

MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

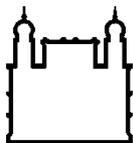
Mestrado em Programa de Pós-Graduação em Biologia Computacional e
Sistemas

Cálculos de Potenciais de Superfície para Análise de Drogabilidade

Gisele Vieira Rocha

Rio de Janeiro

Junho/2015



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

Gisele Vieira Rocha

Cálculos de Potenciais de Superfície para Análise de Drogabilidade

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Computacional e Sistemas

Orientador: Prof. Dr. Floriano Paes Silva-Jr

RIO DE JANEIRO

Junho de 2015

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

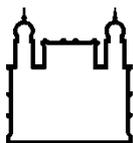
R672 Rocha, Gisele Vieira

Cálculos de potenciais de superfície para análise de drogabilidade /
Gisele Vieira Rocha. – Rio de Janeiro, 2015.
xvii, 126 f. : il. ; 30 cm.

Dissertação (Mestrado) – Instituto Oswaldo Cruz, Pós-Graduação em
Biologia Computacional e Sistemas, 2015.
Bibliografia: f. 122-126

1. Drogabilidade. 2. KVFinder. 3. Cálculo de potencial de superfície
molecular. 4. Lipofilicidade. 5. Densidade de ligações de hidrogênio. I.
Título.

CDD 615.19



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: Gisele Vieira Rocha

Cálculos de Potenciais de Superfície para Análise de Drogabilidade

ORIENTADOR: Prof. Dr. Floriano Paes Silva-Jr

Aprovada em: 24/06/2015

EXAMINADORES:

Prof. Dr. Ernesto Raúl Caffarena - IOC Fiocruz (Presidente)

Prof. Dr. Paulo Sérgio Lopes de Oliveira- - LNBio/CNPEM/SP

Prof. Dr. Ronaldo Alves Pinto Nagem -UFMG-MG

Prof. Dr. Gabriel Limaverde Soares Costa Sousa - IOC/FIOCRUZ/RJ

Prof. Dra. Ana Carolina Ramos Guimarães - IOC/FIOCRUZ/RJ

Rio de Janeiro, 24 de Junho de 2015

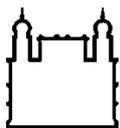
Agradecimentos

Na obra de Guimarães Rosa, “Grande Sertão: Veredas”, há uma passagem que diz “*O correr da vida embrulha tudo, a vida é assim: esquenta e esfria, aperta e daí afrouxa, sossega e depois desinquieta. O que ela quer da gente é coragem*”. Durante a graduação de Física e o período em que cursei o mestrado enfrentei com determinação os desafios que se impuseram. Sob minha perspectiva, digo que falhei comigo mesma algumas vezes, mas nunca desistí e como disse um sábio louco, “o que não nos mata, nos fortalece”. “Aprender” consiste principalmente em “errar”, “acertar” e “errar melhor” no futuro (e assim também caminha a ciência), e sigo em frente com coragem e a certeza de ter dado de mim, o melhor. Agradeço às pessoas que convivi durante estes anos e me ajudaram a concluir mais esta etapa. Agradeço minha mãe, Zezé, meu pai, Arcolino, minha irmã, Lorena, e minha tia Zulma que mesmo distantes me incentivaram e não pouparam esforços para que esta dissertação fosse concluída. Agradeço aos meus tios: Francisco e Sofia que sempre estavam dispostos a resolver quaisquer problemas que surgiam aqui no Rio de Janeiro e os momentos de alegria.

Gostaria de agradecer ao meu orientador o Dr. Floriano Paes por ter me presenteado com um trabalho que tanto me identifiquei, a disponibilidade em auxiliar e guiar o desenvolvimento desta dissertação e principalmente a enorme paciência que demonstrou durante estes meses. Agradeço meus colegas do LABECFar: Dr. Rafael Dantas, Duda, João, Jorge Pina e Mário Senger, os alunos iniciação científica, Ana e Bruno, pelas reuniões que contribuíram para meu aprendizado multidisciplinar e momentos agradáveis que passamos juntos, (obrigada, também, Will e Walter). Quero destacar a ajuda do “espartano” Dr. Mário Senger por ter me auxiliado na escrita da dissertação e me incentivado nos momentos mais difíceis. Agradeço a Pós-Graduação em Biologia Computacional e Sistemas e a agência de fomento da Capes por ter viabilizado o trabalho.

Obrigada à Priscila, Fernanda e a todos os meus amigos do Procc, em especial, Janaína, Lucianna, Vanessa, Rafael, Deborah, Lucas e Amanda pelas divertidas horas e que passamos juntos e as “terapias” que realizamos em grupo.

Por fim, agradeço a meu amado Dunga pelas diversas noites insones e discussões sobre Física, Matemática e um pouco de Biologia para o desenvolvimento das rotinas computacionais, o carinho, incentivo, e o aprendizado em ficção científica que não está relacionado ao trabalho, no entanto, me proporcionou dias felizes e redução da tensão no dia-a-dia.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

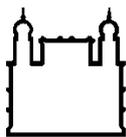
RESUMO

Cálculos de Potenciais de Superfície para Análise de Drogabilidade

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Gisele Vieira Rocha

A seleção de um alvo (macromolécula) com potencial terapêutico é uma etapa chave na maioria dos programas de desenvolvimento de fármacos. A habilidade de prever se uma proteína em particular possui alta afinidade por um determinado fármaco em desenvolvimento não depende somente da estrutura tridimensional dessa: caracterizar suas propriedades físico-químicas é essencial para obter uma alta acurácia na seletividade ao fármaco e prever o processo de reconhecimento deste pela macromolécula. Drogabilidade é, atualmente, o termo empregado para caracterizar a habilidade de uma proteína em ligar-se com alta afinidade e especificidade a um fármaco. O presente trabalho foi executado em duas etapas: (i) implementar um algoritmo presente na literatura de cálculo de potencial lipofílico molecular e de densidade de doadores e aceptores de hidrogênios e (ii) utilizar os resultados obtidos na etapa anterior para dois conjuntos de proteínas, drogáveis e não-drogáveis, de forma a qualificar e quantificar propriedades de cada um dos grupos de macromoléculas. As rotinas computacionais implementadas a partir dos algoritmos para determinar os potenciais citados foram empregadas em conjunto com o KVFinder, programa desenvolvido por pesquisadores do LNBio-CNPq, cuja função é buscar em todos os tipos de proteínas suas cavidades, caracterizando as áreas, os volumes e realizando os mapas eletrostáticos dessas, apresentando grande exatidão em suas soluções. A combinação de programas foi empregada para averiguar cavidades e validar as propriedades físico-químicas de três complexos proteína-fármaco: (i) HIV-1 protease com o VX-478, (ii) a subunidade N-terminal glucoamilase-maltase humana com acarbose e (iii) receptor GABA_B ligado ao agonista baclofen. Os algoritmos apresentaram resultados compatíveis em relação às propriedades dos fármacos e das cavidades quando foram comparados com o *software* de modelagem molecular SYBYL[®]. No que tange às propriedades de drogabilidade, encontramos relações importantes que estabelecem diferenças notáveis entre os dois grupos de estudo. As análises foram baseadas em relações estatísticas simples, mas que originaram resultados que concordam com a literatura e podem ser utilizados no futuro para definições de regras eficazes para discriminar uma proteína drogável de uma não-drogável.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

ABSTRACT

Surface Potential Calculations for Druggability Analysis

MASTER DISSERTATION IN COMPUTATIONAL SYSTEMS BIOLOGY

Gisele Vieira Rocha

Selecting a target (macromolecule) with therapeutic potential is a critical step in most drug development programs. The ability to predict whether a particular protein has high affinity to a specific drug depends on more than just its three-dimensional structure: the description of the physicochemical properties of a drug is essential to obtain with high precision the selectivity of a compound and foresee its recognition process by a macromolecule. The term druggability is employed to characterize the ability of a protein to bind with high affinity and specificity to a drug. The present work was performed in two steps: (i) implementation of an algorithm, found in the literature, that calculates lipophilic potential, hydrogen bond donor and acceptor densities and (ii) usage of the results from the previous step in two sets of proteins, druggable and non-druggable proteins, in order to qualify and quantify properties from each set of macromolecules. The computational routines implemented in the algorithms to determine the mentioned potentials were applied in combination with KVFinder, a program developed by LNBio-CNPEM researchers, whose function is to seek for cavities in all types of protein, describing the area, volume and constructing electrostatic maps of the cavities, showing high accuracy in their outcomes. The combination of programs was employed to evaluate cavities and validate the physicochemical properties of three protein-ligand complexes: (i) HIV-1 protease in complex with VX-478 (1HPV); (ii) the *maltase-glucoamylase N-terminal* subunit in complex with acarbose; and (iii) GABA receptor bound to the agonist Bacoflen. The algorithms showed consistent results regarding the properties of the drugs and cavities when compared to the molecular modelling software SYBYL. Concerning the druggability proprieties, we have found essential associations that distinguish attributes between the two groups. The analyses were based on simple statistical relationships, however capable of originating outcomes that are in agreement with the literature and that might be used in future works to define effective rules to distinguish a druggable protein from a non-druggable one.

Índice	
Resumo.....	viii
Abstract.....	ix
1 Introdução	16
1.1 Planejamento de Fármacos	16
1.1.1 Histórico	16
1.1.2 Mecanismo de ação de um fármaco	19
1.1.3 Limitações e Custos.....	20
1.2 Biologia Estrutural	21
1.3 Integração de técnicas de triagem virtual e triagem biológica no planejamento de fármacos	21
1.3.1 High-Throughput-Screening (HTS)	21
1.3.2 Triagem virtual baseada na estrutura do receptor (TVER)	23
1.4 Atributos <i>in silico</i> na seleção do alvo molecular	24
1.4.1 Busca por cavidades	25
1.5 Propriedades Físico-Químicas	27
1.5.1 Lipofilicidade.....	27
1.6 Potencial de ligação de hidrogênio	33
1.7 Potencial Eletrostático	34
1.8 MOLCAD e SYBYL-X	35
1.9 Drogabilidade	36
1.9.1 Definição	36
1.9.2 Métodos e Limitações	37
1.9.3 Drogabilidade por métodos experimentais	38
1.9.4 Drogabilidade por classes de alvos	38
1.9.5 Avaliação da drogabilidade baseada na estrutura	40
2 Justificativa	43
3 Objetivos	44
3.1 Objetivo Geral.....	44
3.2 Objetivos Específicos:.....	44
4 Materiais e Métodos	45
4.1 Refinamento da superfície molecular do KVFinder	45
4.2 Implementação dos Potenciais de Superfície	47
4.2.1 Potencial Lipofílico.....	47

4.2.2	Densidade de Aceptor e Doador de Hidrogênio	51
4.3	Preparação para análise de drogabilidade	56
4.4	Preparação das proteínas e geração de dados	57
4.5	Seleção das cavidades relevantes para análise dos dados	59
5	Resultados e Discussão.....	60
5.1	Representação da superfície molecular das cavidades protéicas que tangenciam a superfície da proteína.....	60
5.2	Validação dos algoritmos de cálculos de potenciais de superfície molecular	61
5.2.1	Estrutura extracelular do receptor de GABA _B ligado ao agonista baclofen	62
5.2.2	Densidade de aceptor e doador de hidrogênio: Relação Ligante-Cavidade.....	67
5.2.3	Proteína HIV-1 Protease em complexo com o inibidor VX-47871	
5.2.4	Maltase-Glucoamilase Humana e Acarbose.....	79
5.3	Análise do programa em relação ao parâmetro: espaçamento entre as grades.....	88
5.3.1	Eficiência do programa para a execução de grande número de proteínas.....	93
5.4	Análise de Drogabilidade	95
5.4.1	Seleção dos grupos drogáveis e não-drogáveis	95
5.4.2	Análise das características topológicas das cavidades	97
5.4.3	Análise do Potencial Lipofílico.....	99
5.4.4	Análise da Densidade de Ligação de Hidrogênio	109
6	Conclusão	117
7	Referências	119

Índice de Figuras

Figura 1: Representação dos átomos que contribuem para a manifestação do potencial lipofílico.....	32
Figura 2: Função de distância ξ baseada na função de Fermi.....	33
Figura 3: Busca de uma cavidade realizada pelo KVFinder.....	46
Figura 4: Fragmento da tabela contida no artigo de Crippen e colaboradores em que foi realizada a reparametrização dos valores de logP.....	48
Figura 5: Algoritmo de busca de cavidades do KVFinder com ao algoritmo de cálculo do potencial lipofílico.....	50
Figura 6: Fragmento de um arquivo PDB contendo coordenadas dos átomos da metionina (MET) e valina (VAL).	52
Figura 7: Raio de corte em relação a um ponto da superfície.....	53
Figura 8: Representação geométrica da contribuição de cada átomo no cálculo da densidade de hidrogênio em um ponto da superfície.....	54
Figura 9: Complexo proteína-fármaco do servidor PDB contendo o nome genérico do fármaco e o nome comercial, além do link para os bancos de dados DrugBank e Uniprot.....	57
Figura 10: Etapas da preparação das proteínas.	58
Figura 11: Cavidades gerada pelo KVFinder.....	60
Figura 12: Proteína HIV-Protease-1.....	61
Figura 13: Estrutura do baclofen.....	62
Figura 14: Estruturas das subunidades do GABA e seus heterodímeros (subunidades GBR1 e GBR2) com o agonista baclofen.	63
Figura 15: Estrutura do GABA _B e a cavidade onde se liga o agonista baclofen.	64
Figura 16: Potencial lipofílico calculado pelo SYBYL.	66
Figura 17: Potenciais lipofílicos simulados com o KVFinder.....	66
Figura 18: Densidade de acceptor de hidrogênios do Baclofen.....	68
Figura 19: Densidade de doadores de hidrogênio da cavidade.....	68
Figura 20: Densidade de doador de hidrogênios do Baclofen.....	69
Figura 21: Densidade de aceptores de hidrogênio da cavidade.....	69
Figura 22: Influência dos resíduos na densidade de doador de hidrogênios na cavidade.....	70

Figura 23: Influência dos resíduos na densidade de acceptor de hidrogênios na cavidade.....	71
Figura 24: Composto inibidor VX-478.	72
Figura 25: Cavidade e ligante do complexo 1HPV.	73
Figura 26: Potencial lipofílico do ligante VX-478..	74
Figura 27: Potencial lipofílico da cavidade onde se encontra o ligante VX-478.	74
Figura 28: Representação das regiões acceptoras de hidrogênios da molécula VX-478..	75
Figura 29: Densidade de doador de hidrogênios para a cavidade onde se liga o VX-478..	76
Figura 30: Representação do ligante VX-478 em que apresenta densidade de doadores de hidrogênio.....	76
Figura 31: Potencial acceptor de hidrogênio apresentado pela cavidade do ligante VX-478.	77
Figura 32: Resíduos que interagem com o ligante VX-478.....	78
Figura 33: Representação da cavidade e resíduos da macromolécula HIV-1 protease.	79
Figura 34: Estrutura da acarbose.....	80
Figura 35: Região onde se localiza a estrutura da acarbose.	81
Figura 36: Acarvosina presente na cavidade de ligação com a glucoamilase.	82
Figura 37: Cavidade gerada pelo KVFinder onde a acarvosina se liga à glucoamilase humana..	82
Figura 38: Cavidade onde está presente a acarvosina..	83
Figura 39: Molécula de acarbose com densidade de acceptor de hidrogênio mostrada com superfície molecular.....	84
Figura 40: Cavidade da glucoamilase com densidade de doadores de hidrogênio.....	85
Figura 41: Molécula de acarbose (A) com densidade de doador de hidrogênio em sua superfície molecular(B).....	85
Figura 42: Densidade de acceptores gerada pelo KVFinder.....	86
Figura 43: Relação cavidade-resíduo para densidade de acceptores de hidrogênio.....	87

Figura 44: Região da cavidade onde apresenta maior probabilidade de doador de hidrogênio.....	87
Figura 45: Variação da área total de sete cavidades para quatro valores de espaçamento da grade.....	89
Figura 46: O volume da cavidade para seis proteínas diferentes.....	90
Figura 47: O potencial lipofílico total das cavidades de seis proteínas com o ligante.....	91
Figura 48: Hidrofobicidade da cavidade para cada valor da grade. A hidrofobicidade segue um comportamento semelhante à área e volume.	92
Figura 49: Hidrofilicidade para cada cavidade onde o ligante está presente. .	92
Figura 50: Tempo de execução do programa KVFinder em função do número de átomos das proteínas.....	94
Figura 51: Boxplot do Potencial isoeletrico.....	95
Figura 52: Boxplot em relação ao número de átomos das proteínas drogáveis e não drogáveis.....	96
Figura 53: Boxplot do Peso molecular.....	96
Figura 54: Boxplot relativo à área superficial das cavidades das duas amostras.	97
Figura 55: Gráfico da razão área superficial e área total de 193 proteínas drogáveis.....	98
Figura 56: Boxplot do volume das cavidades dos grupos drogáveis e não-drogáveis.....	99
Figura 57: Gráficos da densidade de Kernel para o Potencial Lipofílico para as proteínas não drogáveis.....	100
Figura 58: Gráficos do Potencial Lipofílico para o grupo não-drogável..	101
Figura 59: Lipofilicidade apresentada pelas amostras drogáveis e não-drogáveis	102
Figura 60: Comparação entre grupos drogáveis e não-drogáveis quanto a hidrofobicidade.....	102
Figura 61: Relação entre a hidrofilicidade de cada grupo.	103
Figura 62: Boxplot da área hidrofóbica entre as duas amostras.	104
Figura 63: Relação entre as áreas hidrofílicas de cada amostra.....	105
Figura 64: Relação entre o valor de hidrofobicidade e área total da cavidade..	106

Figura 65: Regressão linear dos valores de hidrofiliidade calculados em relação a área total da cavidade.	106
Figura 66: Regressão linear do descritor "área hidrofóbica" em relação a área superficial total	107
Figura 67: Regressão linear da área hidrofílica em relação à área superficial total.	108
Figura 68: Gráfico da razão da área hidrofóbica das amostras drogáveis em relação à área total da cavidade.	109
Figura 70: Boxplot para densidade de aceptores e doadores de hidrogênio entre os dois grupos.....	109
Figura 71: Distribuição de Kernel para número absoluto de aceptores para os conjuntos drogável e não-drogável.	110
Figura 72: Densidade de Kernel para número absoluto de doadores de hidrogênio para os dois grupos..	111
Figura 73: Boxplot do número absoluto de doadores de hidrogênio para as duas amostras.....	112
Figura 74: Comparação entre número absoluto de aceptores de hidrogênio..	112
Figura 75: Densidade de aceptores e doadores de hidrogênio para os grupos drogáveis e não drogáveis.	114
Figura 76: Relação de linearidade para o número de aceptores e doadores de hidrogênio para o grupo drogável.....	114
Figura 77: Relação entre a área de doadores e aceptores de hidrogênio para grupos drogáveis e não drogáveis.	115
Figura 78: Regressão linear obtida para a área de influência dos doadores e aceptores de hidrogênio.....	116
Figura 79: Regressão linear obtida para a área de influência dos doadores e aceptores de hidrogênio.....	116

1 Introdução

1.1 Planejamento de Fármacos

1.1.1 Histórico

O desenvolvimento da farmacologia, enquanto prática de base científica, se iniciou no século XVIII com o progresso da química e aplicação interdisciplinar dos seus princípios e métodos(1). No século XIX, a hipótese de Avogadro¹ foi confirmada e a tabela periódica dos elementos químicos foi organizada de acordo com a massa atômica, sofrendo correções pela valência, manifestada através das propriedades comuns de cada grupo ou família. Em 1865, August Kekulé formulou sua teoria pioneira sobre a estrutura de moléculas orgânicas aromáticas. A teoria do benzeno deu impulso à evolução de corantes químicos e à síntese de novos compostos orgânicos importantes no estudo da farmacologia(2).

O isolamento e a purificação de ingredientes de plantas medicinais, com o auxílio da química analítica, demonstraram seu valor na rotina médica do século XIX. Em 1815, F. W. Sertürner isolou a morfina do extrato de ópio. A papaverina foi descoberta no mesmo ano, mas suas propriedades antiespasmódicas não foram conhecidas até 1817(3). Devido à disponibilidade dessas e outras substâncias ativas de diversas plantas, as farmácias passaram a fornecer preparações padronizadas, mas estas muitas vezes eram impuras(4).

As mais recentes descobertas que revolucionaram a farmacologia incluem o estudo elaborado por John Newport Langley(5), que estabeleceu o conceito químico de transmissão sináptica – mecanismo sustentado por neurotransmissores que medeia as transferências de informações de uma célula nervosa para outra. Ele determinou que extratos de adrenomedulares, contendo epinefrina e norepinefrina, induziram respostas em diferentes tecidos, sendo essas, semelhantes às provocadas pelos nervos sinápticos. Baseado nestes

¹ “Volumes iguais de quaisquer gases, nas mesmas condições de temperatura e pressão, apresentam a mesma quantidade de matéria em mol ou moléculas”. Avogadro postulou que um mol de quaisquer gases contém a mesma quantidade de moléculas, e o volume ocupado também será o mesmo desde se conservem as mesmas condições de temperatura e pressão. Isso foi determinante para a caracterização *atômica* da matéria.

resultados, Langley postulou em 1905 que uma “substância receptiva” era o local de ação de mediadores químicos liberados pela estimulação do nervo(3).

Paul Ehrlich, em 1900, já havia introduzido o termo “receptor” para designar locais na superfície da célula com características moleculares específicas em que determinadas substâncias interagem de forma pontual, semelhante ao modelo chave-fechadura, introduzindo o termo “farmacologia terapêutica”(3). Em 1910, com base nos trabalhos de Langley, Erlich desenvolveu sua própria teoria de receptor de ligação seletiva de toxinas e substâncias nutritivas, que denominou quimiorreceptores. O conceito revisto tornou-se base para a descoberta da arsfenamina, denominada salvarsan arsênico, primeiro agente quimioterapêutico utilizado para o tratamento da sífilis(3).

A farmacologia, por conseguinte, começou a ganhar seu espaço nas disciplinas médicas. Novas instituições foram criadas para apoiar pesquisas interdisciplinares com o intuito de desenvolver novos medicamentos. O novo caminho de encontrar, caracterizar e elaborar novos fármacos levou à formação de uma nova indústria(4).

A primeira metade do século XX foi marcada por descobertas e novas tecnologias que impulsionaram o desenvolvimento de novos medicamentos e terapias. A penicilina, descoberta por Alexander Fleming em 1929, abriu a porta para uma era de tratamentos às infecções bacterianas. Fleming trabalhou como médico na 2ª Guerra mundial e o predomínio de infecções secundárias de estafilococos em feridas influenciou seu interesse no estudo de antissépticos, investigando uma série de substâncias que interferem no crescimento bacteriano(6). Em 1922, ele isolou uma das vias nasais de um doente que sofria de rinite aguda uma substância antibacteriana que nomeou de lisozima(6).

Alexander Fleming possuía domínio no estudo da lise bacteriana e a descoberta da penicilina, embora classificada como “*Serependity*”, só foi possível devido ao seu conhecimento prévio sobre o assunto².

A grande descoberta de Fleming iniciou-se durante a sua ausência no laboratório, neste período, esporos se estabeleceram em uma placa de Petri que continha estafilococos. Ao retornar, o cientista observou que o mofo sobre a cepa de estafilococos produzia uma substância bactericida. Fleming passou a estudar

² “O acaso só favorece a mente preparada” - Louis Pasteur.

as características das substâncias encontradas e, com a ajuda de C.L. Latouche, descobriu que o fungo pertencia ao gênero *Penicillium*. Fleming cunhou a palavra penicilina e observou que a substância matava estreptococos, pneumococos, meningococos e bacilos, sendo atóxico aos animais(4).

