001.

LAGORCE, D. et al. In Silico ADME/Tox Predictions. **ADMET for Medicinal Chemists: A Practical Guide**, p. 29–124, 2010.

LAMBRINIDIS, G.; TSANTILI-KAKOULIDOU, A. Challenges with multi-objective QSAR in drug discovery. **Expert Opinion on Drug Discovery**, v. 13, n. 9, p. 851–859, 2018.

LE GUILLOUX, V.; SCHMIDTKE, P.; TUFFERY, P. Fpocket: An open source platform for ligand pocket detection. **BMC Bioinformatics**, 2009.

LECLERC, S. et al. Indirubins Inhibit Glycogen Synthase Kinase-3 $\beta$  and CDK5/P25, Two Protein Kinases Involved in Abnormal Tau Phosphorylation in Alzheimer's Disease. **Journal of Biological Chemistry**, v. 276, n. 1, p. 251–260, 5 jan. 2001.

LEI, T. et al. ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. **Journal of Cheminformatics**, 2016.

LEMMEN, C.; LENGAUER, T.; KLEBE, G. FlexS: a method for fast flexible ligand superposition. **Journal of Medicinal Chemistry**, v. 41, n. 23, p. 4502–4520, 1 nov. 1998.

LENGAUER, T.; RAREY, M. Computational methods for biomolecular docking. **Current Opinion in Structural Biology**, v. 6, n. 3, p. 402–406, 1 jun. 1996.

LI, Q.; SHAH, S. Structure-Based Virtual Screening. In: WU, C. H.; ARIGHI, C. N.; ROSS, K. E. (Ed.). **Protein Bioinformatics: From Protein Modifications and Networks to Proteomics**. Methods in Molecular Biology. New York, NY: Springer New York, 2017. 1558p. 111–124.

LI, Y. et al. A Novel Method for Drug Screen to Regulate G Protein-Coupled Receptors in the Metabolic Network of Alzheimer's Disease. **BioMed Research International**, v. 2018, p. 1–10, 2018.

LIEBESCHUETZ, J. W.; COLE, J. C.; KORB, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. **Journal of Computer-Aided Molecular Design**, v. 26, n. 6, p. 737–748, 28 jun. 2012.

LIONTA, E. et al. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. **Current Topics in Medicinal Chemistry**, v. 14, n. 16, p. 1923–1938, 15 out. 2014.

LIPINSKI, C. A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 23, n. 1–3, p. 3–25, jan. 1997.

- LIPINSKI, C. A. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. **Advanced Drug Delivery Reviews**, v. 101, p. 34–41, jun. 2016.
- LLANOS, M. A. et al. Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study. **Journal of Chemical Information and Modeling**, v. 61, n. 8, p. 3758–3770, 23 ago. 2021.
- LORBER, D. M.; SHOICHET, B. K. Flexible ligand docking using conformational ensembles. **Protein Science**, v. 7, n. 4, p. 938–950, abr. 1998.
- LOURENCO, M. V. et al. Exercise-linked FNDC5/irisin rescues synaptic plasticity and memory defects in Alzheimer's models. **Nature Medicine**, v. 25, n. 1, p. 165–175, 7 jan. 2018.
- LOWIS, D. R. HQSAR: a new, highly predictive QSAR technique. **Tripos technical notes**, v. 1, n. 5, p. 1–17, 1997.
- MADHAVI SASTRY, G. 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.
- MAGGIORA, G. et al. Molecular Similarity in Medicinal Chemistry. **Journal of Medicinal Chemistry**, v. 57, n. 8, p. 3186–3204, 24 abr. 2014.
- MARTIN, T. M.; LILAVOIS, C. R.; BARRON, M. G. Prediction of pesticide acute toxicity using two-dimensional chemical descriptors and target species classification. **SAR and QSAR in Environmental Research**, 2017.
- MARTIN, Y. C. et al. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. **Journal of Computer-Aided Molecular Design**, v. 7, n. 1, p. 83–102, fev. 1993.
- MARTINEZ, A. et al. First non-ATP competitive glycogen synthase kinase 3  $\beta$  (GSK-3 $\beta$ ) inhibitors: Thiadiazolidinones (TDZD) as potential drugs for the treatment of Alzheimer's disease. **Journal of Medicinal Chemistry**, v. 45, n. 6, p. 1292–1299, 2002.
- MARTINEZ, A.; PEREZ, D. I.; GIL, C. Lessons Learnt from Glycogen Synthase Kinase 3 Inhibitors Development for Alzheimer's Disease. **Current Topics in Medicinal Chemistry**, v. 13, n. August, p. 1808–1819, 2013.
- MAYBRIDGE. **The Maybridge Screening Collection**. Disponível em: <<https://www.alfa.com/pt/maybridge-pre-plated-screening-compounds-and-fragment-libraries/>>. Acesso em: 15 fev. 2023.
- MCCUBREY, J. A. et al. Roles of GSK-3 and microRNAs on epithelial mesenchymal transition and cancer stem cells. **Oncotarget**, v. 8, n. 8, p. 14221–14250, 21 fev. 2017.
- MCGANN, M. FRED pose prediction and virtual screening accuracy. **Journal of Chemical Information and Modeling**, 2011.
- MCGANN, M. FRED and HYBRID docking performance on standardized datasets. **Journal of Computer-Aided Molecular Design**, v. 26, n. 8, p. 897–906, 5 ago. 2012.
- MCINNES, C.; FISCHER, P. Strategies for the Design of Potent and Selective Kinase Inhibitors. **Current Pharmaceutical Design**, v. 11, n. 14, p. 1845–1863, 1 maio 2005.
- MEDINA, M. An Overview on the Clinical Development of Tau-Based Therapeutics. **International Journal of Molecular Sciences**, v. 19, n. 4, p. 1160, 11 abr. 2018.
- MEIJER, L. et al. Inhibition of cyclin-dependent kinases, GSK-3 $\beta$  and CK1 by hymenialdisine, a marine sponge constituent. **Chemistry and Biology**, v. 7, n. 1, p. 51–63, 2000.
- MELO-FILHO, C. C. et al. QSAR-Driven Discovery of Novel Chemical Scaffolds Active against Schistosoma mansoni. **Journal of Chemical Information and Modeling**, v. 56, n. 7, p. 1357–1372, 25 jul. 2016.
- MENG, E. C.; SHOICHET, B. K.; KUNTZ, I. D. Automated docking with grid-based energy evaluation. **Journal of Computational Chemistry**, v. 13, n. 4, p. 505–524, 1 maio 1992.
- MIGNANI, S. et al. Compound high-quality criteria: a new vision to guide the development of drugs, current situation. **Drug Discovery Today**, v. 21, n. 4, p. 573–584, abr. 2016.
- MIRABELLO, C.; WALLNER, B. InterLig: improved ligand-based virtual screening using topologically independent structural alignments. **Bioinformatics**, v. 36, n. 10, p. 3266–3267, 1 maio 2020.
- MOLPORT. **Chemical Libraries & Screening Compound Database**. Disponível em: <<https://www.molport.com/shop/screening-compound-database>>.
- MORGAN, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. **Journal of Chemical Documentation**, v. 5, n. 2, p. 107–113, 1 maio 1965.
- MORRIS, G. M. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639–1662, 15 nov. 1998.
- MORRIS, G. M. et al. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. **Journal of Computational**