Neste período, também foi introduzido o conceito de enzimas e receptores, que foi empiricamente utilizado para descrever alvos que poderiam se ligar a fármacos. Este foi um conceito bioquímico que auxiliou na descrição e caracterização da anidrase carbônica em 1933(7), uma enzima que possui papel importante no transporte de CO₂ e no controle de pH do sangue. Em seguida foi descoberta a sulfonamida, grupo de antibióticos sintéticos usados no tratamento de doenças infecciosas. Compostos derivados desse grupo originaram fármacos no tratamento de diabetes Mellitos do tipo II, edemas, glaucomas e até hipertensão(4).

O aumento explosivo de novos tratamentos e fármacos aconteceu após a 2ª Guerra Mundial. A descoberta de antibióticos seguros e eficazes foi seguida por eficientes tratamentos de doenças cardíacas, hipertensão e asma. Vale ressaltar o desenvolvimento da psicofarmacologia, instituindo o tratamento relacionado aos medicamentos psiquiátricos refletindo uma projeção de substâncias psicoativas. Essa expansão aprofundou uma nova visão sobre o uso de novos fármacos e suas propriedades farmacológicas que abrangem as propriedades físico-químicas, composição, efeitos bioquímicos, fisiológicos e usos terapêuticos(5).

Em paralelo, tem-se uma das descobertas mais importantes do século XX. James Watson e Francis Crick, no início da década de 50, publicam na *Nature*(8) a obra clássica sobre a estrutura em hélice da molécula de DNA. Este trabalho deve ser visto como contribuição fundamental para o entendimento de moléculas biológicas. O conhecimento estrutural de replicação e tradução do material genético, a partir das bases do DNA, permitiu entender como se dava a organização dos nucleotídeos.

A primeira análise da estrutura tridimensional de proteínas ocorreu com os trabalhos de Linus Pauling e Robert Corey(9), em 1950, e Gopalasamudram N. Ramachandran(10), em 1960. A sistematização do trabalho acerca da estrutura tridimensional, seguido da informação genética, se deve ao papel chave de Margaret Dayhoff, uma das pioneiras no uso de computadores para o estudo de

biomoléculas(11). Dayhoff desenvolveu uma série de matrizes de probabilidade que explicavam as mudanças de aminoácidos ao longo do tempo, baseando na teoria do alinhamento de sequências de proteínas.

Com o progresso computacional tornaram-se possíveis as primeiras análises com programas de dinâmica molecular sobre o enovelamento de proteínas(12,13). O estudo da estrutura de proteínas busca o entendimento das moléculas e eventos mediados por elas. Esse processo inclui a identificação do modo de interação entre moléculas (*docking*), avaliação na estrutura e ambiente molecular na dinâmica, seleção de compostos com maior potencial de inibição e função das moléculas. Estes últimos cresceram decorrentes da conclusão do projeto genoma humano(14). O impacto desta iniciativa resultou em uma melhoria nos processos terapêuticos acessíveis à população, no entanto, guiar o planejamento de um novo composto ainda é uma estratégia complexa(15).

1.1.2 Mecanismo de ação de um fármaco

A ação de um fármaco é resultado da interação do mesmo com uma macromolécula (geralmente, proteína) do organismo. Os ligantes com características de fármacos modulam as funções das proteínas derivando reações bioquímicas e fisiológicas em resposta a esta alteração. É importante ressaltar que o fármaco não cria efeitos, mas modula funções fisiológicas distintas. Os fármacos que se ligam aos receptores fisiológicos e mimetizam os efeitos reguladores dos compostos endógenos de sinalização são denominados agonistas. Os compostos que se ligam sem efeito regulador e bloqueiam a sinalização dos agonistas são denominados antagonistas(2).

A ligação de um fármaco ao seu alvo envolve interações iônicas, ligações de hidrogênio, interações hidrofóbicas, hidrofílicas e covalentes. Adicionado a estes mecanismos, a eficácia geral de um fármaco está relacionada aos efeitos de farmacocinética ou ADMET (absorção, distribuição, metabolismo, excreção e toxicidade). A absorção descreve a taxa de saída do fármaco de seu local de administração e extensão com que este processo ocorre. Após a absorção, a distribuição do fármaco ocorre nos líquidos intersticiais e intracelulares. Esse processo inclui fatores fisiológicos e propriedades físico-químicas específicas de

cada fármaco. A fase de distribuição pode levar minutos ou várias horas, até que haja um equilíbrio entre a concentração do fármaco no tecido e sangue(2).

A conversão metabólica de um fármaco possui natureza enzimática e sua concentração excessiva em um sítio receptor ou promiscuidade, que é a possível interação com várias proteínas, podem acarretar em toxicidade para a célula.

1.1.3 Limitações e Custos

A interação do fármaco com o complexo sistema biológico pode resultar em um efeito diferente do esperado. Um dos maiores problemas na obtenção de um fármaco com eficácia desejável está relacionado principalmente à farmacocinética. Estes efeitos determinam a “concentração da droga em seu sítio de ação e por consequência a intensidade de seus efeitos em função do tempo”(2). No desenvolvimento de um novo fármaco uma das maiores frustrações se baseia na necessidade de uma molécula líder alcançar o local de ação, ou seja, a macromolécula. Dependendo dos recursos disponibilizados, somente parte destes alvos pode ser estudado.

Entre 1975 e 2004, o número de fármacos inseridos no mercado foi de 1556. Esse número sugere um avanço para saúde e medicina, no entanto, não beneficia a todos. De acordo com a pesquisa realizada pela World Health Organization (WHO) em 2007(16), durante o período citado, 179 novos fármacos para doenças cardiovasculares foram inseridos no mercado. Para doenças tropicais e tuberculose foram desenvolvidos 21 fármacos, durante o mesmo período. As companhias farmacêuticas focam em países com altos índices de crescimento econômico e social ignorando países em desenvolvimento e de baixos índices sociais.

Aproximadamente 100 bilhões de dólares são gastos por ano em pesquisa e desenvolvimento em saúde, sendo que dez por cento deste total são gastos em doenças que afetam países em desenvolvimento. O número de indústrias farmacêuticas concentradas na América do Norte, Japão e Europa representa 88% do valor global, visando o mais atrativo potencial de comercialização(16).

1.2 Biologia Estrutural

Sabe-se que o conhecimento das estruturas tridimensionais das proteínas desencadeou o uso da biologia estrutural com a aplicação focada nas globinas, enzimas e hormônios polipeptídeos. A princípio, o emprego destas estruturas guiou a síntese de ligantes da hemácia para diminuir a forma em foice da mesma e melhorar o transporte de O₂ pelo sangue(17), a modificação química da insulina para melhorar a meia-vida na circulação(18) e o planejamento de inibidores de serino-proteases para controlar a coagulação do sangue(19). Estas estruturas foram analisadas através da técnica de cristalografia de raio-x(20).

O conhecimento prévio acerca da estrutura tridimensional das proteínas impulsionava o planejamento de novos fármacos, entretanto, muitas dessas macromoléculas não eram acessíveis diretamente por cristalografia de raios-x. A alternativa encontrada, na década de 80, foi usar os compostos líderes para serem explorados por modelos comparativos baseados na modelagem por homologia(21). O método foi utilizado para modelar a renina a partir da estrutura da aspartil protease. As estruturas foram úteis para obter as topografias das superfícies proteicas complementares ao ligante, podendo ser explorada para otimizar a potência e seletividade. Eventualmente, estruturas cristalinas de alvos com atividade terapêutica relevante tornaram-se disponíveis, fármacos no combate a AIDS como a Agenerase e Viracept, foram desenvolvidos utilizando-se a estrutura cristalina da HIV-protease(22).

1.3 Integração de técnicas de triagem virtual e triagem biológica no planejamento de fármacos

1.3.1 High-Throughput-Screening (HTS)

A indústria farmacêutica como descrita inicialmente sofreu mudanças significativas ao longo dos anos. Na década de 70, os compostos “líderes” eram descobertos usando ensaios *in vitro* e triagem *in vivo*. A fonte destes compostos surgia, principalmente, de produtos naturais e observações clínicas de efeitos colaterais(23). A maioria destes compostos tinha em comum uma atividade

terapêutica(24) que, *a priori*, havia passado por uma considerável investigação científica para ser identificada como uma molécula líder. Do ponto de vista do comportamento químico, os compostos que apresentavam o pior desempenho em uma série de análogos eram eliminados.

Esta situação mudou dramaticamente com o advento do *High Throughput Screening* (HTS) ou triagem de alta vazão que tornou tecnicamente viável rastrear centenas de milhares de compostos *in vitro* através de ensaios que propiciam a avaliação biológica destes compostos contra uma proteína alvo ou um sistema biológico definido(25,26). Além disso, cresce cada vez mais o número de coleções de compostos graças à química combinatória que permitiu a síntese automatizada de números maciços de compostos para a triagem de HTS.

A geração da base de dados é influenciada por diversos fatores, destacando-se as propriedades moleculares e físico-químicas. Neste contexto Lipinski e colaboradores(24) definiram quatro parâmetros que devem estar associados à solubilidade e permeabilidade do composto: o peso molecular, coeficiente logarítmico de partição (LogP), o número de doadores de hidrogênio e o número de aceptores de hidrogênio. A análise resultou em uma técnica mnemônica chamada “regra dos cinco” ou “regra dos cinco de Linpinski” porque o valor de corte dos quatro parâmetros são todos próximos a cinco ou múltiplos de cinco. Para a regra dos cinco, as propriedades que levam à má absorção e menor permeabilidade são:

- i. O peso molecular maior que 500 Da.
- ii. LogP maior que 5;
- iii. Número de doadores de hidrogênio maior que 5.
- iv. Número de aceptores de hidrogênio maior que 10.

As classes de compostos que são ligantes para transportadores biológicos são exceções para estas regras.

Em contrapartida, o conceito de líder-similar (em inglês, *lead-like*) surgiu para identificar moléculas com propriedades semelhantes às de compostos líderes. Essas substâncias bioativas devem apresentar propriedades similares às dos compostos que serviram como ponto de partida para o desenvolvimento de novas entidades químicas passando por um processo de otimização(27).

Um dos métodos mais confiáveis em química medicinal para melhorar a atividade *in vitro* é incorporar grupos lipofílicos adequadamente posicionados. Por exemplo, a adição de um grupo metila que ocupa uma cavidade contribui para diminuir a energia de ligação em 0,7 Kcal/mol(28). Por outro lado, é geralmente difícil melhorar a potência *in vitro* por manipulação de grupos polares que são envolvidos em interações iônicas com o receptor. Grupos polares nos fármacos que não são necessários para a ligação ocorrer podem ser tolerados se ocupam espaço no solvente mas não são adicionados na ligação com o receptor(24).

O HTS, invariavelmente, leva à identificação de grande número de sucessos (*hits*), sendo que apenas uma modesta fração se apresenta como ligantes verdadeiros da enzima alvo. Neste caso, um número expressivo de sucessos não corresponde a mecanismo de inibição da enzima por mecanismos específicos, por exemplo, inibição competitiva, mas pode estar associado à inibição promíscua ou fatores de agregação. Após a identificação destes sucessos, faz-se necessária a confirmação e validação dos resultados por ensaios secundários(27).

1.3.2 Triagem virtual baseada na estrutura do receptor (TVER)

A triagem virtual baseada na estrutura do receptor avalia um extenso banco de dados simulando interações entre os fármacos e alvos moleculares. O método em questão permite a descoberta de ligantes com diversidade estrutural, entretanto, é necessário o conhecimento prévio da estrutura tridimensional da proteína. A macromolécula pode ter sua estrutura elucidada experimentalmente ou através de metodologias computacionais como modelagem comparativa(27).

Esta técnica necessita de uma delicada preparação, com cuidados a serem realizados na preparação do ligante. A base de dados que contém o fármaco necessita do cálculo das cargas parciais atômicas e a representação correta dos estereoisômeros, tautômeros e dos estados de protonação(29,30). A proteína necessita estar no estado adequado de protonação, deve haver a escolha correta do modelo tridimensional e a cavidade para a ligação deve ser previamente definida. A macromolécula geralmente é tratada de forma rígida ou com flexibilidade limitada.

As moléculas de água contidas no sítio ativo podem participar das interações intermoleculares durante a interação proteína-ligante, mas resulta em um grande custo computacional.

Depois de preparados, o ligante e o receptor, realiza-se a docagem molecular na qual são analisadas as interações intermoleculares dos resíduos do sítio de interação da cavidade alvo. As poses ou conformações dos ligantes dentro da cavidade são classificadas de acordo com uma função de pontuação. Existem vários programas de docagem que permitem a realização deste procedimento. As funções de pontuação podem ser baseadas em: (i) princípios físico-químicos, processo em que se analisam cálculos de interações eletrostáticas, van der Waals e de dessolvatação; (ii) funções experimentais em que fatores entrópicos e entálpicos envolvidos no reconhecimento proteína-ligante têm seu peso ajustado para reproduzir um conjunto experimental de dados, e por último; (iii) funções com base em conhecimentos que utilizam dados estatísticos provenientes de complexos já reconhecidos e que estabelecem parâmetros moleculares úteis.

Os estágios particulares que merecem cuidados significativos são a disponibilidade comercial de moléculas para a síntese orgânica e a avaliação experimental, onde os ensaios *in vitro* devem ser padronizados e validados. A conclusão destas etapas fornece *hits* que podem se tornar ligantes reais para a proteína-alvo determinando propriedades como potência, afinidade e seletividade.

1.4 Atributos *in silico* na seleção do alvo molecular

O reconhecimento de um alvo é um passo essencial na exploração de novos medicamentos. A identificação e validação da macromolécula como terapeuticamente significativa irão guiar o reconhecimento de novos fármacos contra estes alvos. No início, a inserção de um fármaco no mercado era conduzida, principalmente, pela química e farmacologia. Com o sequenciamento genômico este processo se expandiu para outras áreas além da Biologia, como Matemática, Física e Computação. Os programas destinados ao planejamento de moléculas bioativas se iniciam com a validação do alvo, que na maioria dos

casos são biomoléculas como proteínas receptoras, enzimas e canais de íons(31).

1.4.1 Busca por cavidades

Um dos processos iniciais para a análise de um alvo com potencialidade terapêutica reside na busca por cavidades nas estruturas tridimensionais. Muitas aproximações para a busca destas cavidades são realizadas. Existem métodos ou algoritmos baseados em geometria como, por exemplo, SURFNET(32), LIGSITE(33) e CAST(34). Simplificando, estes algoritmos partem da premissa que os locais de ligação devem ser invaginações côncavas na superfície das proteínas. Os algoritmos que se baseiam em energia levam em considerações alguns processos físicos por cálculos de potenciais de ligação ou energias de ligação, sendo exemplos o GRID(35) e o DogSite(36).

O programa que utilizamos durante o trabalho é o KVFinder(37), desenvolvido por pesquisadores do Laboratório Nacional de Biociências-LNBio. Esta ferramenta possui um modelo versátil e fácil para usar na prospecção de cavidades e caracterização espacial. O KVFinder apresenta um conjunto customizável de parâmetros que pode ser empregado como um *plugin* do Pymol ou ser utilizado através de linha de comando. O fato de ser um programa com código aberto além da possibilidade de sua utilização em pipeline foram de grande valor para a execução do trabalho. No próximo tópico, o KVFinder será descrito com maiores detalhes.

1.4.1.1 O KVFinder

Esta é uma ferramenta que realiza a busca de cavidades baseando-se em um grid geométrico, que apresenta algumas capacidades distintas como a segmentação das cavidades(37). A segmentação espacial gera a representação destas cavidades com alta resolução. Para validar a ferramenta, os autores utilizaram um *benchmark* com 198 drogas e compararam com outras ferramentas. A simulação consistiu em fazer uma busca por todas as cavidades e a predição foi considerada correta quando o centro geométrico da cavidade se encontrava a uma distância de 4 Å de qualquer átomo ligante. As cavidades foram baseadas no volume e as três principais cavidades foram analisadas(36).

A taxa de sucesso foi avaliada para três casos, classificados como: *Top 1*, *Top 2* e *Top 3* que, no artigo, referia-se a uma, duas e três melhores cavidades, respectivamente. De acordo com os resultados o KVFinder teve a melhor taxa de sucesso com o *Top 1* e *Top 3*, ficando atrás somente do método de predição MetaPocket2.0, que combina o resultado de mecanismo de predição de muitos programas. Em um segundo teste, para avaliar o desempenho do KVFinder usando uma customização do parâmetro *probe out*, que será abordado a seguir, o KVFinder se tornou apto no desempenho para todos os três métodos, alcançando 76% de taxa de sucesso no teste *Top 3*(37). Os resultados podem ser vistos na Tabela 1 onde há a comparação do KVFinder com as outras ferramentas.

Tabela 1: Sucesso na predição de cavidades ligantes do KVFinder, em relação às outras ferramentas, para 198 estruturas alvos de fármacos conhecidos.

Method	TOP 1	TOP 2	TOP 3
KVFinder Combined Prediction	67	74	76
MetaPocket 2.0	61	70	74
KVFinder	51	56	62
LIGSITE(33)	48	57	61
Q-SiteFinder(38)	40	54	62
SURFNET	24	30	34
ConCavity	47	53	56
Fpocket(39)	31	48	57

1.4.1.2 Modelo Teórico

O KVFinder(37) define cavidades geometricamente, baseado na definição da teoria de morfologia matemática(40). A proteína é inserida em uma grade tridimensional e os pontos são tipificados como pertencentes ou não a um espaço com átomos. Duas sondas são utilizadas para definir a superfície molecular da proteína e as cavidades (36). O algoritmo usado no software usa uma sonda A (*probe in*) com raio de 1.4Å para definir espaço entre a proteína e o meio. A sonda percorre toda a grade, os pontos situados nos espaços preenchidos com átomos são definidos como “Falso” (rotulados como 0) e os espaços não preenchidos por átomos (vazios) são definidos como “Verdadeiro”

(rotulados como 1). O mesmo processo é feito com a sonda B (*probe out*) que não possui acesso ao espaço vazio dentro da proteína, gerando dessa forma uma superfície que define o limite entre a cavidade e o espaço exterior à macromolécula(41).

Para o cálculo do volume é feito um somatório dos cubos (voxels) dimensionados pela grade dentro da cavidade. Multiplica-se o número de pontos (**n**) pelo volume de um cubo, ou seja, sendo **h** a distância entre os pontos da grade, o volume da cavidade será $n \cdot h^3$. O cálculo da área é feito pela soma do número (**n**) de pontos que estão na superfície que delimita a cavidade seguido pela multiplicação por h^2 (37,41).

A superfície apresentada pelo KVFinder é descrita pela área da cavidade que tangencia a superfície da proteína e a área da cavidade que fica exposta ao solvente, ou seja, a “tampa” da cavidade. Assim como outros programas (e.g., PocketFinder(42) e Q-SiteFinder(38)) e servidores (e.g., DogSite(43)), a área calculada é representada pela superfície que envolve todo o volume da cavidade (superfície molecular e a “tampa”). Portanto, estão incluídos pontos que tangenciam a proteína e pontos da fronteira entre a cavidade e o ambiente em que a macromolécula está contida.

1.5 Propriedades Físico-Químicas

1.5.1 Lipofilicidade

As primeiras descrições sobre as interações hidrofóbicas(44) já enfatizavam a importância das forças entre os fármacos e seu ambiente durante o transporte, reconhecimento e processos de ligação com receptores. Estas interações são muito importantes no caso de proteínas em um ambiente aquoso ou em uma membrana e são reconhecidas como um dos principais fatores para o enovelamento de proteínas.

A interação fármaco-receptor age de forma similar às forças intermoleculares que agem no particionamento do soluto entre a água e a fase orgânica imiscível(45). O parâmetro que descreve este particionamento,

denominado coeficiente de partição logarítmica (logP), está correlacionado com a atividade biológica(46). O efeito lipofílico pode ser descrito e quantificado utilizando-se a mudança na energia livre para transferir a molécula da fase polar para a apolar.

Na literatura é habitual usar o termo lipofilicidade para se referir somente à hidrofobicidade. No presente trabalho, o termo lipofilicidade pode indicar valores de hidrofobicidade e hidrofilicidade. Enquadrando em uma escala, quando mencionamos que uma superfície possui alta lipofilicidade significa que ela é hidrofóbica ou apolar e possui baixa hidrofilicidade (polaridade). Uma superfície com baixa lipofilicidade significa que a superfície é hidrofílica ou polar.

O termo potencial lipofílico é definido considerando-se uma molécula cercada por um solvente polar e toda a lipofilicidade desta molécula pode ser obtida através das contribuições fragmentais ou atômicas. Estas subdivisões da lipofilicidade são pré-requisito para definir o mapa lipofílico de uma superfície. Desta forma, é razoável postular uma regra dependente da distância que reporta a influência de diferentes fragmentos da lipofilicidade em um determinado ponto da superfície molecular(47).

1.5.1.1 Coeficiente de Partição Logarítmica

O coeficiente de partição de um soluto em dois solventes pode ser aproximado como a razão de sua solubilidade nestes. O coeficiente de partição logarítmico está diretamente relacionado à mudança na energia livre durante a transferência do soluto de um solvente para o outro(48). Este parâmetro está correlacionado com o volume molar, densidade de carga e polarizabilidade.

A utilização de parâmetros atômicos físico-químicos para a definição de lipofilicidade, com intuito de ser amplamente utilizada no mapeamento de estruturas tridimensionais, foi inicialmente proposta por Hansch e colaboradores(49) e Rekker e colaboradores(50,51). Posteriormente, Ghose e Crippen(52) desenvolveram um trabalho com o objetivo de encontrar parâmetros de hidrofobicidade mais refinados e amplamente aplicáveis para serem usados na interação entre ligante-receptor. Os parâmetros, denominados como

coeficiente de partição da molécula no sistema água-octanol, referem-se à medida da lipofilicidade de cada átomo ou fragmento.

Para a parametrização de lipofilicidade de átomos individuais calculam-se as suas contribuições em um conjunto de pequenas moléculas. Enquanto estas propriedades não são estritamente aditivas, as interações intramoleculares podem ser explicadas por classificar átomos em diferentes tipos com base em átomos ligados e vizinhos(46,52). O coeficiente de partição octanol-água é determinado a partir da equação:

$$\log P = \sum_i n_i a_i$$

Equação 1

Onde n_i é o número de átomos do tipo i e a_i é a contribuição correspondente ao átomo do tipo i . Vários tipos de átomos foram estudados sendo que os fatores que resultaram na sua classificação são relativos à hibridização do átomo, a densidade formal de cargas e acessibilidade da molécula solvente em torno do átomo.

1.5.1.2 Potencial Lipofílico Molecular (PLM)

Em estudos preliminares realizados por Crippen e colaboradores(46), o pré-requisito para determinar a lipofilicidade de toda a molécula, medido pelo logaritmo do coeficiente de partição, é obter uma superposição das contribuições atômicas. Para definir o mapa superficial desta quantidade é plausível postular uma regra dependente da distância e encontrar a influência de diferentes fragmentos em um ponto na superfície. Inicialmente, Audry e colaboradores(53) propuseram, em analogia ao Potencial Eletrostático, uma função que denominou “Potencial Lipofílico Molecular” (PLM_A):

$$PLM_A = \sum_i \frac{\sum_i f_i}{(1 + f_i)}$$

Equação 2

Onde f_i é a lipofilicidade parcial do i -ésimo fragmento da molécula e d_i é a distância do ponto na superfície da molécula ao centro do fragmento i . A longa distância, a semelhança da função com o Potencial Coulombiano faz com que um grande número de átomos distantes da superfície seja dominante no cálculo do PLM_A (45). Para pequenas moléculas a função PLM_A é satisfatória. Fület e colaboradores(54) calcularam o perfil do PLM para 20 aminoácidos usando os valores de lipofilicidades de átomos simples postulados por Crippen(52). Fauchère(55) propôs outra forma de determinar o PLM , definindo a contribuição fragmental em relação à superfície da forma $PLM \sim e^{-d}$. Esta função do potencial foi designada com base nos efeitos de proximidades observados por Rekker(51) e Hansch e Leo(49), sendo utilizado para a observação o sistema fármaco-receptor. Ambos os modelos obtiveram sucesso para pequenas moléculas.

Sabendo que as contribuições fragmentais de LogP são baseadas em contribuições locais Heiden e colaboradores(47) definiram uma aproximação de PLM que foi particularmente designada para mapear a hidrofobicidade local na superfície acessível ao solvente de macromoléculas baseados nos valores propostos por Crippen e Ghose(56). Heiden(47) introduziu uma nova forma de calcular o PLM dada por:

$$PLM = N^{-1} \sum_i f_i \xi(d_i)$$

Equação 3

Onde $\xi(d_i)$ é uma função dependente da distância d_i do átomo i à superfície da molécula e N é o fator de normalização:

$$N = \sum_i \xi(d_i)$$

Equação 4

A função $\xi(d)$ deve ser suave e ter valores finitos para $d < d_{corte}$, sendo d_{corte} denominado *distância de proximidade* e deve ser maior que qualquer raio de Van der Waals de um átomo da molécula sob consideração. Para distância $d < d_{corte}$, o valor da função deve cair rapidamente ao valor zero. A normalização da função representada por N garante que os valores de lipofilicidade gerados na superfície sejam coerentes. A função PLM representa um peso médio sobre todos os

valores de f_i , logo, os átomos que estão mais longe da superfície não contribuem significativamente. Isto pode ser demonstrado da seguinte forma (ver Figura 1):

- I. Se somente um átomo está incluído na superfície de corte:

$$\sum_i f_i \xi(d_i) = f_1 \xi(d_1)$$

Equação 5

Como $i=1$, a função é definida como $PLM = f_1$.

- II. Somente quando dois átomos estão dentro do raio de corte sendo iguais à distância dos dois átomos a um ponto da superfície ($d_1=d_2=d$):

$$\sum_i f_i \xi(d_i) = (f_1 + f_2) \xi(d) \quad e \quad N = \sum_i \xi(d_i) = 2 \xi(d)$$

Equação 6

A função de potencial lipofílico molecular é descrita como:

$$PLM = \frac{(f_1 + f_2)}{2}$$

Equação 7

- III. No último caso, dois átomos com distâncias diferentes ($d_1 \neq d_2$) de um ponto da superfície:

$$PLM = \frac{f_1 \xi(d_1) + f_2 \xi(d_2)}{\xi(d_1) + \xi(d_2)}$$

Equação 8

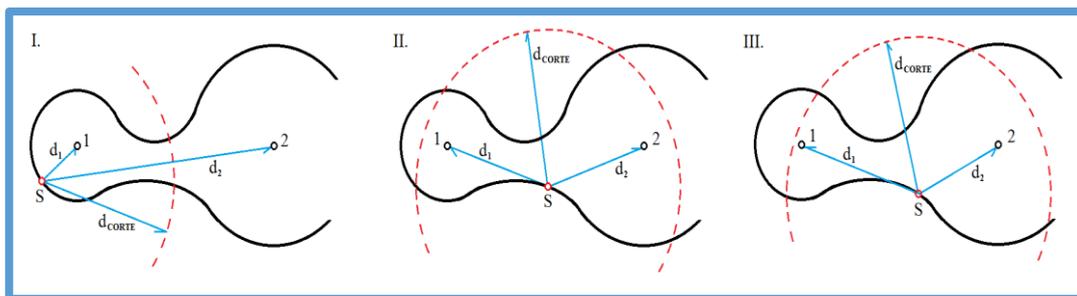


Figura 1: Representação dos átomos que contribuem para a manifestação do potencial lipofílico. Em I, há um átomo centrado no ponto 1 dentro do raio de corte, em II há dois átomos de centros 1 e 2 dentro do raio de corte com a mesma distância (d_1) de um ponto da superfície e, em III dois átomos com diferentes distâncias (d_1 e d_2) de um ponto da superfície, mas dentro do raio de corte.

1.5.1.3 Representação de $\xi(d_i)$

De acordo com Heiden(47) não há uma razão física para a escolha de ξ , esta função deve obedecer as condições citadas na seção anterior em relação à distância d do átomo e d_{corte} . A função apresentada no artigo e também empregada no programa é baseada na função de Fermi (Figura 2):

$$\xi(d) = \frac{1}{e^{a[d-d_{corte}]} + 1}$$

Equação 9

A escolha analítica para esta função reside também no fato de que ela decai rapidamente no intervalo $d_{corte} - \Delta d < d < \Delta d_{corte} + \Delta d$ com $\Delta d = \frac{2}{a}$

A aproximação é flexível e a escolha dos parâmetros a serem considerados foram o d_{corte} igual a 4 e a igual a 2.

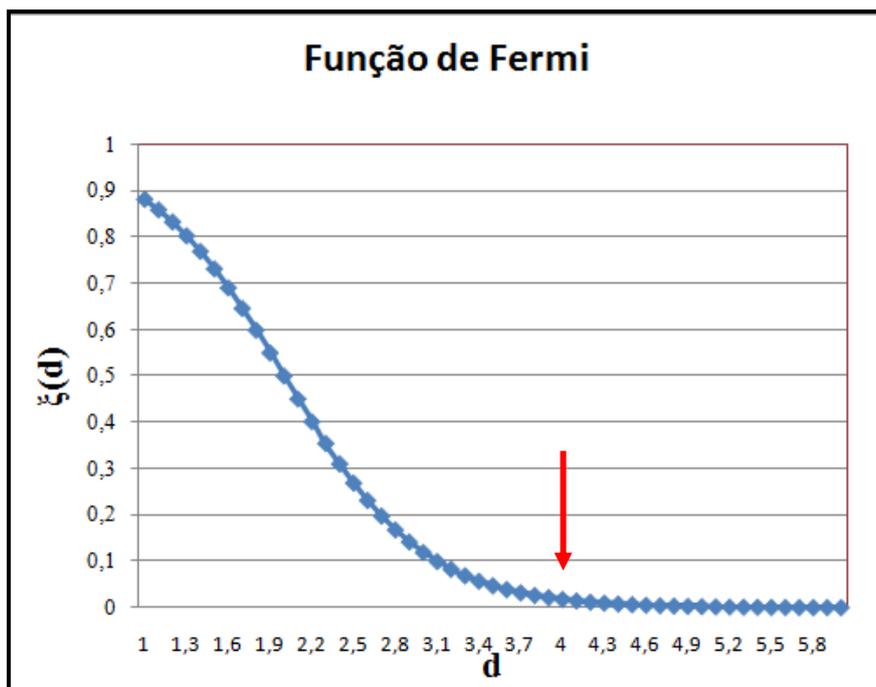


Figura 2: Função de distância ξ baseada na função de Fermi. Os parâmetros utilizados para plotar o gráfico foram: $a=2$ e $d_{corte}=4$. Nota-se que a função é próxima de zero (seta vermelha) quando d é igual a 4.

A função escolhida não tem uma descrição física por si só. De qualquer forma, a aproximação geral é bem flexível, i. e., pode ser considerada como uma razoável escolha dos três parâmetros físicos.

1.6 Potencial de ligação de hidrogênio

Interações entre proteínas como, por exemplo, enzimas e pequenas moléculas dependem da formação e ruptura de ligações de hidrogênio, sendo o meio composto por moléculas de água, o que possibilita a transferência de prótons e/ou elétrons, essencial para as reações químicas(55,57). A energia livre de ligação oriunda do termo relativo à ligação hidrogênio entre o receptor-ligante depende da distância entre o aceptor a doador de hidrogênio, além do ângulo de ligação entre eles.

As ligações de hidrogênio podem ser definidas como uma interação intermolecular entre uma região contendo hidrogênio deficiente de elétrons e uma região de alta densidade de elétrons(58). Esta relação é válida somente

para átomos de hidrogênio ligados a átomos com alta eletronegatividade – são eles: nitrogênio, oxigênio e flúor. A ligação de hidrogênio é direcional não-covalente e fundamental para a estrutura química e reatividade(59). A formação de uma ligação de hidrogênio é baseada no reconhecimento molecular de partes complementares da molécula incluindo grupos aceptores e doadores. Esta interação é mais fraca relativamente a ligações iônicas e covalentes; geralmente as energias de ligação cobrem intervalos entre -0,2 e -40 Kcal.mol⁻¹(58). Devido a estas características as ligações podem ser rompidas em intervalos de flutuações térmicas e processos de reconhecimento molecular podem ocorrer facilmente.

As ligações de hidrogênio são tipos importantes de interações dipolo-dipolo e podem ser consideradas um caso especial de ligação covalente. O grupo composto pelo átomo mais eletronegativo ligado ao hidrogênio é chamado de doador de hidrogênio e o átomo que contribui com o par não compartilhado de elétrons é denominado aceptor de ligação de hidrogênios.

Para a análise do potencial de hidrogênio tendo como referência somente a cavidade da proteína, o descritor foi contabilizado identificando sítios onde há átomos aceptores e doadores de hidrogênio. Para analisar a quantidade e densidade relativa às estruturas aceptoras e doadoras de hidrogênio encontradas em cavidades de proteínas foi utilizado o método apresentado pelo MOLCAD do SYBYL(60).

1.7 Potencial Eletrostático

Processos biológicos englobam uma variedade de processos físicos e químicos, como o transporte de substâncias através de membranas, reações químicas catalisadas por enzimas, e reconhecimentos de ligantes por receptores. A interação eletrostática desempenha importante papel nestes sistemas biológicos.

O modelo teórico da interação eletrostática é relativamente simples. Sabe-se, a priori, fazendo uma aproximação clássica, que a interação entre duas cargas é governada pela lei de Coulomb. Quando temos mais de duas cargas a

energia eletrostática total do sistema se torna um somatório de interações coulombianas. Ao modelar macromoléculas, a aplicação desse princípio se torna mais complexa.

O mapeamento eletrostático realizado neste trabalho é fornecido pelo próprio KVFinder: na ferramenta cada ponto ocupado por um átomo é definido por uma carga específica e estes valores são armazenados em um vetor. Para cada entrada verdadeira da matriz, o algoritmo computa o valor médio das cargas não nula de cada posição vizinha então cada entrada do vetor da superfície recebe o valor médio de cargas recém calculadas(41).

Supondo que a distribuição das cargas dos átomos seja homogênea então a grade irá possuir cargas atômicas parciais nos pontos de fronteiras das cavidades(41).

1.8 MOLCAD e SYBYL-X

Para a validação dos potenciais calculados foi utilizado o módulo MOLCAD pertencente ao programa SYBYL-X(60) que caracteriza propriedades como lipofilicidade, densidade de doadores e aceptores de hidrogênio, potencial eletrostático.

A superfície molecular é gerada, deslizando sobre um conjunto de pontos, uma sonda com raio de 1.4 Å (molécula de água) ao longo do modelo CPK(9,61). Para a busca de canais e cavidades, a molécula é posicionada em uma grade, em que cada ponto é classificado como “in” (dentro) se estiver dentro da superfície molecular e “out” (fora), os demais. Os pontos “out” são selecionados e seus pontos vizinhos que estão a uma distância máxima de 12 Å são rotulados nos sentidos positivo e negativo de cada eixo cartesiano. Se existirem dois eixos com pontos adjacentes rotulados com “in”, nas direções positiva e negativa o ponto central é rotulado como uma cavidade.

Os pontos adjacentes identificados como um ponto de cavidade são clusterizados. Então, são realizadas duas operações: (i) a “contração” que, pelo algoritmo, uma monocamada de pontos nas fronteiras dos clusters é removida. Os clusters que restaram, se possuírem um ou mais vizinhos rotulados como

pontos não pertencentes a uma cavidade, são eliminados, (ii) a segunda operação é a “relaxação”, em que qualquer ponto da grade com um ou mais vizinhos classificados como pontos de uma cavidade são adicionados ao cluster correspondente. Com estas duas operações pequenos clusters são removidos enquanto clusters maiores são ligados por aglomerados menores(62).

O MOLCAD do SYBYL-X possui capacidade de calcular, entre outros, potenciais lipofílicos, densidade de aceptores e doadores de hidrogênio, potencial eletrostático de Poisson-Boltzman com exibição especificada em cores para cada potencial. A validação desse programa foi realizada com substratos específicos de diferentes enzimas P450 e identificação e caracterização do sítio ativo da tripsina(62).

1.9 Drogabilidade

1.9.1 Definição

Os avanços em espectroscopia de ressonância magnética nuclear (RMN) e cristalografia de raios-X aumentaram a habilidade de obter estruturas tridimensionais de proteínas sozinhas ou em complexo com ligantes naturais ou não-naturais. A avaliação destes dados estruturais, especialmente de proteínas complexadas com pequenas moléculas ligantes, permite análises na tentativa de compreender e prever as forças que regem o reconhecimento molecular do ligante. Entender estas forças se torna essencial para a busca de novos fármacos, moléculas que podem ligar-se com alta afinidade e especificidade a proteínas(27).

O processo de reconhecimento de um fármaco por uma macromolécula com alta afinidade e especificidade é conhecido como drogabilidade. A drogabilidade avalia a probabilidade de pequenas moléculas semelhantes a fármacos poderem se ligar a um determinado alvo, com potencial suficiente para alterar sua atividade(63).

O conceito de drogabilidade postula que os sítios de ligação das moléculas biológicas são complementares em termos de volume, topologia e propriedades físico-químicas para seus ligantes. Em trabalhos publicados por Hopkins e

colaboradores(64), introduziu-se o conceito de “*Genoma Drogável*” focando pesquisas de proteínas drogáveis e suas características.

Os recursos necessários para avaliação de drogabilidade baseado em software são, na maioria dos casos, restritos em comparação com técnicas experimentais como espectroscopia por RMN. Métodos experimentais requerem substancial investimento em pesquisas, tempo de purificação de proteínas e coleta de dados. Um esforço que precisa ser considerado em relação ao ganho de conhecimento sobre uma molécula com característica de fármaco e se esta de fato modula a atividade biológica de uma proteína. A espectroscopia por RMN pode ser realizada com um limitado número de proteínas.

Os métodos computacionais que buscam drogabilidade podem apresentar diferentes abordagens. Uma proteína drogável compreende um domínio drogável privilegiado, isto é, um domínio funcional de uma família na qual uma fração significativa dos membros tem sido com sucesso alvo de fármacos(63,64). Outra aproximação é a avaliação da drogabilidade baseada na estrutura, ou seja, a identificação de sítios de ligação na superfície da proteína. O sítio adequado para a ligação de um fármaco possui certas propriedades como tamanho, profundidade da cavidade, forma, distribuição da lipofilicidade por área, número de doadores e aceptores de hidrogênio(30,65).

1.9.2 Métodos e Limitações

Os métodos abordados para avaliar a drogabilidade de proteínas apresentam méritos e limitações. No caso dos métodos computacionais a disponibilidade de programas e os custos devem ser levados em consideração. Para muitas enzimas e receptores pode ser possível ter um primeiro palpite sobre a drogabilidade de sítios ligantes para uma tratabilidade de homólogos. Avaliar os fármacos ligados aos domínios proteicos que pertencem à mesma família de genes pode ser usado para calcular a drogabilidade de um banco de dados estrutural(63).

Métodos baseados em simulação utilizam dinâmica molecular com solvente explícito e requerem um poder computacional muito grande. No entanto, ainda são muito mais rápidos e menos custosos que métodos experimentais.

A limitação do tratamento experimental consiste em produzir um confirmado “*hit*”, ou seja, um ensaio com atividade biologicamente relevante com uma estrutura química tratável, que esteja presente na composição da biblioteca de *screening*(66).

1.9.3 Drogabilidade por métodos experimentais

A forma mais rigorosa de avaliar a drogabilidade é testar experimentalmente a habilidade de compostos semelhantes a fármacos em modular a atividade biológica do alvo de interesse. Considerando isto, examinar a taxa de sucesso do HTS é muito informativo. Uma característica importante do HTS é apontar novos sítios alostéricos em alvos através da seleção de compostos que atuam por mecanismos distintos, ou seja, quando se ligam a cavidades diferentes dos sítios ativos modulando o alvo(27).

Os grandes obstáculos, além dos custos, incluem a ocorrência de falsos positivos, identificados inicialmente como “*hits*” na maioria dos ensaios bioquímicos, e de múltiplos modos de interação ou mecanismos de ação, quando não há a presença de sítios alostéricos.

1.9.4 Drogabilidade por classes de alvos

Investigações sobre alvos drogáveis mostram que eles compartilham características tais como envolvimento em um pequeno número de vias, distribuições em certos órgãos ou localização na superfície celular(67). Alvos drogáveis podem ser considerados um subconjunto do espaço das proteínas drogáveis, mas nem todos os membros da família da proteína drogável se qualificam como um alvo validado.

Para identificar a relação entre alvos drogáveis podem ser utilizados os bancos de dados SCOP(68), que classifica diferentes motivos de proteínas já identificados, e PFAM(69), que possui anotados diferentes domínios proteicos. Esses bancos predizem o número de proteínas que contêm domínios experimentalmente conhecidos para ser alvo de uma pequena molécula com característica de drogas. Aproximadamente 130 “domínios drogáveis

privilegiados” cobrem todos os atuais alvos drogáveis(70), um número que contrasta com o de proteínas e enovelamentos(67,71).

Em trabalho publicado por Hopkins e colaboradores(60) analisou-se a distribuição gene-família de alvos de fármacos, para pequenas moléculas e fármacos biológicos, e descobriu-se que mais de 50% dos alvos fazem parte de quatro famílias-chave: enzimas, GPCR classe I, receptores nucleares, canais de íons dependentes de ligantes e canais de íons dependentes da voltagem. Os alvos com o maior número de fármacos aprovados eram os receptores de glucocorticóides e receptores de Histamina H1. As proteínas quinases são pouco representadas considerando o pequeno número de medicamentos lançados contra esses alvos(70).

O método de suporte de vetores de máquina (SVM) também foi desenvolvido para prever a drogabilidade de proteínas baseado na sequência de aminoácidos independente da sua similaridade(72). No entanto, esse método depende da suposição de que proteínas similares possuem drogabilidades similares, que é comumente aceito em um nível estatístico, mas não necessariamente verdade para alvos particulares de interesse.

A principal limitação destas aproximações vem da seletividade dos membros destas famílias e, paradoxalmente, o desenvolvimento de moduladores alostéricos que são mais seletivos e extraem diversificadas respostas biológicas(73).

Isso demonstra a necessidade de desenvolver métodos de predição de drogabilidade com a finalidade de explorar sítios que não estão ligados à sequência conservada ou ao domínio topológico conservado. Interessantemente, a descoberta de drogas de “primeira da classe” através da triagem fenotípica explora com frequência mecanismos moleculares de ação que diferem da típica inibição competitiva(74).

Embora possivelmente mais difícil de explorar do que sítios ligantes tradicionais, os sítios alostéricos podem ser analisados contra alvos tradicionalmente drogáveis e alvos que não possuem sítios funcionais primários que interagem com pequenas moléculas. A diversidade biológica dos alvos pode apresentar novas oportunidades terapêuticas(75).

1.9.5 Avaliação da drogabilidade baseada na estrutura

A avaliação da drogabilidade baseada na estrutura consiste na identificação do sítio ligante complementar à molécula com característica de fármaco. O primeiro estágio para conduzir esta metodologia inclui a busca por cavidades nos candidatos a alvos. O reconhecimento do sítio de ligação pode se basear somente em considerações geométricas das cavidades (76,77) ou incluir análise de propriedades físico-químicas da superfície da proteína. As estratégias que incorporam propriedades físico-químicas tentam prever interações favoráveis da proteína com o ligante.

Características como cargas, lipofilicidade, regiões doadoras e aceptoras de hidrogênios, potencial eletrostático e aspectos topológicos, como área e volume de cavidades, são importantes para a predição da drogabilidade. A definição de um índice de drogabilidade deriva da ponderação adequada destes parâmetros e comparações quantitativas entre várias proteínas(78).

Recentemente, uma variedade de métodos tem sido reportada para prever ou estimar a drogabilidade de uma proteína. Cheng e colaboradores(79) propuseram um modelo que estima a máxima afinidade de um sítio ligante e uma proteína dada a restrição de que a molécula líder deve possuir propriedades farmacológicas consistentes com a biodisponibilidade oral. O modelo se baseia no princípio de que um fármaco para ser absorvido deve possuir peso molecular menor que 500 Da, área de superfície polar menor que 140 Å² e um valor pequeno de cargas formais uma vez que as moléculas neutras se difundem melhor através das membranas, e por último o alvo deve ter a máxima afinidade de ligação. Com tais propósitos a afinidade pode ser calculada através do entendimento de força molecular presente durante a ligação(79).

A relação entre propriedades físico-químicas e a drogabilidade do alvo foi explorada por Halgreen(80). Com a finalidade de obter informações sobre proteínas que podem ser possíveis alvos terapêuticos, foi desenvolvido pela *Schrödinger Inc* o programa chamado SiteMap®(80). Esta ferramenta computacional encontra cavidades em proteínas e caracteriza regiões

hidrofóbicas, hidrofílicas e o grau de englobamento do sítio. Estas três propriedades estimam quantitativamente a drogabilidade referida pelo *DScore*, uma pontuação que quantifica o grau de drogabilidade. O SiteMap® é uma ferramenta paga, não possui a vantagem de ser utilizada em linha de comando e sua estratégia de análise de drogabilidade ainda é discutida na literatura(63,81).

1.9.5.1 Servidor da Web DogSite

O Dogsite(43) é um servidor da Web que realiza predição automática de sítios de ligação, e analisa e avalia a drogabilidade de uma macromolécula. Na primeira etapa de detecção de cavidade, uma grade é gerada englobando a proteína e os pontos são marcados de acordo com a sobreposição ou não com qualquer átomo. Em sequência, um filtro baseado em diferença de gaussianas (DoG) é aplicado ao grid. Com essa operação, as posições nas superfícies das proteínas são identificadas onde a localização de um objeto como uma esfera é favorável. Com base em um limite de densidade estas posições são agrupadas para potenciais subcavidades. Finalmente, subcavidades vizinhas se fundem formando cavidades(82).

Os descritores de relevância para a detecção de drogabilidade incluem a forma da cavidade. A casca, que envolve o volume, e a “tampa” da cavidade descrevem o quociente denominado invólucro (em inglês, *enclosure*). Quanto maior a taxa entre a “tampa e os pontos que envolvem este volume”, mais exposta ao ambiente é a cavidade. O número de átomos do bolso (C,N,H,O,S e outros) são somados e normalizados assim como o número de regiões hidrofílicas. Os aminoácidos pertencentes às cavidades são analisados por tipo e agrupadas em subconjuntos: apolar e polar(76).

Para a predição de drogabilidade a técnica de aprendizado de máquina supervisionada (em inglês, SML) é incorporada no DogSite. O modelo foi testado para uma versão de banco de dados drogáveis não-redundantes(43).