**Chemistry**, 2009.

MUEGGE, I. PMF Scoring Revisited. **Journal of Medicinal Chemistry**, v. 49, n. 20, p. 5895–5902, out. 2006.

MÜLLER, S.; KNAPP, S. CHAPTER 3. Targeting Catalytic and Non-Catalytic Functions of Protein Kinases. In: **Allosterism in Drug Discovery**. Cambridge: Royal Society of Chemistry, 2017. p. 40–64.

MURATOV, E. N. et al. QSAR without borders. **Chemical Society Reviews**, v. 49, n. 11, p. 3525–3564, 8 jun. 2020.

MYSINGER, M. M. et al. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. **Journal of Medicinal Chemistry**, v. 55, n. 14, p. 6582–6594, 26 jul. 2012.

MYSINGER, M. M.; SHOICHET, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. **Journal of Chemical Information and Modeling**, v. 50, n. 9, p. 1561–1573, 27 set. 2010.

NASSER, M. et al. Feature Reduction for Molecular Similarity Searching Based on Autoencoder Deep Learning. **Biomolecules**, v. 12, n. 4, p. 508, 27 mar. 2022.

NAYARISSERI, A. et al. Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery. **Current Drug Targets**, v. 22, n. 6, p. 631–655, 5 abr. 2021.

NEVES, B. J. et al. Best Practices for Docking-Based Virtual Screening. In: **Molecular Docking for Computer-Aided Drug Design**. [s.l.] Elsevier, 2021. p. 75–98.

NICHOLLS, A. What do we know and when do we know it? **Journal of Computer-Aided Molecular Design**, v. 22, n. 3–4, p. 239–255, 2008.

NICHOLLS, A. et al. Molecular Shape and Medicinal Chemistry: A Perspective. **Journal of Medicinal Chemistry**, v. 53, n. 10, p. 3862–3886, 27 maio 2010.

NOGUEIRA, L. J.; MONTANARI, C. A.; DONNICI, C. L. The history, evolution and importance of lipophilicity in medicinal chemistry: from Hippocrates and Galeno to Paracelsus and the contributions of Overton and Hansch. **Revista Virtual de Química**, v. 1, n. 3, 2009.

OECD. **Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models**. Paris: OECD, 2007. v. 69

OPENEYE. **Scientific Software. FRED 3.5.0.4**Santa Fe, NM, 2020a. .

OPENEYE. **Scientific Software. Documentation - Applications 2020.0.4.** Disponível em: <<https://docs.eyesopen.com/applications/index.html>>. Acesso em: 7 fev. 2020b.

OPENEYE. **Scientific Software. OMEGA 4.0.0.4**Santa Fe, NM, 2020c. .