A desvantagem deste servidor reside na maneira como o usuário irá utilizá-lo. No DogSite faz-se a análise de uma proteína por vez, o que impossibilita qualquer trabalho em paralelo além do estudo de um grande grupo de proteínas simultaneamente, o que demandaria muito tempo para ser realizado. Além disso,

fica-se na dependência da recepção dos resultados via mensagem eletrônica e os possíveis erros decorrentes desse sistema de envio de dados (e.g, erro ou falha de endereçamento).

2 Justificativa

O conhecimento do número de alvos moleculares viáveis apresenta grande importância para intervenção terapêutica. Na Era Pós-Genômica é, para a pesquisa, uma estratégia fundamental no planejamento de fármacos.

A predição de um alvo com potencial terapêutico se torna importante para a busca de compostos bioativos que apresentem afinidade ao alvo em questão. A viabilidade destes alvos para a análise *in silico*, em relação à drogabilidade, está, principalmente, presente na investigação das suas propriedades topológicas e físico-químicas.

Saber as regras que governam a drogabilidade se torna fundamental em programas que tentam produzir novos medicamentos para o mercado. Um novo fármaco, para doenças negligenciadas ou não, demora muitos anos para ser disponibilizado para a população. De acordo com as ferramentas que aqui citamos, a busca *in silico* de novos alvos apresenta uma alternativa, mas com algumas limitações. O tema é bastante discutido e novas regras são apresentadas na literatura para que alcance máxima eficiência.

Neste trabalho utilizamos o KVFinder, um programa com código aberto, que demonstra eficácia na busca por cavidades em proteínas, e introduzimos algoritmos que calculam os potenciais lipofílicos e de densidade de doadores e aceptores de hidrogênio validando as propriedades apresentadas pelos programas com complexos sabidamente drogáveis. Em sequência, combinamos os aspectos físico-químicos e topológicos, realizando uma comparação entre grupos de proteínas drogáveis e não-drogáveis.

Esses primeiros passos, que incluem uma análise estatística simples, irão fornecer dados no futuro, para desenvolvimento de regras robustas que governam a drogabilidade.

3 Objetivos

3.1 Objetivo Geral

Desenvolver uma ferramenta que forneça, através de cálculos de potencial de superfície molecular, uma análise primária da drogabilidade de cavidades de proteínas.

3.2 Objetivos Específicos:

- I. Implementar algoritmos para cálculo de potenciais lipofílico, de densidade de ligação de hidrogênio no programa KVFinder.
- II. Validar os novos algoritmos implementados comparando-os com o programa SYBYL-X.
- III. Investigar a correlação entre parâmetros topológicos e os potenciais calculados das cavidades para proteínas drogáveis e não-drogáveis.

4 Materiais e Métodos

O KVFinder incrementado com os potenciais lipofílico e densidade de ligação de hidrogênio foram compilados em ambiente UNIX testados somente em sistema operacional Linux. A linguagem foi totalmente implementada em ambiente ANSI C.

Para o tratamento de dados e a validação dos potenciais utilizou-se a ferramenta de visualização Pymol, para gerar a superfície definida pelo PLM (Potencial Lipofílico Molecular) e compará-la com a superfície fornecida pelo software SYBYL-X. As proteínas selecionadas para a demonstração e discussão serão mencionadas no final deste capítulo. Nesta fase, o software Pymol foi utilizado para limitar o espaço de busca da cavidade. Para isso, uma caixa foi delimitada na região onde se encontrava o ligante, o espaçamento da grade determinada como 0,2 Å/pts, a sonda interna e a sonda externa eram de valores iguais a 1,4Å e 5,0Å, respectivamente.

O conjunto de proteínas para análise de drogabilidade foi extraído do banco de dados PDB e todo o processo de análise foi realizado através de programação em Shell Script, desde a preparação das proteínas até a análise dos conjuntos proteicos drogáveis e não-drogáveis.

Nas próximas seções serão detalhados os passos realizados para agregar ao KVFinder as rotinas implementadas para os cálculos dos potenciais de superfície.

4.1 Refinamento da superfície molecular do KVFinder

A proteína é inserida em uma grade tridimensional e os átomos são representados pelos seus respectivos raios de van der Waals. Devido a esta representação, a superfície apresenta pequenos espaços vazios, uma vez que, os raios de van der Waals não preenchem totalmente o espaço proteico. O espaço de busca definido pela grade é representado através de uma matriz booleana, onde cada entrada da matriz indica se o espaço representado por ela está vazio ou não.

Para definir a superfície de acesso ao solvente uma sonda de raio definida pelo usuário (neste caso 1,4 Å) percorre todos os pontos da grade. Atribui-se o valor 0 (Falso) aos pontos em que os átomos estão presentes e o valor 1 (Verdadeiro) onde não existem átomos, ou seja, onde não há a intersecção entre os raios dos átomos e da sonda. Após realizar o processo para todos os átomos contidos no espaço de busca do algoritmo, se obtém uma matriz que representa os espaços vazios e não vazios.

Para obter a superfície molecular é definido um ponto de fronteira com uma entrada de valor 1 e que possui no mínimo um ponto adjacente com valor 0. Quando o ponto de fronteira é identificado, centraliza-se nele uma esfera limítrofe de raio 1,4 Å, a rotina atribui o valor 1 a todas as entradas que representam os pontos contidos na esfera.

O mesmo processo é repetido com uma sonda maior gerando outra superfície de pontos. A sonda de 5,0 Å restringe o acesso ao espaço vazio dentro da proteína, gerando uma superfície usada para definir a fronteira entre a cavidade e o meio externo.

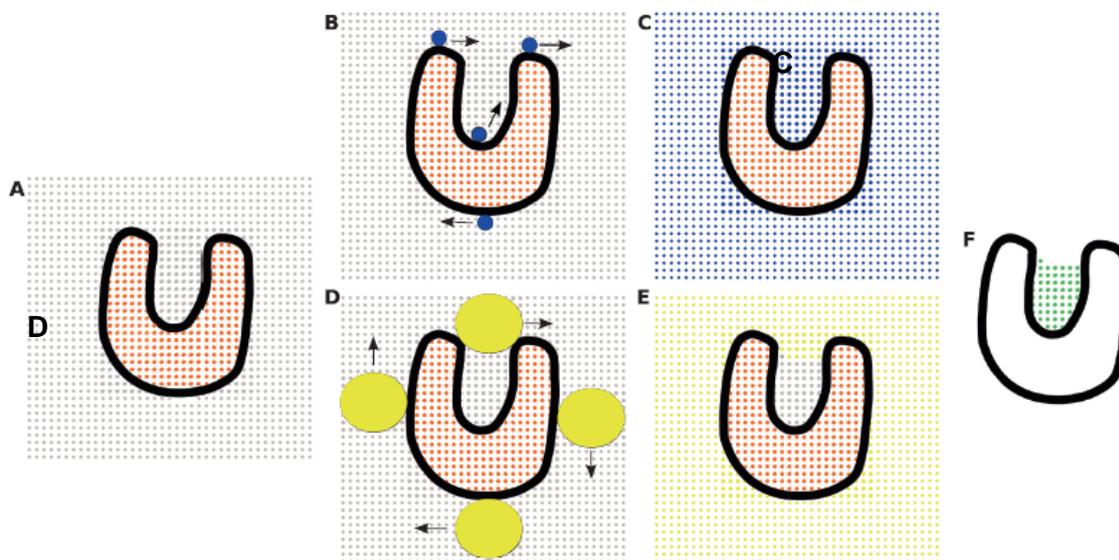


Figura 3: Busca de uma cavidade realizada pelo KVFinder. Figura extraída do artigo que apresenta a ferramenta(37). Em (A), temos o espaço, com pontos vermelhos, onde se encontra os átomos da proteína e onde não há átomos da proteína, círculos em azul. Em (B), a sonda de raio 1,4 Å (azul) percorre a superfície da proteína. Em (C), todos os pontos que foram sobrepostos pela sonda são marcados em azul. Em (D), uma sonda de raio 5,0 Å (amarelo) desliza pela superfície limitando a região da cavidade. Em (E), temos os pontos da grade onde, em vermelho, se encontra a proteína, em rosa claro, o espaço vazio e, em verde os pontos da cavidade. Por fim, em (F) temos somente demonstrado os pontos da cavidade.

Para obter somente a superfície molecular que tangencia a proteína, foram adicionadas novas rotinas à ferramenta. Durante o preenchimento dos espaços vazios e não-vazios, um terceiro vetor foi implementado para identificar o espaço externo à proteína e à cavidade. Foram atribuídos a estes pontos valor igual a -1. Cada ponto que representa a superfície da proteína é analisado quanto aos seus primeiros e segundos vizinhos. O ponto da cavidade que apresenta mais que dois pontos adjacentes com valor igual -1 é descartado de forma que a cavidade apresente somente pontos que tangenciam a proteína.

4.2 Implementação dos Potenciais de Superfície

4.2.1 Potencial Lipofílico

4.2.1.1 Entrada de Dados e atribuição do coeficiente logarítmico de partição (LogP)

O KVFinder possui três arquivos de entrada: (i) um arquivo contendo os parâmetros do programa, que são determinados pelo usuário, (ii) um arquivo PDB contendo informações sobre a estrutura da proteína alvo, (iii) arquivo de dicionário que possui o raio de van der Waals de cada átomo e cargas parciais. Para o cálculo do potencial lipofílico o arquivo de entrada ParamLogP foi construído com os valores de logP dos átomos de cada aminoácido. O arquivo ParamLogP é composto por três colunas: i) a primeira contém os símbolos de cada átomo distribuídos de acordo com cada resíduo; ii) a segunda coluna possui os valores re-parametrizados por Crippen(46) e; ii) a terceira coluna contém os valores de LogP parametrizados por Ghose e Viswanadhan(56). O usuário pode escolher quais parâmetros utilizar.

A interação de qualquer átomo da molécula com o solvente é uma função de várias forças moleculares, assim como a acessibilidade ao solvente. A classificação dos átomos é realizada, por conseguinte, para diferenciar parcialmente: (i) a distribuição eletrônica ao redor do átomo, (ii) a acessibilidade do solvente para o átomo. A natureza dos átomos ligados ao átomo em questão, bem como a seus vizinhos mais próximos influenciam estes fatores. Os valores de LogP são armazenados em uma tabela e conferidos a cada tipo de átomo do

resíduo. Na rotina de extração das coordenadas atômicas, os valores de raio, carga e logP, armazenado inicialmente pelos arquivos de entrada ParamLog, são conferidos a cada átomo da proteína. Para a implementação do potencial lipofílico foi atribuído um vetor ponteiro que armazena: (i) o símbolo de cada tipo de átomo da proteína (C, N, O, S, H) contido no espaço de busca, (ii) o raio, (iii) posição x, y, z indicada pelo arquivo PDB, (iv) valor de LogP.

A Figura 4 indica os parâmetros atômicos físico-químicos parametrizados por Crippen e colaboradores(46). A Figura 5 é um esquema que une os algoritmos do KVFinder associado às rotinas implementadas para o cálculo do Potencial Lipofílico Molecular.

type	descriptions	SMARTS ^a	log P
C1	1°, 2° aliphatic	'[CH4]', '[CH3]C', '[CH2](C)C'	0.1441
C2	3°, 4° aliphatic	'[CH](C)(C)C', 'C(C)(C)C'	0.0000
C3 ^b	1°, 2° heteroatom	'[CH3][(N,O,P,S,F,Cl,Br,I)]'	-0.2035
C4 ^c	3°, 4° heteroatom	'[CH2X4](N,O,P,S,F,Cl,Br,I)' '[CH1X4][(N,O,P,S,F,Cl,Br,I)]' '[CH0X4][(N,O,P,S,F,Cl,Br,I)]'	-0.2051
C5	C = heteroatom	'[C] = [A#X]'	-0.2783
C6	C = C aliphatic	'[CH2] = C', '[CH1](=C)A', '[CH0](=C)(A)A', '[C](=C)=C'	0.1551
C7	acetylene, nitrile	'[CX2]#A'	0.00170
C8	1° aromatic carbon	'[CH3]c'	0.08452
C9	1° aromatic heteroatom	'[CH3][a#X]'	-0.1444
C10	2° aromatic	'[CH2X4]a'	-0.0516
C11	3° aromatic	'[CHX4]a'	0.1193
C12	4° aromatic	'[CH0X4]a'	-0.0967
C13 ^d	aromatic heteroatom	'[cH0]-[(C,N,O,S,F,Cl,Br,I)]'	-0.5443
C14	aromatic halide	'[c][#9]'	0.0000
C15	aromatic halide	'[c][#17]'	0.2450
C16	aromatic halide	'[c][#35]'	0.1980
C17	aromatic halide	'[c][#53]'	0.0000
C18	aromatic	'[cH]'	0.1581
C19	aromatic bridgehead	'[c](a)(a):a'	0.2955
C20	4° aromatic	'[c](a)(a)-a'	0.2713
C21	4° aromatic	'[c](a)(a)-C'	0.1360
C22	4° aromatic	'[c](a)(a)-N'	0.4619
C23	4° aromatic	'[c](a)(a)-O'	0.5437
C24	4° aromatic	'[c](a)(a)-S'	0.1893
C25	4° aromatic	'[c](a)(a) = C', '[c](a)(a) = N', '[c](a)(a) = O'	-0.8186
C26	C = C aromatic	'[C](=C)(a)A', '[C](=C)(c)a', '[CH](=C)a', '[C] = c'	0.2640
C27 ^e	aliphatic heteroatom	'[CX4]![(C,N,O,P,S,F,Cl,Br,I)]'	0.2148
CS	carbon supplemental	'[#6] not matching any basic C type	0.08129

Figura 4: Fragmento da tabela contida no artigo de Crippen e colaboradores(46) em que foi realizada a reparametrização dos valores de logP. Nesta tabela há os valores de LogP para o carbono.

4.2.1.2 Cálculo do PLM para cada Cavidade

As rotinas do KVFinder são executadas até as cavidades serem encontradas e os pontos que tangenciam a superfície da proteína são exportados para o arquivo de saída de extensão PQR. As cavidades geradas pelo KVFinder são exportadas para um novo arquivo de saída apresentando as posições x , y , z de cada ponto da cavidade e a posição i , j e k relativo à grade. As posições i , j e k serão necessárias para o cálculo da área referente à lipofilicidade, o que será discutido mais adiante.

Ao fim da busca pela cavidade foram inseridas as rotinas que calculam o potencial lipofílico para cada ponto da cavidade e o potencial lipofílico total da mesma.

O algoritmo define um raio de corte que seja relevante para a influência da lipofilicidade em cada ponto da superfície. O raio de corte indicado pela literatura é de 4Å(47). Nesta etapa não foi utilizada o artifício da grade.

Para cada ponto da superfície, o algoritmo realiza uma busca pelos átomos que possuem uma distância espacial menor ou igual a 4Å. As distâncias destes átomos são então guardadas em um vetor assim como o valor do LogP. Ao final da busca e com as informações armazenadas, a cada ponto da superfície é aplicado a função de distância, equação 3, seguida do fator de normalização. Os valores de lipofilicidade encontrados, sendo menor que zero, serão considerados pontos com características hidrofílicas, e aqueles que possuem valores positivos serão considerados pontos com características hidrofóbicas. Essas duas grandezas são somadas para saber o quão hidrofóbico ou hidrofílico uma cavidade pode se apresentar. Este cálculo é realizado para todas as cavidades encontradas pelo KVFinder, e os valores dos potenciais impressos em um arquivo com extensão PQR, que possui as coordenadas x , y , z , de cada ponto e o valor do potencial lipofílico para cada ponto das cavidades.

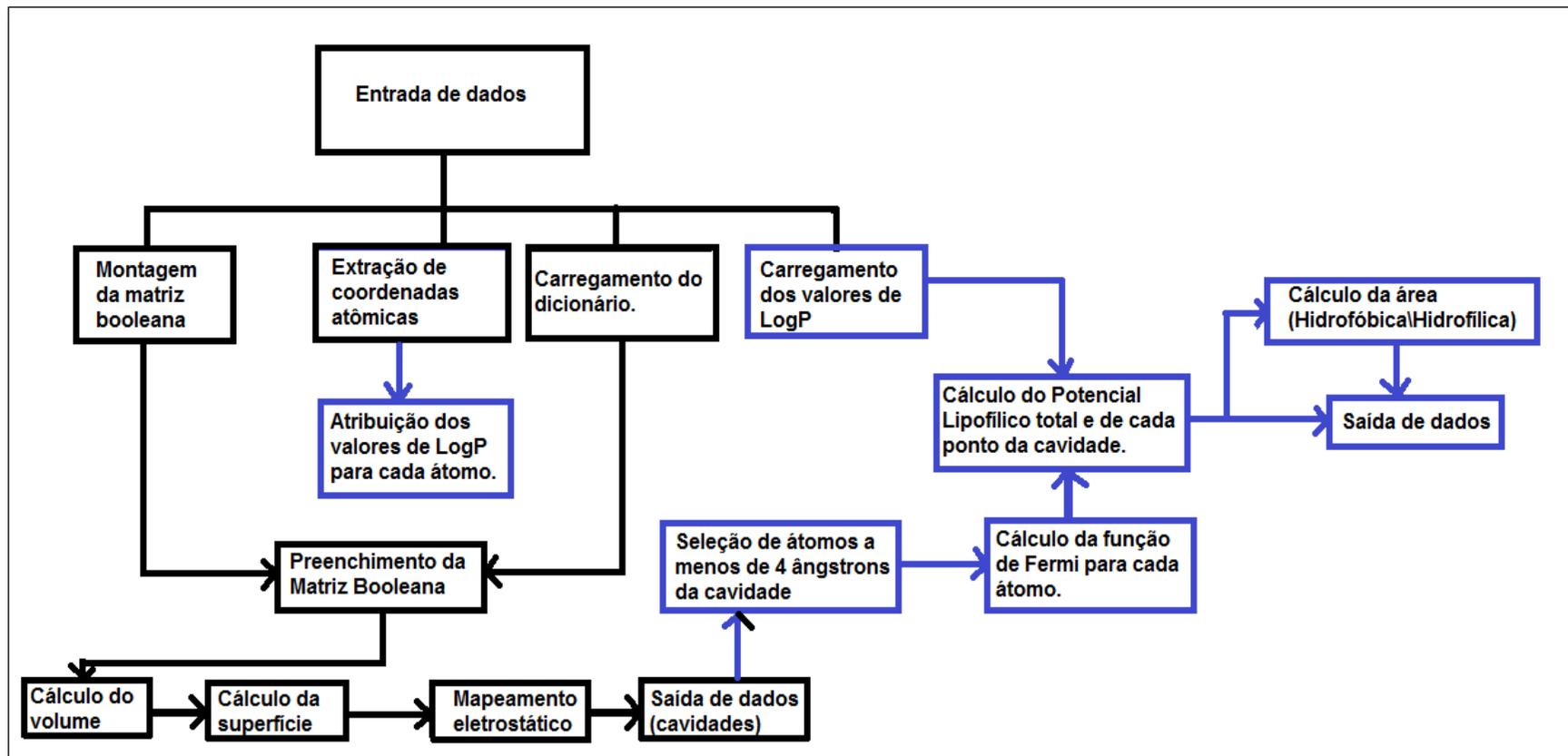


Figura 5: Algoritmo de busca de cavidades do KVFinder com o algoritmo de cálculo do potencial lipofílico(em azul).

O Potencial Lipofílico total para uma cavidade é alcançado somando todos os valores de potenciais encontrados, positivos e negativos, de cada ponto da superfície. Nesta etapa, um segundo arquivo é gerado, identificando a cavidade e o valor total de hidrofobicidade e hidrofilicidade.

4.2.1.3 Determinação do valor da área hidrofóbica e hidrofílica de cada cavidade

O arquivo PQR que contém as coordenadas espaciais x , y , z de cada ponto da cavidade foi modificado para incluir o valor correspondente às posições da grade (i,j,k) e respectivos valores de lipofilicidade. Os pontos com características hidrofóbicas recebem valor +1 e pontos com características hidrofílicas recebem valor -1.

As informações sobre os pontos da cavidade eram enviadas para o cálculo da área referente às regiões hidrofóbicas e hidrofílicas. Para obter o valor da área da superfície considera-se quadrados de lado h e área $h^2(41)$, sendo que a área para cada uma das duas características é calculada multiplicando o número de pontos de cada conjunto por h^2 .

As informações são exportadas para um arquivo de saída de extensão TXT, indicando a área com propriedades hidrofóbicas e hidrofílicas apresentadas por cada cavidade.

4.2.2 Densidade de Aceptor e Doador de Hidrogênio

4.2.2.1 Entrada de dados e identificação de átomos aceptores e doadores de Hidrogênio

Para o cálculo da densidade de ligação de hidrogênio uma tabela que indica os átomos que fazem ligação de hidrogênio foi adicionada para identificação dos átomos nitrogênio (N), oxigênio (O). O arquivo de entrada com a tabela denominada "hbond_tab" possui uma coluna com os átomos que realizam ligações de hidrogênio e uma coluna com o máximo de ligações que ele pode realizar.

Átomos	Ligações
N	3
O	2

Tabela 2: Número máximo de ligações de hidrogênio por átomo aceptor.

Os valores do arquivo são importados e armazenados em uma tabela. Para assinalar a cada hidrogênio seu respectivo átomo ligante o algoritmo se baseou nas posições que os átomos se encontram no arquivo PDB.

ATOM	1	N	MET	A	1	12.776	8.296	37.659	1.00	51.22	N
ATOM	2	CA	MET	A	1	13.289	9.437	38.274	1.00	50.25	C
ATOM	3	C	MET	A	1	12.309	10.123	39.121	1.00	46.70	C
ATOM	4	O	MET	A	1	11.190	9.650	39.262	1.00	47.07	O
ATOM	5	CB	MET	A	1	13.605	10.793	37.021	1.00	54.09	C
ATOM	6	CG	MET	A	1	12.880	11.029	35.426	1.00	57.61	C
ATOM	7	SD	MET	A	1	12.445	12.767	34.707	1.00	61.46	S
ATOM	8	CE	MET	A	1	13.971	13.309	33.819	1.00	61.18	C
ATOM	9	HA	MET	A	1	14.039	9.143	38.867	1.00	12.00	H
ATOM	10	HG2	MET	A	1	13.520	10.600	34.788	1.00	12.00	H
ATOM	11	HG3	MET	A	1	12.031	10.502	35.450	1.00	12.00	H
ATOM	12	HE1	MET	A	1	14.178	12.636	33.108	1.00	12.00	H
ATOM	13	HE2	MET	A	1	13.855	14.210	33.402	1.00	12.00	H
ATOM	14	HE3	MET	A	1	14.724	13.340	34.477	1.00	12.00	H
ATOM	15	N	VAL	A	2	12.488	11.430	39.074	1.00	41.83	N
ATOM	16	CA	VAL	A	2	11.533	12.403	39.573	1.00	36.64	C
ATOM	17	C	VAL	A	2	11.669	13.701	38.764	1.00	31.31	C
ATOM	18	O	VAL	A	2	12.762	14.020	38.319	1.00	30.86	O
ATOM	19	CB	VAL	A	2	11.814	12.649	41.059	1.00	38.45	C
ATOM	20	CG1	VAL	A	2	13.263	13.116	41.284	1.00	41.32	C
ATOM	21	CG2	VAL	A	2	10.797	13.614	41.689	1.00	38.59	C
ATOM	22	H	VAL	A	2	13.336	11.772	38.669	1.00	12.00	H
ATOM	23	HA	VAL	A	2	10.612	12.042	39.427	1.00	12.00	H
ATOM	24	HB	VAL	A	2	11.705	11.776	41.535	1.00	12.00	H
ATOM	25	HG11	VAL	A	2	13.896	12.415	40.954	1.00	12.00	H
ATOM	26	HG12	VAL	A	2	13.421	13.969	40.788	1.00	12.00	H
ATOM	27	HG13	VAL	A	2	13.412	13.266	42.262	1.00	12.00	H
ATOM	28	HG21	VAL	A	2	9.879	13.228	41.601	1.00	12.00	H
ATOM	29	HG22	VAL	A	2	11.009	13.752	42.656	1.00	12.00	H
ATOM	30	HG23	VAL	A	2	10.837	14.491	41.210	1.00	12.00	H

Figura 6: Fragmento de um arquivo PDB contendo coordenadas dos átomos da metionina (MET) e valina (VAL). A demonstração tem por objetivo exemplificar o algoritmo empregado para o cálculo da densidade de hidrogênio, mostrando como cada hidrogênio é ligado a cada átomo N, O, C.