OPENEYE. **Scientific Software. ROCS 3.4.0.4**Santa Fe, NM, 2020d. .

OPENEYE. **Scientific Software. EON 2.3.3.4**Santa Fe, NM, 2020e. .

OPENEYE. **OEDOCKING 4.2.0.2: OpenEye Scientific Software, Inc.**, 2023a. .

OPENEYE. **Documentation: EON Theory - OpenEye Scientific Software, Inc.** Disponível em: <[https://docs.eyesopen.com/applications/eon/theory/eon\\_theory.html](https://docs.eyesopen.com/applications/eon/theory/eon_theory.html)>. Acesso em: 9 fev. 2023b.

OVER, B. et al. Natural-product-derived fragments for fragment-based ligand discovery. **Nature Chemistry**, v. 5, n. 1, p. 21–28, 2 jan. 2013.

PALOMO, V. et al. Exploring the binding sites of glycogen synthase kinase 3. identification and characterization of allosteric modulation cavities. **Journal of Medicinal Chemistry**, v. 54, n. 24, p. 8461–8470, 2011.

PALOMO, V. et al. Subtly Modulating Glycogen Synthase Kinase 3  $\beta$ : Allosteric Inhibitor Development and Their Potential for the Treatment of Chronic Diseases. **Journal of Medicinal Chemistry**, v. 60, n. 12, p. 4983–5001, 2017.

PALOMO, V.; MARTINEZ, A. Glycogen synthase kinase 3 (GSK-3) inhibitors: a patent update (2014-2015). **Expert Opinion on Therapeutic Patents**, v. 27, n. 6, p. 657–666, 2017.

PANJKOVICH, A.; DAURA, X. PARS: A web server for the prediction of Protein Allosteric and Regulatory Sites. **Bioinformatics**, 2014.

PATTERSON, C. The state of the art of dementia research: New frontiers. **World Alzheimer Report**, p. 48, 2018.

PAWAR, G. et al. In Silico Toxicology Data Resources to Support Read-Across and (Q)SAR. **Frontiers in Pharmacology**, v. 10, n. JUN, p. 1–26, 11 jun. 2019.

PEDREGOSA FABIANPEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

- PEITSCH, M. C. Manuel Peitsch discusses knowledge management and informatics in drug discovery. **Drug Discovery Today: BIOSILICO**, v. 2, n. 3, p. 94–96, maio 2004.
- PERKINELMER INFORMATICS INC. **Chemdraw Professional 18.2.0.48**, 2021. .
- PICCIRILLO, E.; DO AMARAL, A. T. Virtual screening of bioactive compounds: Concepts and applications. **Química Nova**, v. 41, n. 6, p. 662–677, 2018.
- PINHIERO, A. et al. In silico Identification of Novel Potential BACE-1 Inhibitors for Alzheimer's Disease Treatment: Molecular Docking, Pharmacophore Modeling and Activity and Synthetic Accessibility Predictions. **British Journal of Pharmaceutical Research**, v. 7, n. 3, p. 217–229, 2015.
- POIANI, J. G. C. **Planejamento racional de candidatos a fármacos inibidores de glicogênio sintase cinase - 3 beta (GSK-3B) em doença de Alzheimer**. 2017. Universidade de São Paulo, Ribeirão Preto, 2017.
- PRADO-PRADO, F. J. et al. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug–drug complex networks. **Bioorganic & Medicinal Chemistry**, v. 17, n. 2, p. 569–575, jan. 2009.
- PRINCETON. **Chemistry Princeton University Library**. Disponível em: <<https://library.princeton.edu/databases/subject/chemistry>>.
- RANKOVIC, Z. CNS Drug Design: Balancing Physicochemical Properties for Optimal Brain Exposure. **Journal of Medicinal Chemistry**, v. 58, n. 6, p. 2584–2608, 26 mar. 2015.
- RANKOVIC, Z. CNS Physicochemical Property Space Shaped by a Diverse Set of Molecules with Experimentally Determined Exposure in the Mouse Brain. **Journal of Medicinal Chemistry**, v. 60, n. 14, p. 5943–5954, 2017.
- RAREY, M. et al. A Fast Flexible Docking Method using an Incremental Construction Algorithm. **Journal of Molecular Biology**, v. 261, n. 3, p. 470–489, ago. 1996.
- RDKIT: OPEN-SOURCE CHEMINFORMATICS. <http://www.rdkit.org>, 2021. .
- REICHMAN, W. E. Current pharmacologic options for patients with Alzheimer's disease. **Annals of general hospital psychiatry**, v. 2, n. 1, p. 1, 29 jan. 2003.
- REN, H. C.; SAI, Y.; CHEN, T. Evaluation of Generic Methods to Predict Human Pharmacokinetics Using Physiologically Based Pharmacokinetic Model for Early Drug Discovery of Tyrosine Kinase Inhibitors. **European Journal of Drug Metabolism and Pharmacokinetics**, v. 44, n. 1, p. 121–132, 2019.
- RICHMOND, N. J. et al. GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D. **Journal of Computer-Aided Molecular Design**, v. 20, n. 9, p. 567–587, 19 set. 2006.
- RINIKER, S.; LANDRUM, G. A. Open-source platform to benchmark fingerprints for ligand-based virtual screening. **Journal of Cheminformatics**, v. 5, n. 1, p. 26, 30 dez. 2013.
- RODRIGUES, R. P.; SILVA, C. H. T. P. da. Discovery of potential neurodegenerative inhibitors in Alzheimer's disease by casein kinase 1 structure-based virtual screening. **Medicinal Chemistry Research**, v. 26, n. 12, p. 3274–3285, 22 dez. 2017.
- ROGERS, D.; HAHN, M. Extended-Connectivity Fingerprints. **Journal of Chemical Information and Modeling**, v. 50, n. 5, p. 742–754, 24 maio 2010.
- ROGERS, D. J.; TANIMOTO, T. T. A Computer Program for Classifying Plants. **Science**, v. 132, n. 3434, p. 1115–1118, 21 out. 1960.
- ROY, K.; KAR, S.; DAS, R. N. Statistical Methods in QSAR/QSPR. In: **A Primer on QSAR/QSPR Modeling**. Cham: Springer, 2015. p. 37–59.
- RUSYN, I. et al. Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. **Toxicological sciences : an official journal of the Society of Toxicology**, v. 127, n. 1, p. 1–9, maio 2012.
- RUVINSKY, A. M. Role of binding entropy in the refinement of protein-ligand docking predictions: Analysis based on the use of 11 scoring functions. **Journal of Computational Chemistry**, v. 28, n. 8, p. 1364–1372, jun. 2007.
- SAZONOVAS, A.; JAPERTAS, P.; DIDZIAPETRIS, R. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). **SAR and QSAR in Environmental Research**, 2010.
- SCHNEIDMAN-DUHOVNY, D. et al. PharmaGist: a webserver for ligand-based pharmacophore detection. **Nucleic Acids Research**, v. 36, n. Web Server, p. W223–W228, 19 maio 2008.
- SCHRÖDINGER. **LigPrep - Suite 2015-2**, 2018a. .
- SCHRÖDINGER. **Protein Preparation Wizard - Suite 2015-2**, 2018b. .
- SCHRÖDINGER. **Glide Schrödinger Suite 2019-2**, 2019. .
- SCHRÖDINGER. **Schrödinger Release 2021-1: MaestroLLC**, New York, NY, 2021. .