Na Figura 6, há um fragmento de um arquivo PDB com as denominações dos resíduos metionina (MET) e valina (VAL) e os átomos que compõem cada aminoácido, além das demais informações como coordenadas x, y, z, resíduo e o tipo de átomo. Cada resíduo se inicia com o nitrogênio (N) componente do grupamento amino.

Os hidrogênios ligados ao nitrogênio e oxigênio são armazenados em um vetor. O número de hidrogênios ligados a cada N e O são contabilizados e comparados com a tabela 2. Se o número de ligação é menor, ou concorda com a tabela os átomos de N e O recebem uma indicação de aceptores de hidrogênio.

Restrições:

- (i) Se o número de ligações exceder o valor da tabela e a carga é positiva, os átomos são retirados dos grupos de receptores de H.
- (ii) A histidina possui um nitrogênio com ligação dupla em uma cadeia e apresenta ressonância. Este nitrogênio é marcado como já possuidor três ligações.

4.2.2.2 Algoritmo implementado para o cálculo da densidade deceptor e doador de hidrogênio

Assim como no algoritmo de cálculo do PLM, somente os átomos que possuem uma distância menor que 4Å do ponto da superfície irão contribuir para o valor de densidade de hidrogênio.

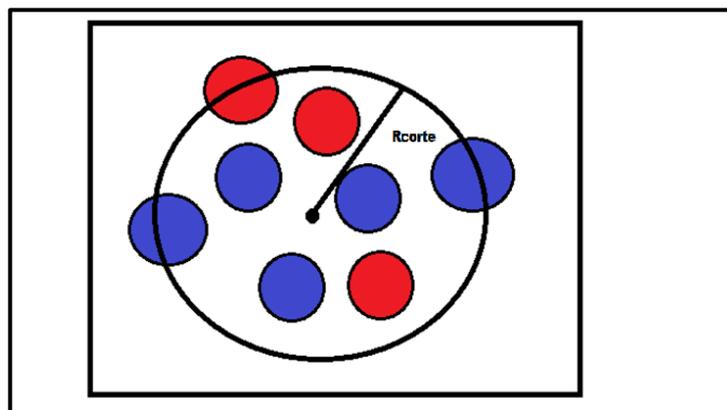


Figura 7: Raio de corte em relação a um ponto da superfície. A esfera originada contém aceptores e doadores de hidrogênio.

De acordo com a Figura 7 os átomos representados de cores azul e vermelho que estão parcialmente dentro do raio de corte também irão contribuir para o cálculo da densidade de aceptores e doadores de H. Para cada ponto na superfície da cavidade é calculada a porcentagem do volume do átomo que está

compreendida dentro da esfera. O cálculo da densidade de doadores (ρ_{don}) e aceptores de hidrogênio (ρ_{acep}) é dado pela seguinte equação:

$$\rho_{acep/don} = \frac{\sum_i n_{acep/don}}{V_{total}}$$

Equação 10

sendo n a fração de volume de cada átomo doador e acceptor compreendido dentro do volume total. Como o raio de corte (R_{corte}) e o raio (r) dos átomos são da mesma ordem de grandeza, $R_{corte} \approx r$, é necessário usar artifícios matemáticos para o cálculo da densidade. Se um átomo está totalmente dentro do raio de corte da esfera consideramos n igual a 1, se somente parte do átomo está contido no raio de corte da esfera, utilizamos os cálculos descritos a seguir.

Detalhes da Implementação

Para ilustrar o cálculo da contribuição parcial de átomos para o cálculo da densidade de hidrogênio segue a Figura 8.

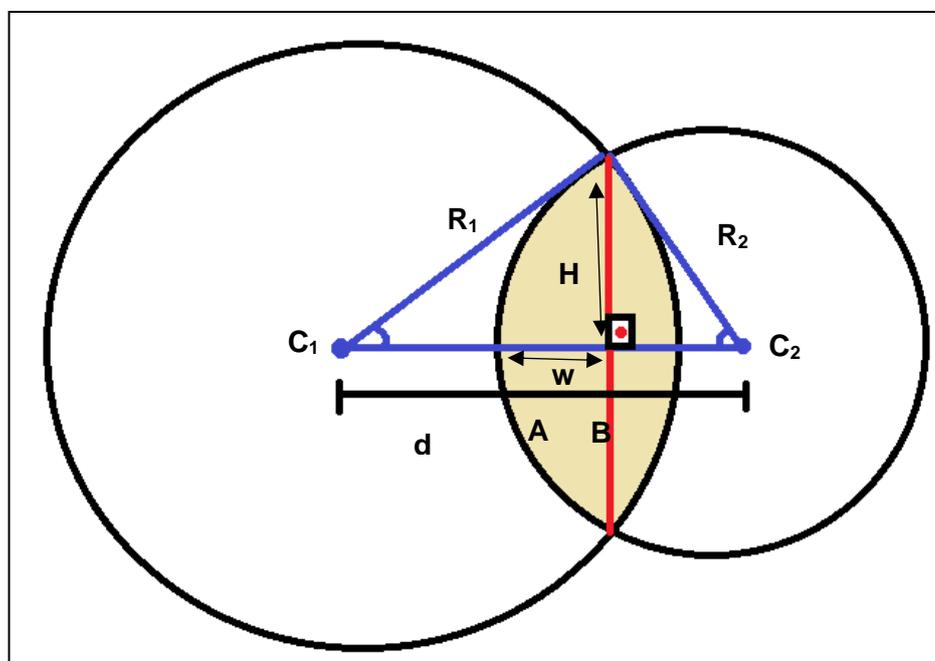


Figura 8: Representação geométrica da contribuição de cada átomo no cálculo da densidade de hidrogênio em um ponto da superfície, C_1 . O raio do átomo é representado por C_2 . A distância entre os dois centros é determinada por d . R_1 e R_2 representam o raio de corte e o raio do átomo, respectivamente. A e B é a região do átomo contida dentro da esfera com raio de corte de 4\AA , w é a altura da cunha esférica A gerada pela intersecção entre as duas esferas e H é o raio desta cunha esférica.

Na Figura 8, C_1 representa um ponto da superfície, R_1 é o raio de corte, R_2 é o raio do átomo de centro C_2 , e d a distância entre os centros das duas esferas. O volume do átomo contido dentro da esfera maior é dado pela região colorida em amarelo claro. Para calcular este volume o algoritmo se baseia em operações geométricas. A região colorida é uma intersecção composta por duas calotas esféricas, o que permite construir um triângulo com lados R_1 , R_2 , d e altura H . Usando o Teorema de Carnot ou também conhecido como lei dos cossenos é possível encontrar o valor de θ :

$$\cos \theta = \frac{(R_1)^2 + d^2 - (R_2)^2}{2R_1d}$$

Equação 11

Identificando-se o $\cos \theta$ é possível saber o valor de H , uma vez que, podemos usar a identidade $\sin^2 \theta + \cos^2 \theta = 1$.

$$H = R_1 \sin \theta$$

Equação 12

Sendo H o raio da calota esférica e w a distância entre a o centro da calota esférica e extremidade do átomo. Desta forma, conclui-se que o volume da calota:

$$V_C = \frac{1}{6} \pi w^2 (3H^2 + w^2)$$

Equação 13

Da mesma forma podemos encontrar o volume da segunda calota C_1 . Somando os dois volumes A e B e dividindo o resultado pelo volume total do átomo obtemos o valor de $n_{acep/donor}$. O cálculo é realizado para cada átomo acceptor e doador a 4Å de distância da superfície. A densidade é calculada somando os valores de n para aceptores e doadores e dividindo o resultado pelo volume da esfera com raio de 4Å.

Outra abordagem realizada consistiu em encontrar o número absoluto de aceptores e doadores de hidrogênio contidos próximos a cavidade.

As quantidades referentes à densidade e ao número de átomos aceptores e doadores para cada cavidade são exportados para um arquivo de saída. Os

valores de densidade de cada ponto da cavidade são exportados em dois arquivos com extensão PQR.

4.2.2.3 Determinação do valor da densidade de aceptores e doadores de hidrogênio de cada cavidade

A área referente à densidade de doadores de aceptores de ligação de hidrogênio é calculada da mesma forma que se calculam as áreas hidrofóbica e hidrofílica. Neste caso, cada ponto da superfície da cavidade pode apresentar ambas as características de densidade, por isso a necessidade de dois vetores que armazenam informações sobre doadores e aceptores e calculam a área relativa a cada propriedade.

4.3 Preparação para análise de drogabilidade

A seleção do conjunto de alvos drogáveis foi realizada utilizando o banco de dados *Protein Data Bank* (PDB). O servidor possui uma tabela que contém informações sobre os fármacos e seus respectivos alvos e o endereçamento ao banco de dados *Drugbank*. O banco de dados *Drugbank* (<http://www.drugbank.ca/>) é um recurso que combina informações detalhadas farmacológicas e químicas com informações sobre alvos drogáveis. Este banco de dados possui um número considerável de alvos: 4282 proteínas, com sequências não-redundantes, para fármacos aprovados disponíveis em estruturas cristalográficas.

A tabela que mapeia os fármacos e alvos das estruturas cristalográficas compreende: (i) nome genérico do fármaco e o nome patenteado, (ii) o identificador da molécula no *Drugbank*, (iii) a identidade do ligante e o nome do alvo, (iv) as três melhores estruturas que contêm o fármaco ou seu estereoisômero e, (v) o grau de similaridade dos três melhores alvos encontrados em relação à sequência de aminoácidos do alvo e, por fim, (iv) entrada no PDB.

Drugs Bound to Primary Target Table

Drug Bounds to Primary Targets Primary Drug Targets

The top 3 PDB IDs for each drug based on the **drug target sequence similarity search** and **ligand id search**. If the sequence identities are same, the PDB ID with better

Generic Name	Brand Name	DrugBank ID	ATC Codes	Ligand ID	Target Name	UniProt ID	PDB ID 1	Seq. Identity 1	PDB ID 2	Seq. Identity 2	PDB ID	Seq. Id	All PDB IDs
Acarbose	• Glucobay • Prandase • Precose	DB00284	• A10BF01	ACR, QPS	Maltase-glucoamylase, intestinal	O43451	2QMJ	99%	3TOP	97%			Search
Acarbose	• Glucobay • Prandase • Precose	DB00284	• A10BF01	ACR, QPS	Lysosomal alpha-glucosidase	P10253	2QMJ	44%	3TOP	39%			Search
Acarbose	• Glucobay • Prandase • Precose	DB00284	• A10BF01	ACR, QPS	Sucrase-isomaltase, intestinal	P14410	3TOP	60%	2QMJ	58%			Search
Acarbose	• Glucobay • Prandase • Precose	DB00284	• A10BF01	ACR, QPS	Pancreatic alpha-amylase	P04746	1KXH	45%					Search

Filter Reload Page 1 of 1 20

Figura 9: Complexo proteína-fármaco do servidor PDB contendo o nome genérico do fármaco e o nome comercial, além do link para os bancos de dados *DrugBank* (para o fármaco) e *Uniprot* (para o alvo). As informações incluem o nome do alvo e a identidade no PDB.

Os alvos drogáveis escolhidos foram aqueles que possuem o maior grau de similaridade com as estruturas primárias dos alvos. O objetivo foi escolher a amostra mais diversificada possível e que apresentasse maior grau de similaridade com alvos primários. Dessa forma, o valor de corte de similaridade obtido foi de 37%, para que o conjunto apresentasse maior número de alvos, correspondendo nesse caso a 199 proteínas.

A seleção para o grupo não-drogável também foi obtida pelo PDB. Neste caso não há a certeza da drogabilidade da estrutura e para contornar este problema, foi realizada uma busca por artigos (63,81,82) que abordam drogabilidade. A classificação como alvos não drogáveis foi baseada nos resultados computacionais publicados, selecionando proteínas que apresentavam baixos índices de drogabilidade. Os alvos escolhidos foram das famílias NPRT, Endoglucanase_c, TPK, Caspase_1, Malato_desidrogenase e PISPLP.

4.4 Preparação das proteínas e geração de dados

As proteínas extraídas do banco de dados PDB foram, através de um *shell script*, isoladas do ligante e protonadas usando o programa PROPKA automatizado pelo pdb2pqr(83), que é uma ferramenta empregada na preparação de proteínas, como exemplo, otimizando as macromoléculas para

ligações de hidrogênio favoráveis. O programa retorna um arquivo com as mesmas características de um arquivo PDB. Os parâmetros definidos para esta etapa incluem a determinação do pH padronizado para todas as proteínas como 7,4 e o Amber como campo de forças(84).

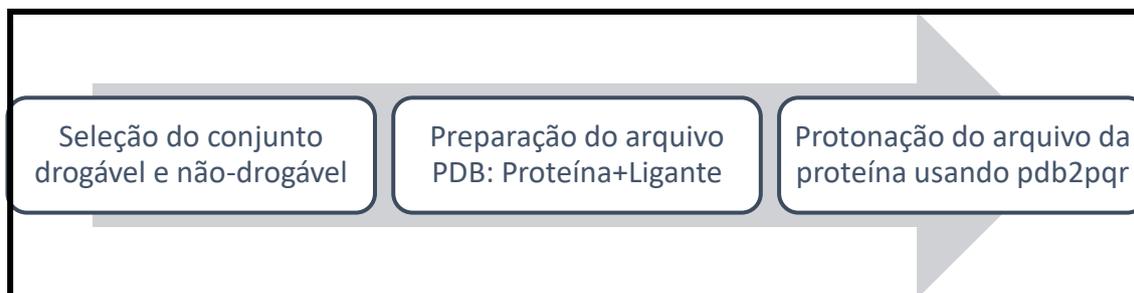


Figura 10: Etapas da preparação das proteínas.

Depois de preparados, os arquivos foram submetidos ao KVFinder para a busca das cavidades e cálculos dos potenciais. Nesta segunda etapa, o programa Pymol foi utilizado para obter o arquivo de entrada padrão (*input*) para todas as proteínas. Todos os ligantes presentes no arquivo pdb com a proteína foram selecionados para a análise.

Após a preparação da proteína com o arquivo de entrada padrão do KVFinder, e os arquivos de entrada: "ParamLogP.dat" e "hbond_tab.dat", todas as proteínas foram submetidas ao programa, de forma automatizada, utilizando um computador Intel Core, com 8GB de memória e sistema operacional Scientific Linux 6.5.

O principal arquivo de saída do KVFinder com as informações topológicas das cavidades consiste de um arquivo de texto de extensão PQR, que possui as coordenadas de cada ponto da cavidade. Os arquivos cav_lipo, cav_acep e cav_don apresentam informações sobre a lipofilicidade, densidade de aceptores e doadores de hidrogênio de cada cavidade, respectivamente. Foram gerados dois arquivos de texto com informações sobre lipofilicidade total, densidade total, densidade de aceptores e doadores de ligação de hidrogênio e o número absoluto de aceptores e doadores para cada cavidade individualmente.

As informações sobre todas as cavidades estão presentes em cada arquivo citado, ou seja, um arquivo cav_lipo.pqr possui informações espaciais (x,y,z), o

nome da cavidade, e o valor lipofílico de cada ponto da cavidade, o arquivo cav_hb.pqr possui as mesmas informações acima mas ao invés do valor de lipofilicidade encontra-se o valor de densidade de ligação de hidrogênio.

Distinguir a posição na proteína e encontrar em qual cavidade está localizado um ligante para um conjunto de mais de 199 proteínas geraria um trabalho exaustivo se analisadas visualmente em programas como Pymol e VMD. O próximo tópico descreve como foi realizado o refinamento da busca pelos bolsos proteicos de interesse.

4.5 Seleção das cavidades relevantes para análise dos dados

Para encontrar a cavidade onde está presente o fármaco nos baseamos no seguinte método: a distância entre as médias geométricas da posição de cada átomo do fármaco e dos pontos que compõem a cavidade onde se encontra o ligante, deve possuir o menor valor. Para encontrar estas cavidades um programa foi elaborado obedecendo o seguinte algoritmo:

- 1) As coordenadas x, y e z de cada cavidade e do ligante eram lidas a partir do arquivo inicial gerado pelo KVFinder, onde cavidade é uma superfície fechada, e do arquivo PDB.

- 2) É calculada a média geométrica espacial de cada cavidade e do ligante.

- 3) A cavidade selecionada será aquela que apresentar menor distância entre sua média geométrica e do ligante.

5 Resultados e Discussão

5.1 Representação da superfície molecular das cavidades proteicas que tangenciam a superfície da proteína.

O KVFinder possui o diferencial de caracterizar todos os tipos de cavidades dentro de uma proteína, incluindo fendas, túneis, depressões e demais elementos topológicos vazios dentro de uma proteína. A cavidade encontrada mostra uma estrutura bem definida. No entanto, assim como outros programas de busca de cavidades, os bolsos são demonstrados como um contorno fechado. A ferramenta mostra a cavidade oca e completamente fechada ([Figura 11A](#)).

O primeiro passo foi retirar a região que não tangencia a cavidade, que também podemos designar como uma tampa, para que se torne possível a caracterização dos potenciais de superfície ([Figura 11B](#)). A [Figura 11](#) e a [Figura 12](#) mostram as proteínas cruzaina e HIV-1 Protease, respectivamente, com cavidades totalmente fechadas e somente com os pontos tangentes à superfície molecular.

Incluindo estas alterações que foram descritas no capítulo de materiais e métodos, uma nova área foi calculada utilizando o mesmo princípio empregado pelo KVFinder.

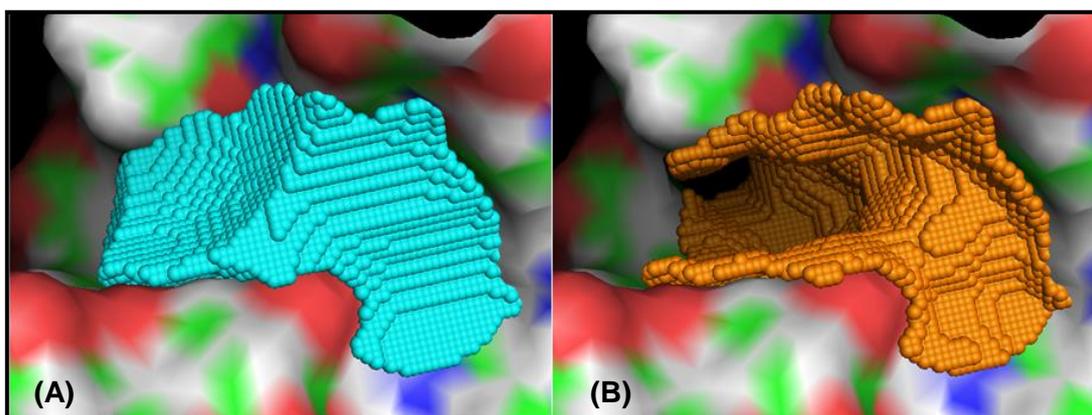


Figura 11: Cavidades geradas pelo KVFinder. O volume está totalmente coberto em A e foi gerado pelo arquivo de saída original do KVFinder, em B tem-se a mesma cavidade, no entanto, somente os pontos que tangenciam a superfície molecular é representado, conforme modificações introduzidas no programa.

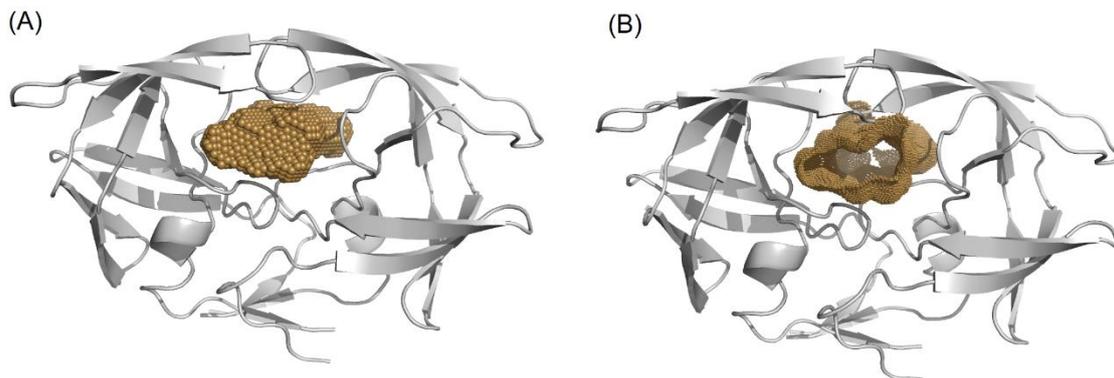


Figura 12: Proteína HIV-Protease-1. Em A, a cavidade encontrada pelo KVFinder totalmente fechada. Em B, a mesma proteína e cavidade, no entanto, somente os pontos que tangenciam a superfície molecular estão representados.

5.2 Validação dos algoritmos de cálculos de potenciais de superfície molecular

A etapa de validação foi realizada com o KVFinder atuando nos complexos: 4MS4 (receptor de GABA_B ligado ao agonista baclofen)⁽⁸⁵⁾, 1HPV (HIV-protease-1 e VX-478)⁽⁸⁶⁾, 2QMJ (Glucoamilase-maltase humana e acarbose)⁽⁸⁷⁾. O SYBYL-X foi a ferramenta utilizada para comparar com os resultados gerados pelo KVFinder.

Os potenciais das cavidades foram comparados com os potenciais do ligante e discutidos em relação à complementaridade de cada complexo, indicando afinidade entre grupos químicos de fármacos e as superfícies das cavidades. A segunda análise se refere à semelhança entre os potenciais das cavidades apresentadas pelo SYBYL-X e KVFinder. A visualização dos potenciais nas cavidades das proteínas geradas pelo KVFinder foi realizada com o programa Pymol.

5.2.1 Estrutura extracelular do receptor de GABA_B ligado ao agonista baclofen

O receptor humano GABA_B é um receptor acoplado à proteína-G para inibir a neurotransmissão no sistema nervoso central. O receptor GABA_B medeia uma lenta e prolongada inibição sináptica através das proteínas G_i e G_o. O mau funcionamento deste receptor pode resultar em várias desordens neurológicas incluindo espasticidades³ e epilepsia(88). O Baclofen, um agonista seletivo do GABA_B, é usado clinicamente no tratamento da espasticidade muscular associada com múltiplas escleroses, paralisia cerebral e lesão medular(85).

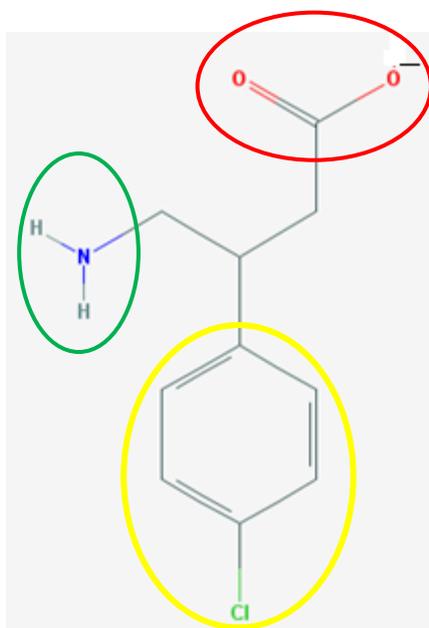


Figura 13: Estrutura do baclofen. Em amarelo o grupo clorofenil, em vermelho o grupo carboxilílico, em verde o grupamento amino.