SCHRÖDINGER LLC. **The PyMol Molecular Graphics System 1.3**, 2010. .

SCHWAB, C. H. Conformations and 3D pharmacophore searching. **Drug Discovery Today: Technologies**, v. 7, n. 4, p. e245–e253, dez. 2010.

SCIOR, T. et al. Recognizing pitfalls in virtual screening: A critical review. **Journal of Chemical Information and Modeling**, v. 52, n. 4, p. 867–881, 2012.

SHINTANI, E. Y.; UCHIDA, K. M. Donepezil: an anticholinesterase inhibitor for Alzheimer's disease. **American Journal of Health-System Pharmacy**, v. 54, n. 24, p. 2805–2810, 15 dez. 1997.

SILVA, G. M. et al. Allosteric Modulators of Potential Targets Related to Alzheimer's Disease: a Review. **ChemMedChem**, v. 14, n. 16, p. 1467–1483, 20 ago. 2019.

SILVA, G. M. et al. In Silico Methods to Predict Relevant Toxicological Endpoints of Bioactive Substances. In: [s.l.] Springer, Cham, 2021a. p. 649–676.

SILVA, G. M. et al. Revisiting the Proposition of Binding Pockets and Bioactive Poses for GSK-3 $\beta$  Allosteric Modulators Addressed to Neurodegenerative Diseases. **International Journal of Molecular Sciences** 2021, Vol. 22, Page 8252, v. 22, n. 15, p. 8252, 31 jul. 2021b.

SIVAPRAKASAM, P. et al. Discovery of new acylaminopyridines as GSK-3 inhibitors by a structure guided in-depth exploration of chemical space around a pyrrolopyridinone core. **Bioorganic & Medicinal Chemistry Letters**, v. 25, n. 9, p. 1856–1863, maio 2015.

SLIWOSKI, G.; MENDENHALL, J.; MEILER, J. Autocorrelation descriptor improvements for QSAR: 2DA-Sign and 3DA-Sign. **Journal of Computer-Aided Molecular Design**, v. 30, n. 3, p. 209–217, 2016.

SOUSA, S. F.; FERNANDES, P. A.; RAMOS, M. J. Protein-ligand docking: Current status and future challenges. **Proteins: Structure, Function, and Bioinformatics**, v. 65, n. 1, p. 15–26, 21 jul. 2006.

SPITZER, R.; JAIN, A. N. Surflex-Dock: Docking benchmarks and real-world application. **Journal of Computer-Aided Molecular Design**, v. 26, n. 6, p. 687–699, 9 jun. 2012.

STANDAERT, D. G.; YOUNG, A. B. **Tratamento dos distúrbios degenerativos do sistema nervoso central**. 10. ed. New York: The McGraw-Hill Companies, 2005.

STEPHENSON, J. et al. Inflammation in CNS neurodegenerative diseases. **Immunology**, v. 154, n. 2, p. 204–219, 2018.

STERLING, T.; IRWIN, J. J. ZINC 15 - Ligand Discovery for Everyone. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2324–2337, 2015.

SU, M. et al. Comparative Assessment of Scoring Functions: The CASF-2016 Update. **Journal of Chemical Information and Modeling**, v. 59, n. 2, p. 895–913, 25 fev. 2019.

SUPERSTAR. **Cambridge Crystallographic Data Centre - SuperStar User Guide and Tutorials - Release update 3** Cambridge Crystallographic Data Centre, , 2018. .

SUTHERLAND, C. What Are the bona fide GSK3 Substrates? **International Journal of Alzheimer's Disease**, v. 2011, p. 1–23, 2011.

SVETNIK, V. et al. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. **Journal of Chemical Information and Computer Sciences**, v. 43, n. 6, p. 1947–1958, 1 nov. 2003.

TEAGUE, S. J. et al. The Design of Leadlike Combinatorial Libraries. **Angewandte Chemie International Edition**, v. 38, n. 24, p. 3743–3748, 16 dez. 1999.

TETKO, I. V.; TANCHUK, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 5, p. 1136–1145, set. 2002.

THE SCRIPPS RESEARCH INSTITUTE. **Autodock 4** (<http://autodock.scripps.edu/>), 2020. .

TIAN, S. et al. The application of in silico drug-likeness predictions in pharmaceutical research. **Advanced Drug Delivery Reviews**, v. 86, p. 2–10, jun. 2015.

TODESCHINI, R.; CONSONNI, V. **Handbook of Molecular Descriptors**. Weinheim: Wiley, 2000.

TOROPOV, A. A.; TOROPOVA, A. P. QSPR/QSAR: State-of-Art, Weirdness, the Future. **Molecules**, v. 25, n. 6, p. 1292, 12 mar. 2020.

TOTROV, M. Atomic Property Fields: Generalized 3D Pharmacophoric Potential for Automated Ligand Superposition, Pharmacophore Elucidation and 3D QSAR. **Chemical Biology & Drug Design**, v. 71, n. 1, p. 15–27, 7 dez. 2007.

TRESADERN, G.; BEMPORAD, D.; HOWE, T. A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor. **Journal of Molecular Graphics and Modelling**, v. 27, p. 860–870, 2009.

TRIBALLEAU, N. et al. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. **Journal of Medicinal Chemistry**, v. 48, n.

7, p. 2534–2547, abr. 2005.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, 12 jul. 2010.

TROPSHA, A.; GOLBRAIKH, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. **Current pharmaceutical design**, v. 13, n. 34, p. 3494–504, jan. 2007.

TRUCHON, J.-F.; BAYLY, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. **Journal of Chemical Information and Modeling**, v. 47, n. 2, p. 488–508, 1 mar. 2007.

VAN DER SCHYF, C. J. The use of multi-target drugs in the treatment of neurodegenerative diseases. **Expert Review of Clinical Pharmacology**, v. 4, n. 3, p. 293–298, 2011.

VANOMMESLAEGHE, K.; MACKERELL, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. n. 1, 2012.

VARESI, A. et al. Non-Enzymatic Antioxidants against Alzheimer’s Disease: Prevention, Diagnosis and Therapy. **Antioxidants**, v. 12, n. 1, p. 180, 12 jan. 2023.

VÁZQUEZ, J. et al. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. **Molecules**, v. 25, n. 20, p. 4723, 15 out. 2020.

VEDANI, A. et al. Multiple-conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. **Journal of Medicinal Chemistry**, 2000.

VEDANI, A.; DOBLER, M. Multidimensional QSAR: Moving from three- to five-dimensional concepts. **Quantitative Structure-Activity Relationships**, v. 21, n. 4, p. 382–390, out. 2002.

VEMULA, D. et al. CADD, AI and ML in drug discovery: A comprehensive review. **European Journal of Pharmaceutical Sciences**, v. 181, p. 106324, 1 fev. 2023.

VERDONK, M. L. et al. Improved protein-ligand docking using GOLD. **Proteins: Structure, Function, and Bioinformatics**, v. 52, n. 4, p. 609–623, 1 ago. 2003.

VERÍSSIMO, G. C. et al. Designing drugs when there is low data availability: one-shot learning and other approaches to face the issues of a long-term concern. **Expert Opinion on Drug Discovery**, v. 17, n. 9, p. 929–947, 2 set. 2022.

VOLKAMER, A. et al. Analyzing the topology of active sites: On the prediction of pockets and subpockets. **Journal of Chemical Information and Modeling**, v. 50, n. 11, p. 2041–2052, 2010.