O receptor de GABA_B funciona como um heterodímero das subunidades GBR1 e GBR2. O GBR₁ é responsável pelo reconhecimento do ligante através do domínio extracelular. Apesar do GBR2 não ligar a qualquer ligante do GABA_B, seu ectodomínio interage diretamente com o ectodomínio do GBR1 para melhorar a afinidade agonista que é requerida para a ativação do receptor. O domínio transmembranar do GBR2 é responsável pelo acoplamento da proteína-

³ Hipertonia dos músculos, acompanhada de rigidez e exagero nos reflexos

G. Cada subunidade GABA_B contém dois domínios distintos, LB1 e LB2. Os domínios LB1 e LB2, das duas subunidades, são altamente correlacionados.

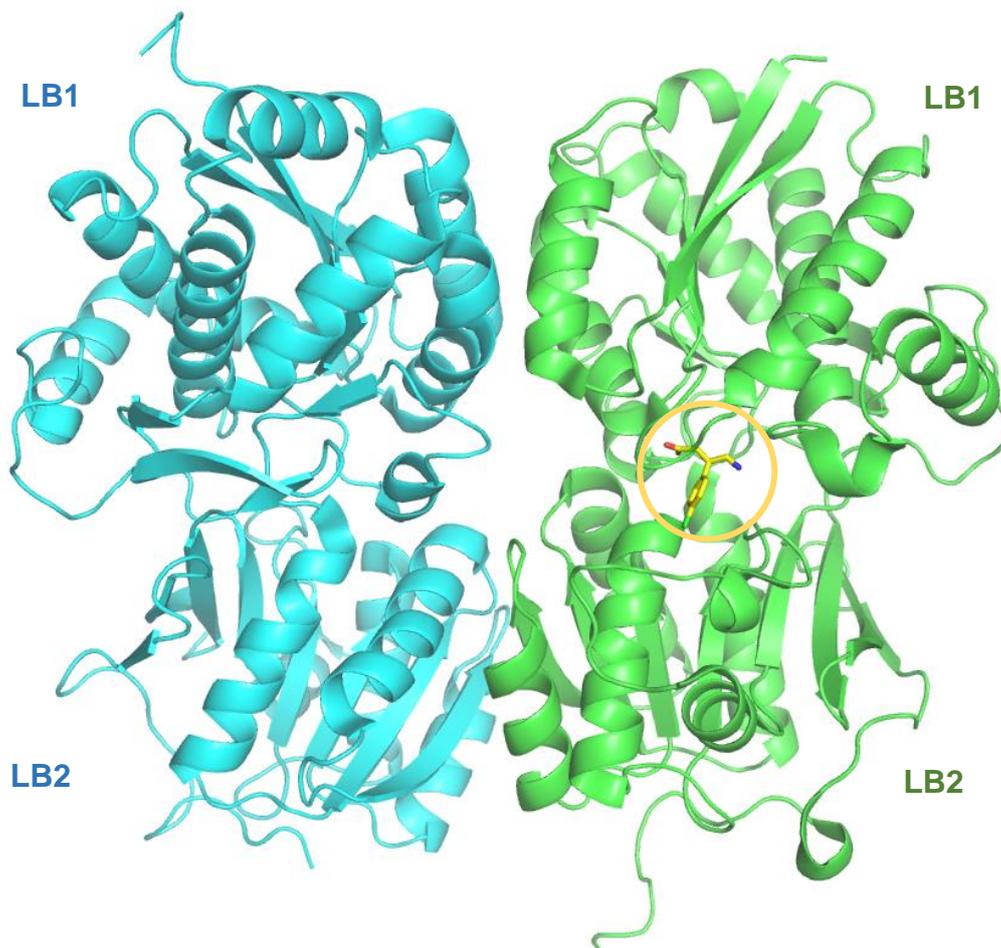


Figura 14: Estruturas das subunidades do GABA e seus heterodímeros (subunidades GBR1 e GBR2) com o agonista baclofen (círculo em laranja).

A subunidade GBR2 possui mais afinidade por ligantes agonistas que por ligantes antagonistas(88). O fármaco (R)-Baclofen contém um substituinte clorofenol na posição β e é reconhecido pelo GBR2 através de ligações de hidrogênio. O agonista faz contato diretamente com os resíduos Tyr250 e Trp278 e se liga à estrutura por uma cavidade interna inacessível ao solvente. O (R)-Baclofen é quase totalmente enterrado na subunidade da macromolécula, enquanto os fármacos antagonistas ligados ao receptor GABA ficam em uma cavidade com abertura maior, que os deixam em contato com o solvente(85).

A área da superfície molecular da cavidade, calculada pelo KVFinder, onde o agonista (R)-Baclofen se liga, é de 257\AA^2 . Por outro lado, a área que envolve todo o volume da cavidade, ou seja, uma superfície totalmente fechada com uma tampa, calculado originalmente pelo KVFinder, é de 323\AA^2 . A razão entre a superfície molecular da cavidade e a cavidade fechada é de $\sim 0,8$ (80%). Esta razão indica que o fármaco possui mais interações com os átomos da proteína do que com o solvente.

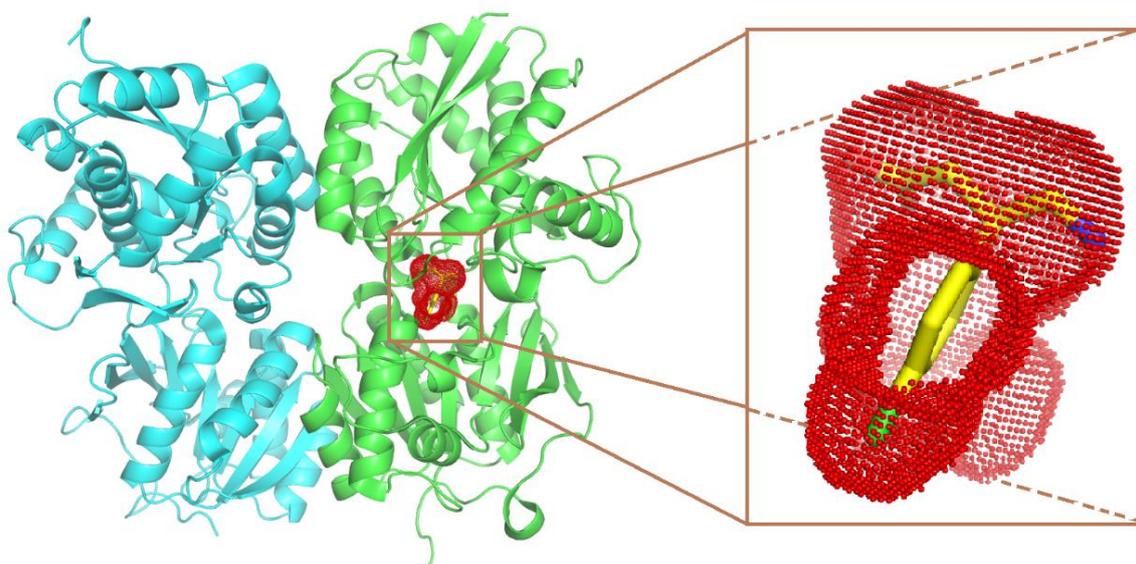


Figura 15: Estrutura do GABA_B e a cavidade onde se liga o agonista baclofen. Podemos observar que a cavidade engloba quase que totalmente o agonista.

5.2.1.1 Comparação Potencial Lipofílico KVFinder e SYBYL em relação à cavidade e ligante

O baclofen possui a denominação pela IUPAC de ácido 4-amino-3-(4-clorofenil) butanóico, sua estrutura possui peso molecular de 230,66 g/mol e logP igual a 1,3 (<https://pubchem.ncbi.nlm.nih.gov/>). Utilizou-se o programa SYBYL para a análise correspondente à superfície molecular do composto. Os grupos amino e carboxilato apresentam regiões mais hidrofílicas. O local onde está o clorofenil apresenta duas características distintas: (i) a região em que está o grupo fenil exibe cor mais esverdeada denotando uma transição entre as regiões

hidrofílica e hidrofóbica (Figura16B), (ii) o espaço hidrofóbico se concentra próximo ao cloro.

De acordo com os valores de hidrofobicidade atômica, o resultado do baclofen está correto, pois todos os valores são positivos para os halogênios implicando que estes elementos aumentam a lipofilicidade da molécula. Em geral, adicionar substituintes de halogênios irá aumentar a lipofilicidade das moléculas: a molécula torna-se maior, é mais polarizada e as forças de dispersão de London são aumentadas em conformidade(89).

O KVFinder mostrou características da cavidade que complementavam com as características do composto. Na Figura 17, pode-se ver, de acordo com a orientação da Figura, regiões hidrofóbicas indicadas em vermelho no topo da cavidade e próximas ao grupo clorofenil. A base da cavidade se mostrou mais hidrofílica, em azul, indicando uma superfície que combina com a propriedade polar dos grupamentos amino e carboxilato. Nestas regiões os átomos do ligante possuem maior proximidade com a superfície da cavidade. A região em verde caracteriza uma região em transição mais próxima de uma propriedade hidrofílica. A cavidade no Pymol recebeu as escalas de cores: azul, branco e vermelho, referente aos valores de lipofilicidade que vão de -1 até 1. Sendo o azul menos lipofílico (hidrofílico) e o vermelho mais lipofílico(hidrofóbico).

Comparando com a Figura 16B, podemos observar as semelhanças quanto à hidrofobicidade do ligante. A diferença mais sensível é a parte hidrofílica, em azul, que se sobressai quando é aplicado o KVFinder.

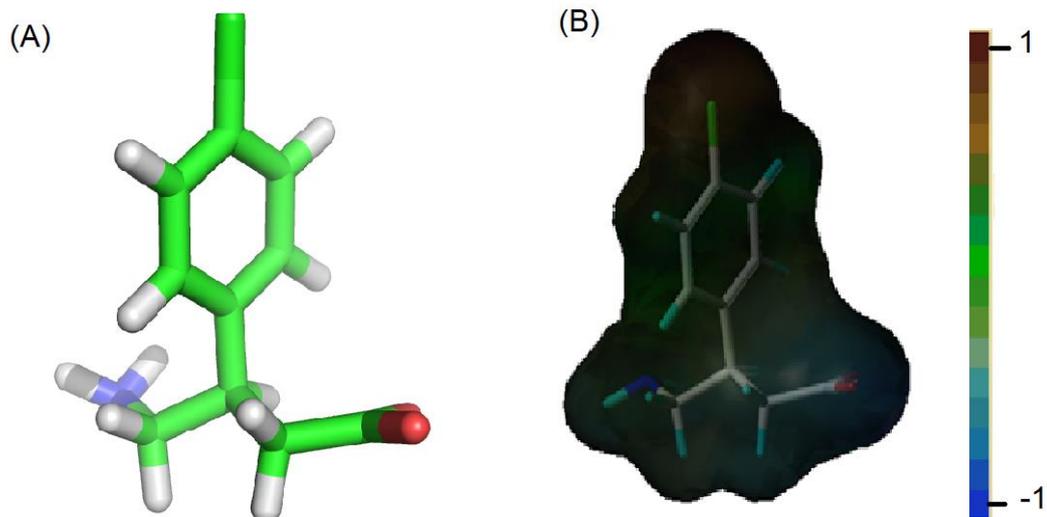


Figura 16: Potencial lipofílico calculado pelo programa SYBYL. Em A, temos a representação da estrutura do Baclofen, gerada pelo programa Pymol. Em B, a superfície molecular do baclofen com o potencial lipofílico. A cor azul indica hidrofiliabilidade e vermelho, hidrofobicidade.

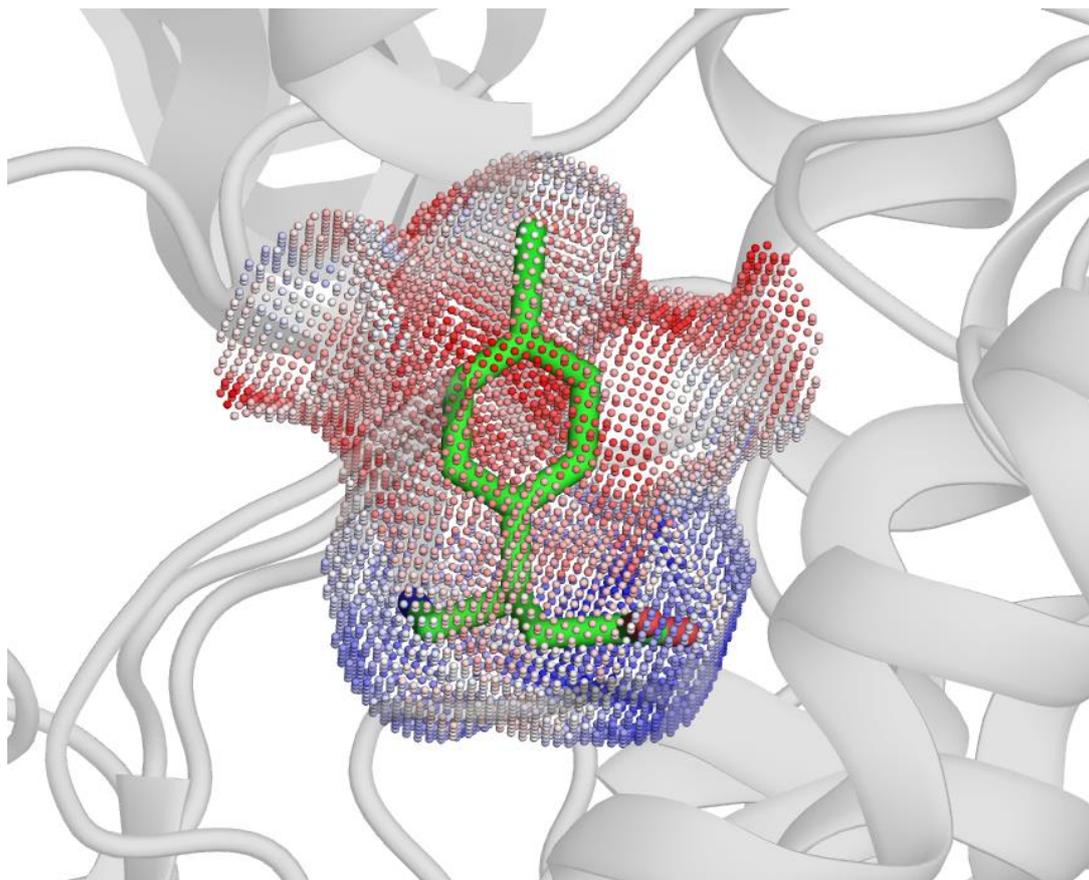


Figura 17: Potenciais lipofílicos simulados com o KVFinder. O azul retrata a hidrofiliabilidade e o vermelho a hidrofobicidade. A região em azul na cavidade é mais próxima ao grupamento amino e a região em vermelho (menos lipofílico) está próxima ao grupo clorofenil. A visualização foi gerada usando o software Pymol.

5.2.2 Densidade de acceptor e doador de hidrogênio: Relação Ligante-Cavidade

A densidade de acceptor e doador de hidrogênio também foi analisada quanto à sua complementaridade ligante-cavidade. Neste caso, regiões do ligante onde há aceptores foram comparadas com regiões da cavidade onde encontramos doadores de hidrogênio. No caso do agonista baclofen, o reconhecimento pela macromolécula também ocorre pelas ligações de hidrogênio(90). O agonista realiza duas ligações de hidrogênio, uma em cada extremidade e está diretamente ligado aos resíduos Tyr250 e Trp248.

A propriedade relativa à densidade de acceptor e doador de hidrogênio, utilizando o cálculo do SYBYL para o baclofen, apontou regiões que possuem estes potenciais. Na Figura 18, a região acceptora é demonstrada em azul e está localizada próxima ao grupamento carboxilato. Os resultados obtidos pelo KVFinder, Figura 19, demonstram que os pontos da cavidade que apresentam maior densidade de doadores de hidrogênio se concentram junto ao carboxilato, indicando a complementariedade cavidade-ligante.

A Figura 20 exhibe, em rosa, onde há maior densidade de doadores de hidrogênios que se concentram na superfície próxima ao grupo amino da molécula baclofen. O algoritmo de detecção de densidade de acceptor de hidrogênio incluído no KVFinder consegue localizar regiões onde há alta densidade de aceptores de hidrogênio, demonstrado pela cor rosa na superfície próxima aos grupos amino e carboxila da molécula de baclofen (Figura 21).

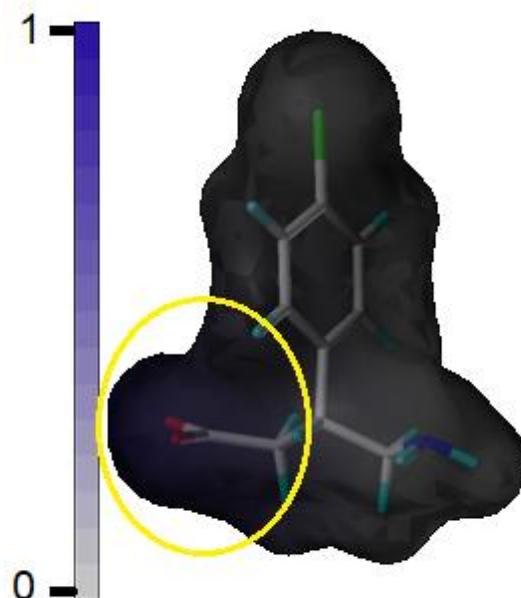


Figura 18: Densidade de aceptor de hidrogênios do Baclofen. A região da superfície que está em azul, círculo amarelo, apresenta maior potencial aceptor de hidrogênio representado pelo grupamento carboxilato. Visualização gerada pelo SYBYL.

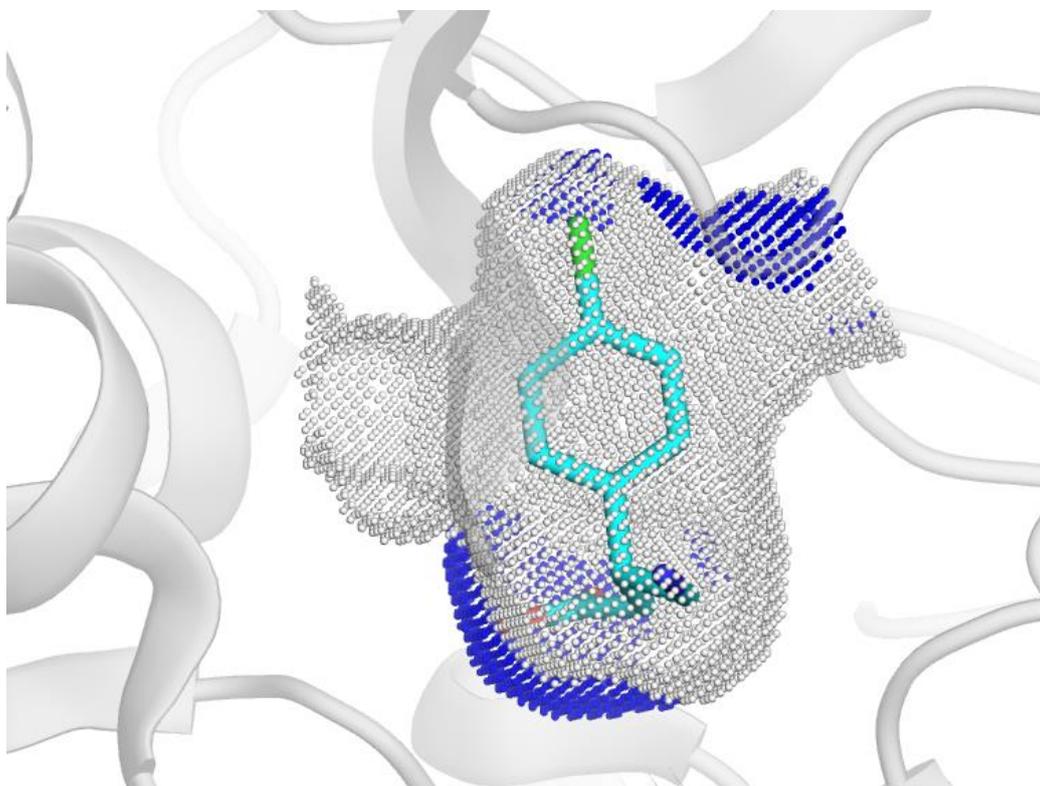


Figura 19: Densidade de doadores de hidrogênio da cavidade. As regiões com densidade de **doadores** de hidrogênio são representadas em azul. Visualização gerada pelo Pymol.

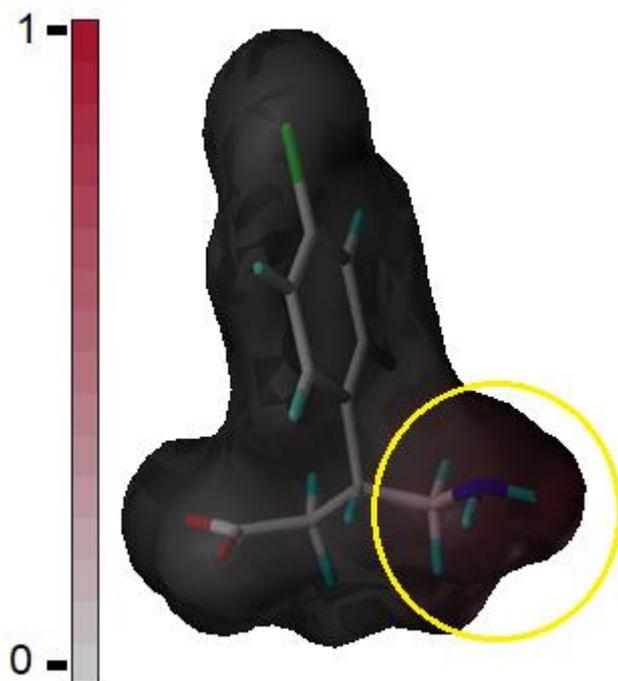


Figura 20: Densidade de doador de hidrogênios do Bacoflen. A região com maior potencial doador de hidrogênio está localizada próximo ao grupo amino, círculo em amarelo. Visualização gerada pelo SYBYL.

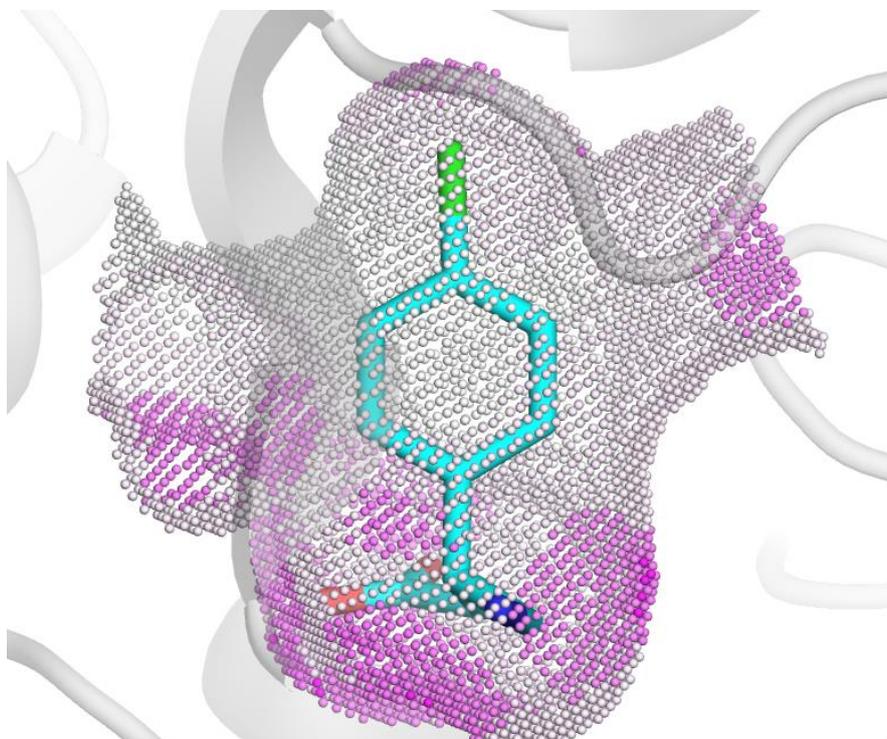


Figura 21: Densidade de aceptores de hidrogênio da cavidade. A região próxima do grupamento amino e carboxila apresenta maior potencial de aceptores de hidrogênio. Representação obtida baseando no programa Pymol.

Representamos a relação entre a densidade de hidrogênio e os resíduos próximos ao ligante. Para isso, foi realizada a avaliação quanto à densidade de doadores de hidrogênio, em azul (Figura 22). As regiões onde estão os resíduos Ser131, Tyr150, Thr252 apresentam maior probabilidade de ocorrer ligação de hidrogênio.

Na Figura 23, a presença dos resíduos Cys129, Ser130, Ser152, Ser153, Tyr250 e Tyr279 influenciam na densidade de acceptor de hidrogênio. Vale ressaltar que os cálculos foram realizados incluindo resíduos a uma distância de 4Å da cavidade.

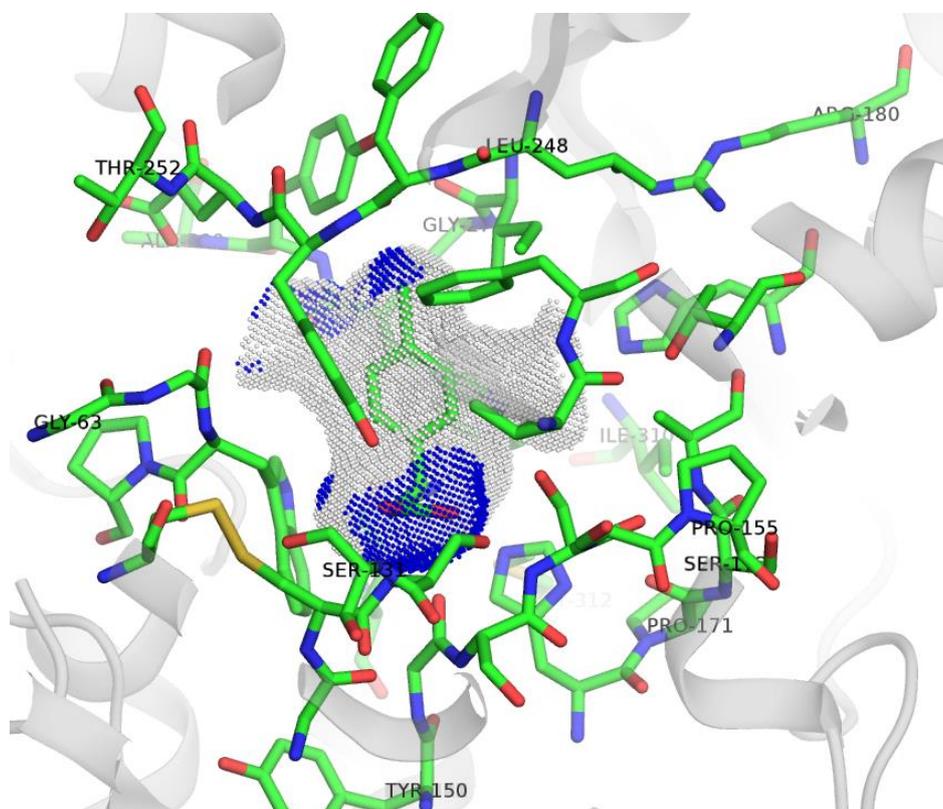


Figura 22: Influência dos resíduos na densidade de doador de hidrogênios na cavidade. Em azul estão as regiões com maior potencial doador de hidrogênio influenciado pelos resíduos: Ser131, Tyr150, Thr252.

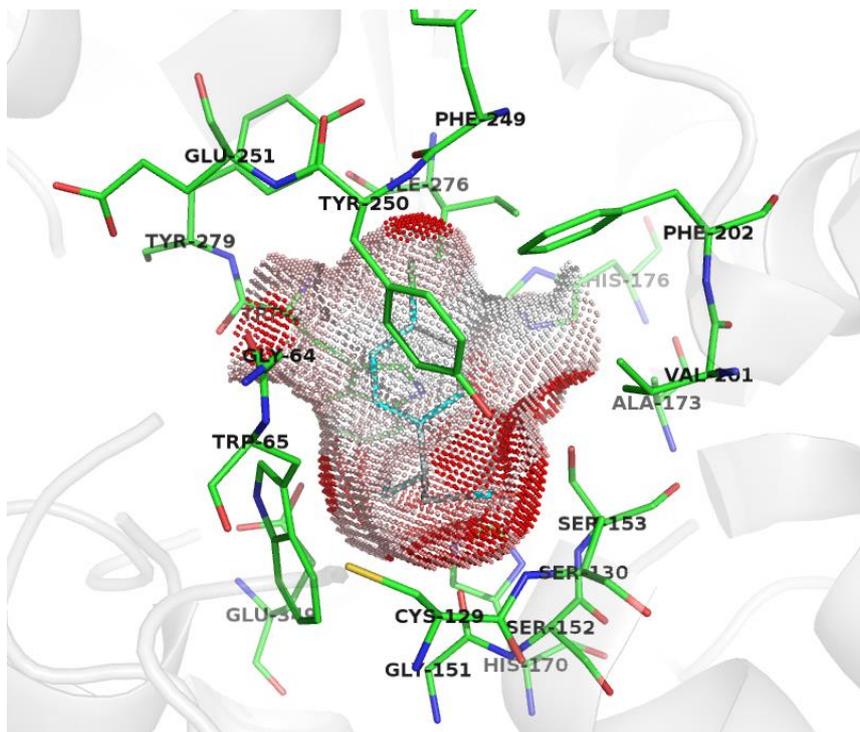


Figura 23: Influência dos resíduos na densidade de aceitador de hidrogênios na cavidade. Próximo da cavidade encontram-se entre os resíduos: Cys129 Ser130, Ser152, Ser153. Os resíduos Tyr250 e Tyr 279 se apresentam na parte superior da cavidade.

5.2.3 Proteína HIV-1 Protease em complexo com o inibidor VX-478

A proteína HIV-1 protease é um alvo terapêutico conhecido para o tratamento de infecção pelo HIV. O composto VX-478 é um potente inibidor oralmente biodisponível da enzima HIV-1 protease(91). O ligante possui peso molecular de 506Da e inibe, além da protease do HIV-I, a protease do HIV-II possuindo uma constante de inibição de 0,6nM(91).

No inibidor VX-478 é possível observar grupos fenil e sulfonilanilinas e em outra porção da molécula um grupo THF Carbamato (Figura 24).

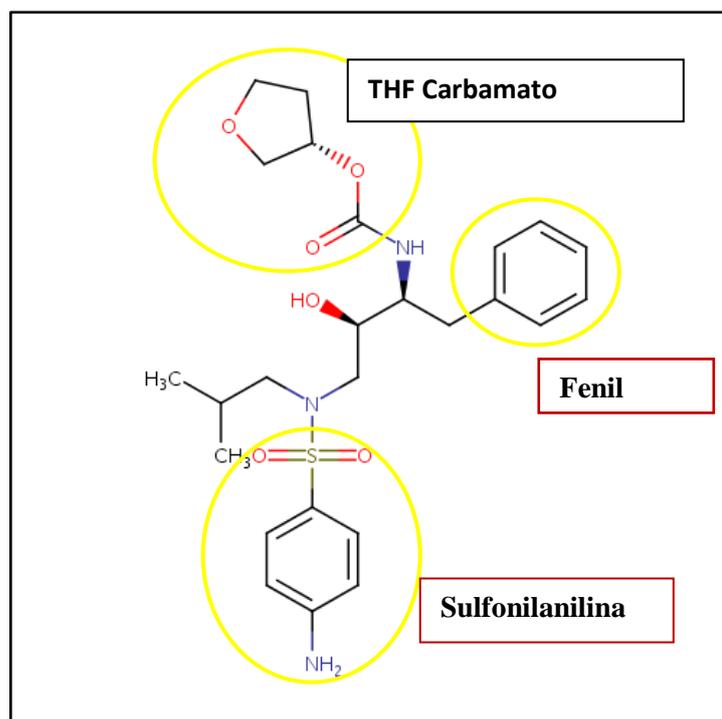


Figura 24: Composto inibidor VX-478. Em destaque compostos como benzeno, sulfonamida e THF Carbamato na sua composição.

O composto possui alta potência antiviral e baixa toxicidade celular. Podemos ver a cavidade gerada pelo KVFinder em que se encontra o ligante na Figura 25. A área calculada da superfície da cavidade da macromolécula foi de 700Å^2 enquanto que a área que envolve o volume total da cavidade apresentou 1049Å^2 . A razão entre a área que está em contato com a superfície da molécula e a área que envolve todo o volume é de 0,68, ou seja, 68% do espaço onde se encontra o inibidor pertence à proteína. Este valor concorda com dados encontrados na literatura, que aponta o valor em torno de 60%(91).

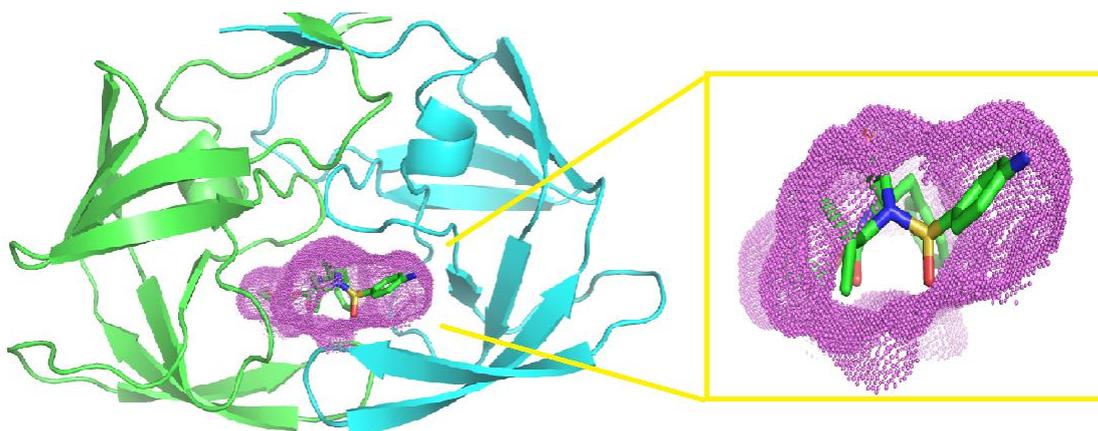


Figura 25: Cavidade e ligante do complexo 1HPV. Pode-se observar que a cavidade engloba o ligante em sua estrutura deixando uma pequena abertura para o ambiente. Representação obtida baseando no programa Pymol.

5.2.3.1 Comparação Potencial Lipofílico do KVFinder e SYBYL em relação à cavidade e ligante

O composto VX-478 possui peso molecular igual a 506Da e é um inibidor com alta disponibilidade oral(91). O composto, quando ligado à proteína, tem o grupo sulfonilânina parcialmente englobado pela macromolécula em um ambiente hidrofóbico formado pela Ile50 e Ile84. Quanto à lipofilicidade de cada parte da estrutura, pode-se observar que o grupo fenil é a porção que apresenta maior hidrofobicidade (Figura 26). Esta característica é visualizada pela cor avermelhada (círculo em amarelo). A parte mais hidrofílica se concentra no grupo derivado do THF carbamato.

Correlacionando o ligante com a cavidade do KVFinder podemos ver que a região onde estão os grupos sulfonilânina e fenil (Figura 27), apresenta uma região mais lipofílica (hidrofóbica). O grupo THF carbamato define uma região com menor lipofilicidade (hidrofílica). A análise das cavidades mostra que há predomínio das regiões hidrofílicas.

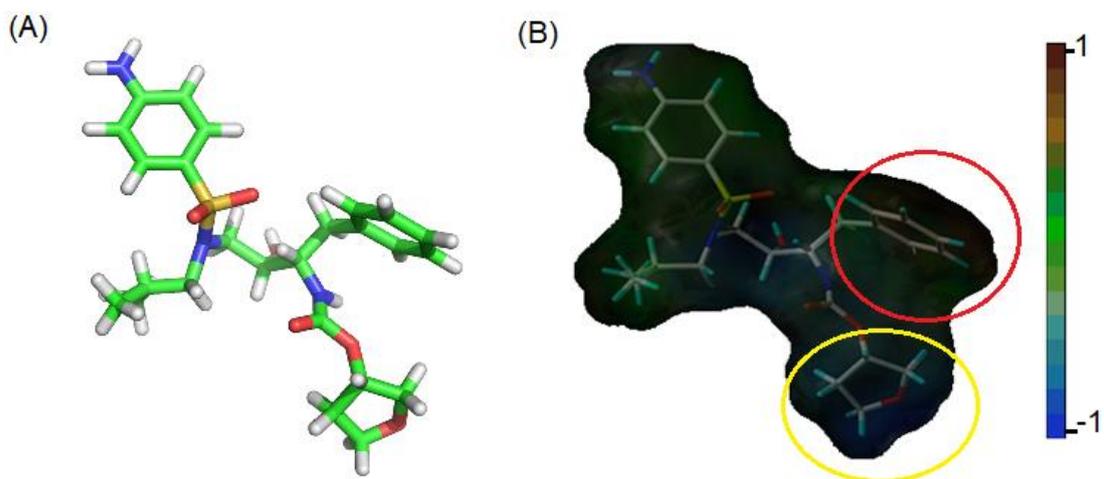


Figura 26: Potencial lipofílico do ligante VX-478. (A) Estrutura do VX-478. (B) região onde se localiza o grupo fenil (círculo em vermelho) e o THF carbamato (círculo em amarelo) possui, respectivamente, o maior e o menor potencial lipofílico.

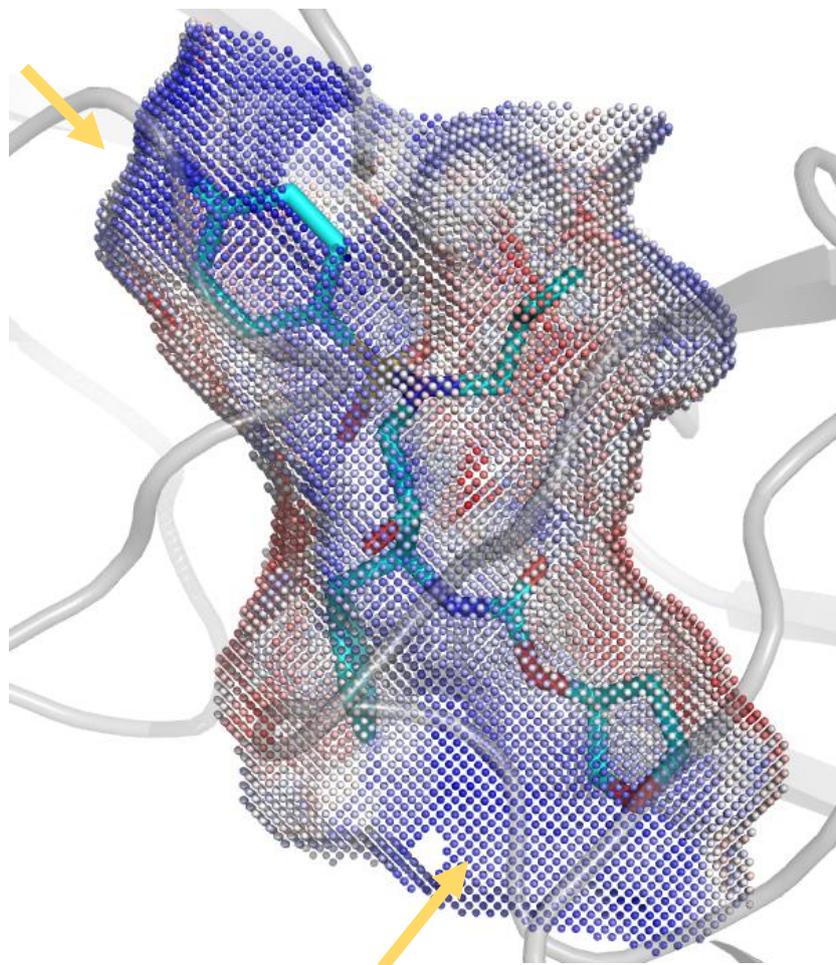


Figura 27: Potencial lipofílico da cavidade onde se encontra o ligante VX-478. O topo da cavidade, onde se localiza o grupo sulfonilânilina e a região inferior onde está o grupo THF carbamato (indicados pelas setas amarelas), apresentam menor potencial lipofílico. As regiões hidrofóbicas se localizam próximas ao grupo fenil.

5.2.3.2 Densidade de acceptor e doador de hidrogênio: Relação Cavityde-Ligante

A ligação do inibidor VX-478 é marcada por ligações de hidrogênio envolvendo os grupos sulfonilânilina e o oxigênio do THF carbamato. Ainda na estrutura é possível observar vários grupos como amino e hidroxila que podem fazer ligações de hidrogênio(91). As Figuras 28 e 29 apresentam a complementaridade entre as superfícies aceptores de hidrogênio do ligante e doadoras de hidrogênio da cavidade. O VX-478 possui grande área em sua superfície que podem receber hidrogênios ao ligar-se a uma macromolécula (Figura 28A e Figura 28B). Logo o KVFinder demonstra uma extensa região de doadores de hidrogênio (Figura 29), em acordo com a potencialidade do composto de realizar estas ligações.

A superfície do inibidor com potencial doador de hidrogênio se concentra próxima ao grupamento sulfonilamina e região central onde há grupos OH e NH (Figura 30). A superfície da cavidade, mostrada em rosa, Figura 31, compreende alta densidade de acceptor de hidrogênio que pode realizar interações com o composto.

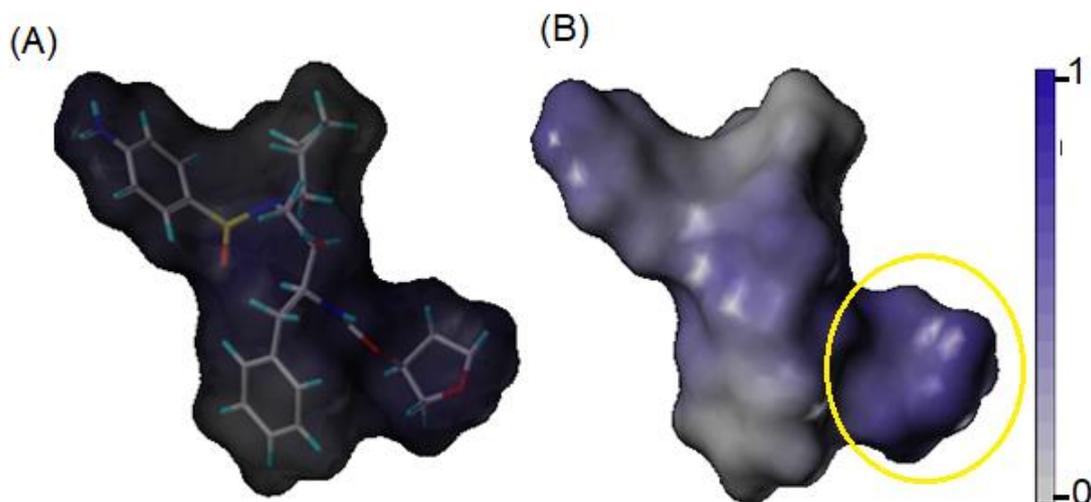


Figura 28: Representação das regiões aceptoras de hidrogênios da molécula VX-478. Em A, a superfície molecular transparente da molécula, em B, a representação opaca mostrando que as regiões, em azul, principalmente próximo do THF carbamato (círculo em amarelo) possui potencial acceptor de hidrogênio.

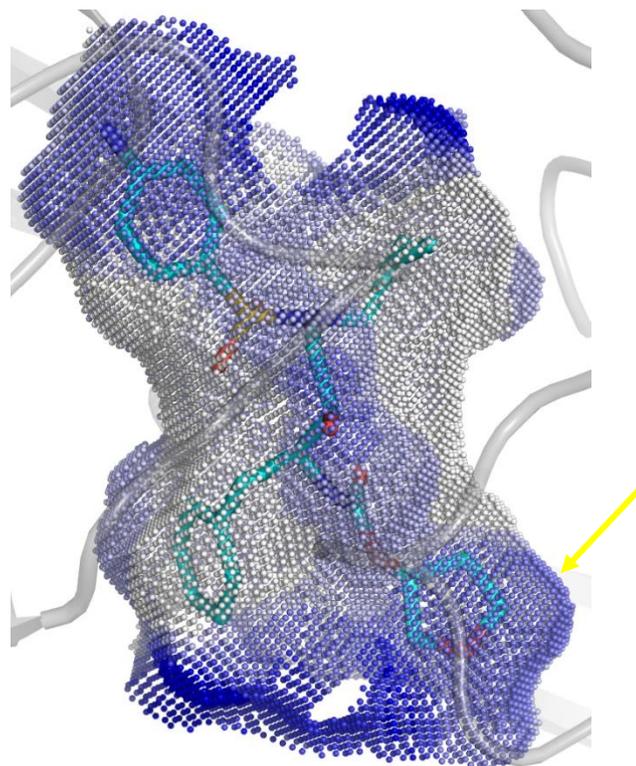


Figura 29: Densidade de doador de hidrogênios para a cavidade onde se liga o VX-478. Os locais, em azul, da cavidade próximos ao grupo THF carbamato (seta em amarelo), possui maior potencial de doador de hidrogênios.

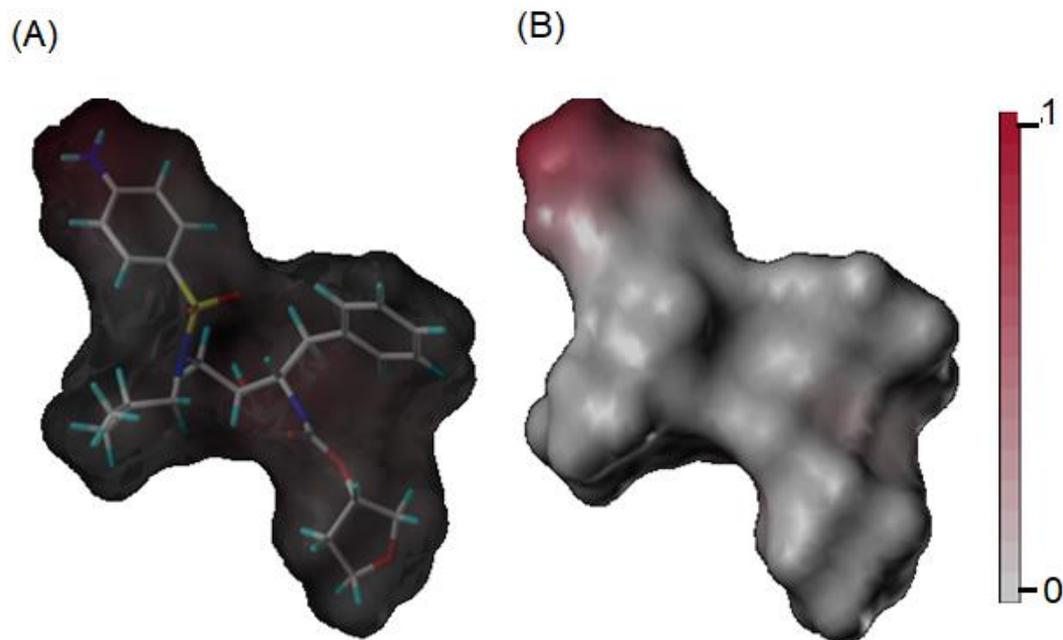


Figura 30: Representação do ligante VX-478 em que apresenta densidade de doadores de hidrogênio. A Figura A apresenta o ligante e sua superfície molecular transparente, enquanto em B, a superfície molecular é apresentada de forma opaca e possui no topo coloração avermelhada indicando a região doadora de hidrogênio. Ilustração gerada pelo SYBYL.