VOLKAMER, A. et al. Combining global and local measures for structure-based druggability predictions. **Journal of Chemical Information and Modeling**, v. 52, n. 2, p. 360–372, 2012.

VUKOVIC, K.; GADALETA, D.; BENFENATI, E. Methodology of AiQSAr: A group-specific approach to QSAR modelling. **Journal of Cheminformatics**, 2019.

WAGER, T. T. et al. Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. **ACS Chemical Neuroscience**, v. 1, n. 6, p. 435–449, 2010.

WAGER, T. T. et al. Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. **ACS Chemical Neuroscience**, v. 7, n. 6, p. 767–775, 2016.

WANG, W. et al. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. **Annual Review of Biophysics and Biomolecular Structure**, v. 30, n. 1, p. 211–243, jun. 2001.

WANG, W.; OUYANG, D. Opportunities and challenges of physiologically based pharmacokinetic modeling in drug delivery. **Drug Discovery Today**, v. 27, n. 8, p. 2100–2120, ago. 2022.

WARR, W. A. et al. Exploration of Ultralarge Compound Collections for Drug Discovery. **Journal of Chemical Information and Modeling**, v. 62, p. 29, 2022.

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, 1 out. 2006.

WESTBROOK, J. D. et al. The chemical component dictionary: Complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. **Bioinformatics**, 2015.

WILLETT, P. Similarity-based virtual screening using 2D fingerprints. **Drug Discovery Today**, v. 11, n. 23–24, p. 1046–1053, dez. 2006.

WILLETT, P. Chemoinformatics: A history. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 1, p. 46–56, 2011.

- WOLBER, G.; LANGER, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 160–169, 1 jan. 2005.
- WOODGETT, J. R. Molecular cloning and expression of glycogen synthase kinase-3/factor A. **The EMBO journal**, v. 9, n. 8, p. 2431–8, 1990.
- WORLD HEALTH ORGANIZATION. Global status report on the public health response to dementia. **Geneva: World Health Organization**, p. 251, 2021.
- WORTH, ANDREW; FUART-GATNIK, MOJCA; LAPENNA, SILVIA; SERAFIMOVA, R. SCIENTIFIC REPORT submitted to EFSA Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk. n. June, p. 174, 2011.
- YAP, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. **Journal of Computational Chemistry**, 2011.
- YUSOF, I.; SEGALL, M. D. Considering the impact drug-like properties have on the chance of success. **Drug Discovery Today**, v. 18, n. 13–14, p. 659–666, jul. 2013.
- ZAGÓRSKA, A. et al. Drug Discovery and Development Targeting Dementia. **Pharmaceuticals**, v. 16, n. 2, p. 151, 19 jan. 2023.
- ZENG, X. et al. NPASS: Natural product activity and species source database for natural product research, discovery and tool development. **Nucleic Acids Research**, 2018.
- ZHANG, D. et al. Biological evaluation and energetic analyses of novel GSK-3 $\beta$  inhibitors. **Journal of Cellular Biochemistry**, v. 119, n. 4, p. 3510–3518, 2018.
- ZHANG, P. et al. Design, synthesis and biological evaluation of benzothiazepinones (BTZs) as novel non-ATP competitive inhibitors of glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ). **European Journal of Medicinal Chemistry**, v. 61, p. 95–103, mar. 2013.
- ZHANG, P. et al. Novel benzothiazinones (BTOs) as allosteric modulator or substrate competitive inhibitor of glycogen synthase kinase 3 $\beta$  (GSK-3 $\beta$ ) with cellular activity of promoting glucose uptake. **Bioorganic and Medicinal Chemistry Letters**, v. 24, n. 24, p. 5639–5643, 15 dez. 2014.
- ZHANG, S. et al. A Novel Automated Lazy Learning QSAR (ALL-QSAR) Approach: Method Development, Applications, and Virtual Screening of Chemical Databases Using Validated ALL-QSAR Models. **Journal of Chemical Information and Modeling**, v. 46, n. 5, p. 1984–1995, 1 set. 2006.
- ZHANG, Y. et al. Application of Computational Biology and Artificial Intelligence in Drug Design. **International Journal of Molecular Sciences**, v. 23, n. 21, p. 13568, 5 nov. 2022.
- ZHANG, Z. et al. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. **Bioinformatics**, 2011.

## Produção Científica

### Principais publicações relacionadas a este Projeto de Doutorado:

#### Artigos

1. **Silva, G. M.**, Borges, R. S., Santos, K. L. B., Federico, L. B., Francischini, I. A. G., Gomes, S. Q., et al. (2021). [Revisiting the Proposition of Binding Pockets and Bioactive Poses for GSK-3 \$\beta\$  Allosteric Modulators Addressed to Neurodegenerative Diseases](#). *Int. J. Mol. Sci.* 22, 8252.
2. **Silva, G. M.**, Barcelos, M. P., Poiani, J. G. C., Hage-Melim, L. I. da S., and da Silva, C. H. T. de P. (2019). [Allosteric Modulators of Potential Targets Related to Alzheimer's Disease: a Review](#). *ChemMedChem* 14, 1467–1483.

#### Pôsteres

1. **Silva, G. M.**; Alves, V. M.; Hochuli, J. E.; Silva, C. H. T. P.; Muratov, E. N.; Tropsha, A. *Integrated computational approaches for the discovery of putative GSK-3 $\beta$  allosteric modulators*. In: ACS Spring 2022, San Diego. Bonding Through Chemistry, MEDI3650944.
2. **Silva, G. M.**; Silva, C. H. T. P. *Should cannabidiol and tetrahydrocannabinol act as allosteric modulators of GSK-3?* In: 10th Brazilian Symposium in Medicinal Chemistry, 2022, Poços de Caldas. *BrazMedChem*, 153260.
3. **Silva, G. M.**; Alves, V. M.; Silva, C. H. T. P. *Chemical space exploration and machine learning models for discovering multi-target anti-Alzheimer agents*. In: 45ª Reunião Anual da Sociedade Brasileira de Química - SBQ - Química para o Desenvolvimento Sustentável e Soberano, Maceió, 2022, MED110.
4. **Silva, G. M.**; Silva, C. H. T. P. *Improving the prediction of binding site and pose of allosteric modulators of GSK-3 $\beta$  related to Alzheimer's disease*. In: 9th Brazilian Symposium in Medicinal Chemistry, 2019, Pirenópolis. *BrazMedChem*, SDD27.
5. **Silva, G. M.**; Silva, C. H. T. P. *Structure-Activity Relationship Studies of Allosteric GSK-3 $\beta$  Inhibitors for Future Potential Therapeutics in Alzheimer's disease*. In: 12th International Conference of Pharmaceutical Sciences, 2019, Ribeirão Preto. *CIFARP*, 2019. p. 145.
6. **Silva, G. M.**; Poiani, J. G. C.; Silva, C. H. T. P. *Planejamento computacional de inibidores tradicionais e alostéricos de GSK-3 $\beta$  com interesse em doença de Alzheimer*. In: SBQ - Sul - 2018, Londrina. XXV Encontro de Química da Região Sul, FQ59.

#### Apresentações orais:

1. **Silva, G. M.**; Alves, V. M.; Silva, C. H. T. P. "Chemical space exploration and machine learning models for discovering multi-target anti-Alzheimer agents". Talk at the 45ª RASBQ – 2022, Maceió, AL, Brazil
2. **Silva, G. M.**; Poiani, J. G. C.; Silva, C. H. T. P. "Planejamento computacional de inibidores tradicionais e alostéricos de GSK-3 $\beta$  com interesse em doença de Alzheimer". Talk at the XXV SBQSul – 2018, Londrina, PR, Brazil.

#### Premiações:

1. Winner in 1<sup>st</sup> place of [BioSolveIT's Scientific Challenge Winter 2021](#) with the proposal *Ensemble methods to discover putative multi-target anti-Alzheimer agents*. Received funding to attend an international event 2023-24.
2. Research productivity award by publishing paper classified as A1/A2 (Qualis, CAPES). Received funding to attend the international event CUP XX, organized by OpenEye at Santa Fe, NM, USA, on March 2020. Award issued by Programa de Pós-Graduação em Química (FFCLRP-USP) and CAPES (PROEX).
3. Received Poster Prize from the *Royal Society of Chemistry* (Organic & Biomolecular Chemistry) at the event *BrazMedChem* (Pirenópolis, GO), in 2019.

### Outras publicações (de Set/2018 até o presente):

#### Artigos

1. Santos, I. V., **Silva, G. M.**, Lima, L. R., Bastos, R. S., Ramos, R. S., et al. (2022). [Hierarchical Virtual Screening Based on Rocaglamide Derivatives to Discover New Potential Anti-Skin Cancer Agents](#). *Front. Molec. Biosc.*, 9, 836572.
2. Correia, L. C., Ferreira, J. V., Lima, H. B., **Silva, G. M.**, et al. (2022). [Pharmacophore-based virtual screening from phytocannabinoids as antagonist r-CB1](#). *Journal of Molecular Modeling*, 28, 258.
3. Barros, S. E. L., Lima, H. B., Goncalves, L. K. A., Correia, L. C., Brito, M. F., Barcelos, M. P., **Silva, G. M.** et al. (2022). [Chromatographic profile, in silico and in vivo study of the pharmacokinetic and toxicological properties of major constituent present in kefir: the kefiran](#). *Toxicology Research*, 1, tfac032.

4. Federico, L. B., **Silva, G. M.**, Gomes, S. Q., Francischini, I. A. G., Barcelos, M. P., et al. (2021). [Potential colchicine binding site inhibitors unraveled by virtual screening, molecular dynamics and MM/PBSA](#). *Comput. Biol. Med.* 137, 104817.
5. Rodrigues, D. M., Portapilla, G. B., **Silva, G. M.**, Duarte, A., Rotta, C. G., da Silva, C. H. T. de P., et al. (2021). [Synthesis, antitumor activity and in silico analyses of amino acid derivatives of artemisinin, drupanin and baccharin from green propolis](#). *Bioorg. Med. Chem.* 47, 116372.
6. **Silva, G. M.**, Federico, L. B., da Silva Hage-Melim, L. I., Gomes, S. Q., Barcelos, M. P., Francischini, I. A. G., et al. (2021). [Identification of known drugs as potential SARS-CoV-2 Mpro inhibitors using ligand- and structure-based virtual screening](#). *Future Med. Chem.* 13, 1353–1366.
7. Federico, L. B., **Silva, G. M.**, de Fraga Dias, A., Figueiró, F., Battastini, A. M. O., dos Santos, C. B. R., et al. (2020). [Identification of novel  \$\alpha\beta\$ -tubulin modulators with antiproliferative activity directed to cancer therapy using ligand and structure-based virtual screening](#). *Int. J. Biol. Macromol.* 165, 3040–3050.
8. Castro, L. L., Picanço, L. C. S., Silva, J. V., Souza, L. R., Sousa, K. P. A., Pinheiro, A. A., **Silva, G. M.**, et al. (2020). [Proposition of Potential GSK-3 \$\beta\$  Inhibitors for the Treatment of Alzheimer's Disease: A Molecular Modeling Study](#). *Curr. Comput. Aided. Drug Des.* 16, 541–554.
9. Souza, L. R., Picanço, L. C. S., Brito, M. F. B., Almeida, M. R. S., Marino, B. L. B., Sousa, K. P. A., **Silva, G. M.**, et al. (2020). [Theoretical Study of Monoamine Oxidase B Inhibitors as Drug Candidates for Treatment of Parkinson's Disease](#). *Cent. Nerv. Syst. Agents Med. Chem.* 20, 128–143.
10. Oliveira, N. K. S., Almeida, M. R. S., Pontes, F. M. M., Barcelos, M. P., **Silva, G. M.**, de Paula da Silva, C. H. T., et al. (2019). [Molecular Docking, Physicochemical Properties, Pharmacokinetics and Toxicity of Flavonoids Present in Euterpe oleracea Martius](#). *Curr. Comput. Aided. Drug Des.* 17, 589–617.

### Capítulos de livros

1. Gomes, S. Q., **Silva, G. M.**, Barcelos, M. P., Hage-Melim, L. I. S., Silva, C. H. T. P. (2022). ["Structural Aspects of Organic Compounds as Proteasome Inhibitors Addressed to Several Diseases"](#) in *Research Topics in Bioactivity, Environment and Energy* (Springer Nature).
2. Barcelos, M. P., Gomes, S. Q., Federico, L. B., Francischini, I. A. G., Hage-Melim, L. I. S., **Silva, G. M.**, Silva, C. H. T. P. (2022). ["Lead Optimization in Drug Discovery"](#) in *Research Topics in Bioactivity, Environment and Energy* (Springer Nature).
3. Silva, R. C., Santos, C. B. R., Taft, C. A., **Silva, G. M.**, Silva, C. H. T. P. (2022). ["Advances in Computational Techniques for Discovery and Development of Drugs Against Leishmaniasis: a Brief Review"](#) in *Research Topics in Bioactivity, Environment and Energy* (Springer Nature).
4. **Silva, G. M.**, Federico, L. B., Alves, V. M., and de Paula da Silva, C. H. T. (2021). ["In Silico Methods to Predict Relevant Toxicological Endpoints of Bioactive Substances"](#) in *Functional Properties of Advanced Engineering Materials and Biomolecules* (Springer), 649–676.
5. Federico, L. B., Barcelos, M. P., **Silva, G. M.**, Francischini, I. A. G., Taft, C. A., and da Silva, C. H. T. de P. (2021). ["Key Aspects for Achieving Hits by Virtual Screening Studies"](#) in *Functional Properties of Advanced Engineering Materials and Biomolecules* (Springer), 455–487.

### Pôsteres

1. **Silva, G. M.**; Sullivan, H.; Rath, M. K.; Alves, V. M.; Muratov, E. N.; Silva, C. H. T. P.; Tropsha, A. *Development of Robust and Predictive Machine Learning QSAR Models for Hepatic Stability*. In: *Translating Ideas in Therapies*, 2021, Virtual Poster. The British Pharmacological Society & ELRIG Joint Meeting, P014.
2. **Silva, G. M.**; Federico, L. B.; Barcelos, M. P.; Gomes, S. Q.; Francischini, I. A. G.; Silva, C. H. T. P. *Prospective virtual screening of natural products as potential Sars-Cov-2 Mpro Inhibitors*. In: *EFMC-ISMIC & EFMC-YMCS, 2020*, Virtual Poster Session. Book of Abstracts, P021.
3. Federico, L. B.; **Silva, G. M.**; Campos, J.; Silva, C. H. T. P. *Consensus docking studies for the development of new colchicine binding site inhibitors with potential activity tubulin polymerization destabilizers*. In: *EFMC-ISMIC & EFMC-YMCS, 2020*, Virtual Poster Session. Book of Abstracts, P028.