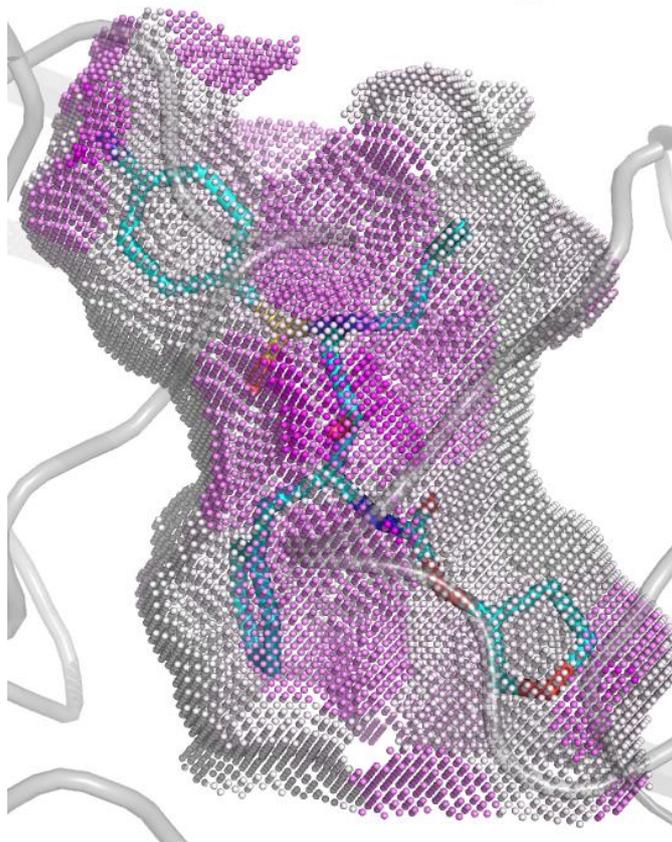


Figura 31: Potencial aceptor de hidrogênio apresentado pela cavidade do ligante VX-478. A coloração em rosa indica o potencial de regiões aceptoras de hidrogênio. Na parte superior da cavidade, onde se encontra o grupo sulfonilânilina, possui maior densidade de aceptor de hidrogênio. Ilustração gerada pelo pymol.

Através da ferramenta Pymol foram identificados os resíduos que estão a 4Å de distância do ligante. Na Figura 32, os resíduos Asp25, Gly27, Ala28, Asp29, Asp30, influenciam no potencial de aceptor de hidrogênio da cavidade.

As regiões em branco são regiões com pouca ou nenhuma probabilidade de ocorrer ligações de hidrogênio. As cores somente se destacaram em regiões específicas devido à metodologia utilizada para o cálculo da densidade. Ao calcular os valores da densidade, a soma de aceptores e doadores foi dividida pelo volume total da esfera que circunda cada ponto da cavidade. A divisão pelo volume total de um raio de 4Å fez com que a densidade de aceptores das outras regiões se reduzisse consideravelmente. Além disso, devemos considerar que somente parte do átomo pode estar incluída no cálculo, diminuindo ainda mais a sua influência na cavidade.

A Figura 33 apresenta os valores relativos à densidade de doadores de hidrogênio que possuem maior domínio na superfície da cavidade. Os resíduos Asp29, Asp30, Thr31 e Ile84, próximos à cavidade aumentam a probabilidade de interação de ligação de hidrogênio com o ligante.

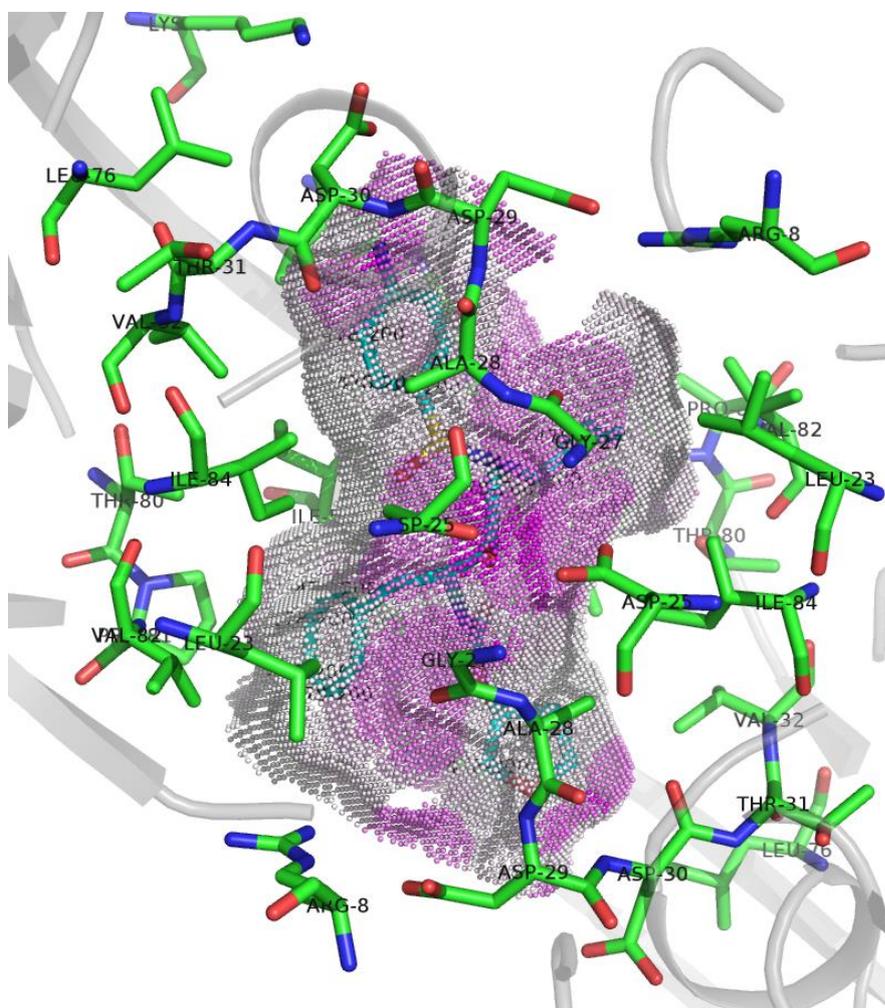


Figura 32: Resíduos que interagem com o ligante VX-478. A Figura representa a cavidade onde está ligado o composto e, em rosa, as regiões com maior potencial de aceitor de hidrogênios próximo aos resíduos Asp25, Asp29, Asp30.

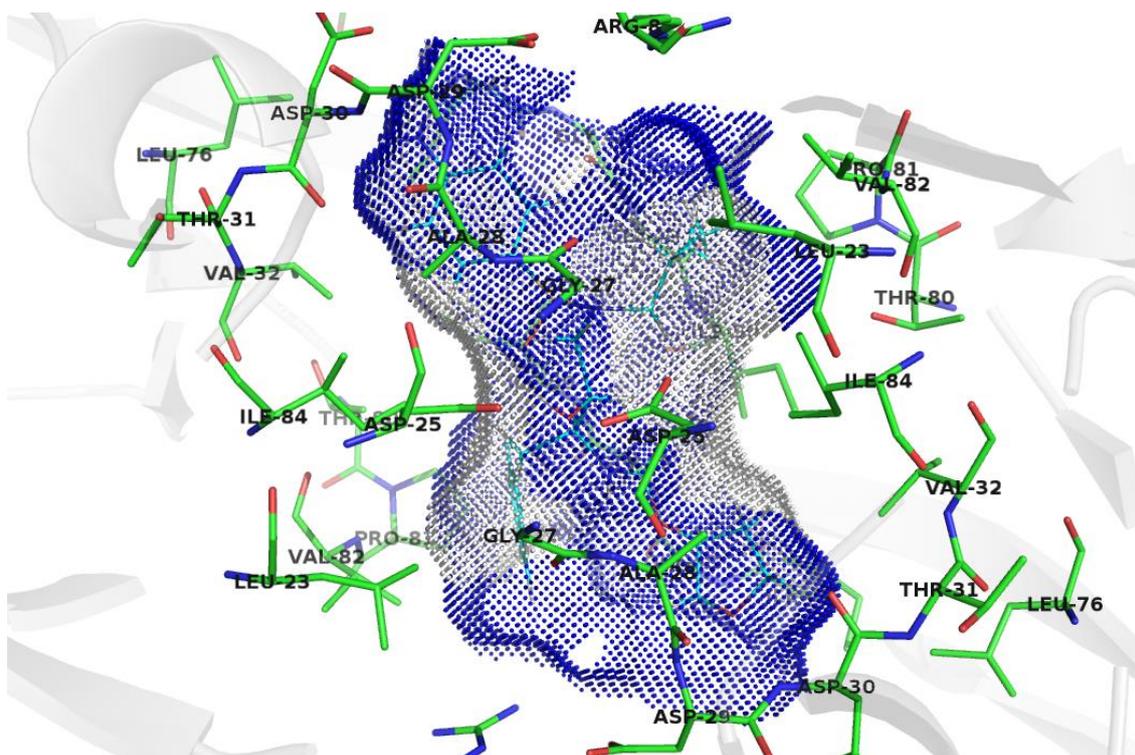


Figura 33: Representação da cavidade e resíduos da macromolécula HIV-1 protease. Podemos observar que a densidade de doador de hidrogênio é superior à densidade de aceptores de hidrogênio. Na parte superior e inferior da cavidade onde se encontram os resíduos Arg8, Asp28, Asp29, Asp30 e Thr31 o potencial doador de hidrogênio é maior.

5.2.4 Maltase-Glucoamilase Humana e Acarbose

A dieta humana baseia-se principalmente em amido como fonte de glicose. Em humanos as α -amilases pancreáticas e salivares são enzimas que clivam as ligações internas $\alpha(1-4)$ de amido em cadeias pequenas e lineares de dextrina. A mistura de dextrinas é então hidrolisada nas suas extremidades não redutoras em glicose por duas exo-hidrolases: maltase-glucoamilase (MGAM) e a sucrase-isomaltase (SI)(92).

O MGAM e SI são enzimas ligadas às membranas, cada uma contendo duas subunidades catalíticas: uma subunidade N-terminal e uma C-terminal.

Do ponto de vista terapêutico, devido a limitadas informações sobre o processamento intestinal dos produtos da α -amilase, é importante entender o mecanismo catalítico destas enzimas como inibidores das α -glucosidases que são alvos destas enzimas na via de digestão do amido. Estes inibidores são

atualmente usados para retardar a produção de glicose e auxiliar no tratamento de diabetes tipo II. Acarbose, Figura 34, é um inibidor atualmente no mercado que quando testado com a subunidade recombinante N-terminal da MGAM (NtMGAM) apresenta maior atividade inibitória com $K_i=62\pm 13 \mu\text{M}$ (92).

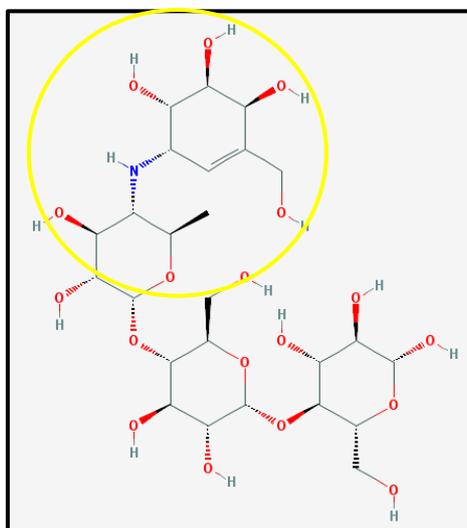


Figura 34: Estrutura da acarbose. O círculo em destaque (amarelo) indica a acarvosina.

O sítio ativo do NtMGAM é uma cavidade formada principalmente pelos resíduos de Folhas- β C-terminal da estrutura barrel $(\beta/\alpha)_8$. Os dois anéis, destacados em amarelo na Figura 34, representam a unidade acarvosina e ficam inseridos dentro do sítio catalítico. Em comparação com os anéis de acarvosina na ligação da molécula de acarbose, os outros dois anéis que não estão no sítio ativo, na forma reduzida formam interações com o NtMGAM, por intermédio do solvente. Isto inclui essencialmente ligação de hidrogênio com o resíduo Tyr605 através de interações indiretas com a molécula de água.

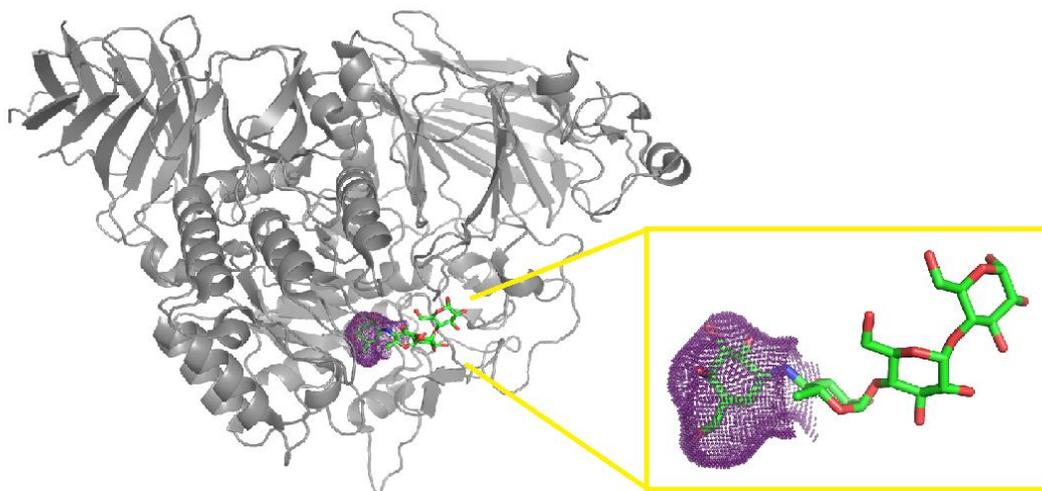


Figura 35: Região onde se localiza a estrutura da acarbose. Somente a parte que corresponde à acarvosina se localiza dentro da cavidade.

5.2.4.1 Potencial Lipofílico da cavidade do complexo glucoamilase humana e acarbose

As interações que se destacam na ligação dos resíduos do sítio ativo da glucoamilase e a acarvosina são as ligações de hidrogênio. Na Figura 36, geradas pelo SYBYL, o baixo potencial lipofílico é predominante nas cavidades. Na Figura 36B, a parte mais externa da cavidade ou, sua extremidade, é uma região mais hidrofílica devido aos resíduos ácido aspártico, arginina, triptofano e tirosina. Os resíduos desta cavidade estão expostos ao solvente o que explica a característica de hidrofiliidade. Na Figura 36A, há uma pequena região com característica hidrofóbica.

A cavidade encontrada pelo KVFinder demonstra características semelhantes às achadas pelo programa SYBYL. De acordo com os resultados alcançados a partir do cálculo dos potenciais, no bolso, em sua maior extensão, predomina a baixa lipofilicidade devido a resíduos que estão em contato com o solvente. Na Figura 37, há uma região hidrofóbica mais extensa. Dados da literatura mostram que existem resíduos hidrofóbicos, o Trp e Phe, próximos à região onde a acarvosina se liga na macromolécula(92).

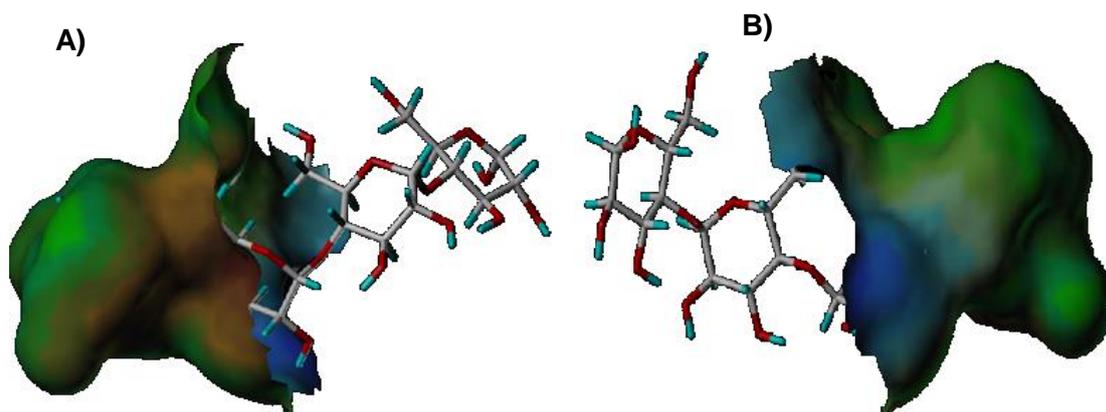


Figura 36: Acarvosina presente na cavidade de ligação com a glucoamilase. Cavidade gerada pelo SYBYL demonstrando a posição do ligante por dois ângulos. O baixo potencial lipofílico predomina na maioria das regiões. Em A, existe uma pequena região hidrofóbica, em marrom, na extremidade da cavidade devido aos resíduos Trp e Phe.

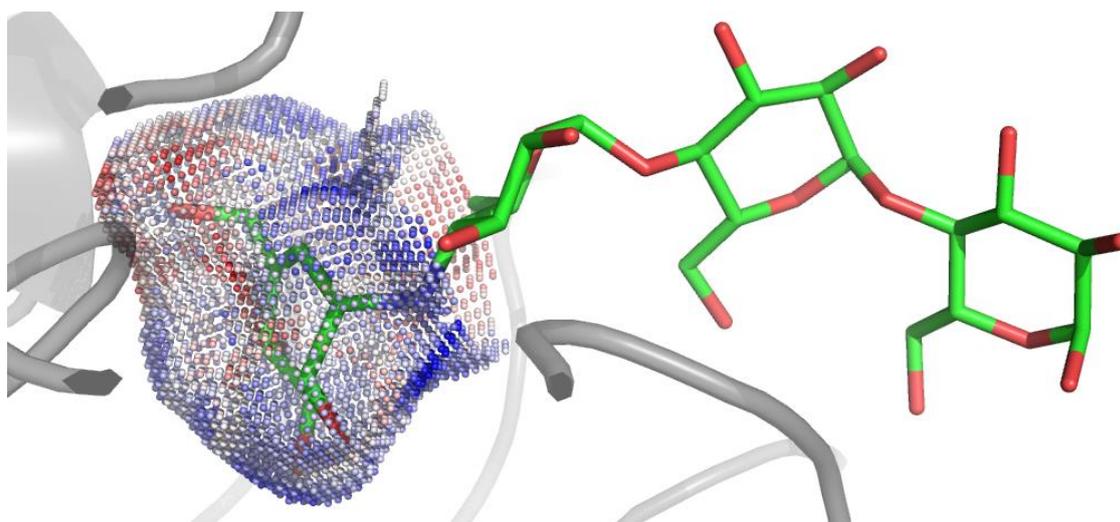


Figura 37: Cavidade gerada pelo KVFinder onde a acarvosina se liga à glucoamilase humana. Nesta posição temos regiões hidrofílicas mais extensas, predominado por toda a cavidade.

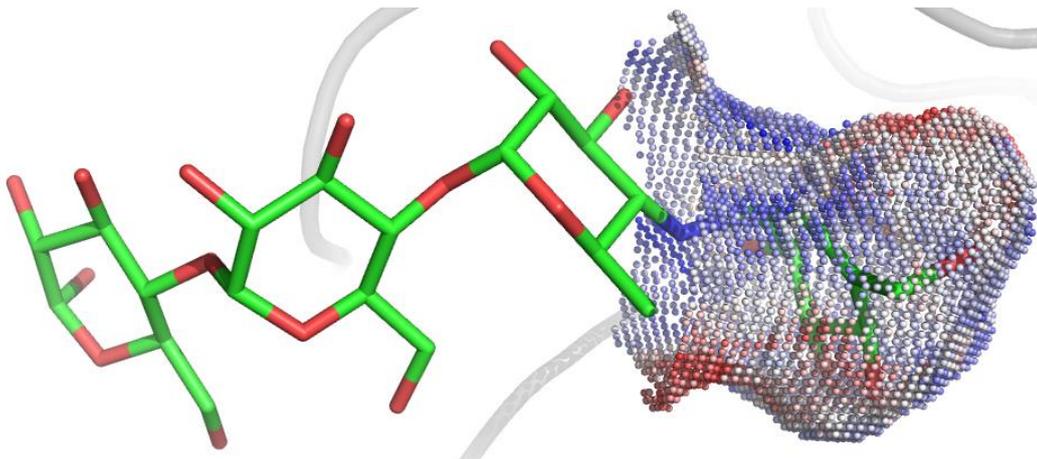


Figura 38: Caverna onde est presente a acarvosina. Vemos que a caverna, por outro ngulo, tambm , em sua maioria, hidroflica, apresentando poucos pontos em vermelho indicando hidrofobicidade.

5.2.4.2 Densidade de receptor e doador de hidrognio: Relaco Caverna-Ligante

A prxima discusso  relativa  densidade de doador de hidrognio da molcula de acarbose. A caracterstica apresentada  influenciada pelo nmero de oxignios e nitrognios presentes na estrutura do frmaco. A acarvosina que est presente dentro do stio ativo possui dois anis que contm estes tomos, responsveis pela complementaridade receptor-doador de hidrognio na interao com a protena. A Figura 39 exibe a molcula de acarbose com superfcie opaca e densidade de receptores de hidrognio. A superfcie com cor azul um pouco mais intensa est localizada nos dois anis que esto fora do stio ativo e prximas das regies que possuem tomos de nitrognio e oxignio da acarvosina.

A Figura 40 mostra as regies da caverna em que a densidade de doadores de hidrognio  mais acentuada. A densidade calculada pelo KVFinder  influenciada por tomos doadores presentes nos resduos Asp203, Arg256, Asp542, His600.

A ligao da acarbose no stio ativo envolve numerosas ligaes de hidrognio. Na Figura 41, a cor rosa demonstra o potencial de doadores de

hidrogênio. O colorido mais robusto demonstra os átomos que interagem com resíduos do sítio ativo.

A cavidade encontrada pelo KVFinder e visualizada pelo Pymol mostra somente uma região com alto valor de densidade de aceptores de hidrogênio. A Figura 42, a cor em azul refere-se à extremidade onde há maior número de resíduos capazes de interagir com o hidrogênio. Estas interações ocorrem através dos resíduos Asp443, Asp542, Arg526 e Tyr605, muitas são realizadas por intermédio de moléculas de água(81). Essa é uma característica deste complexo enzimático em que o fármaco está parcialmente dentro da cavidade. A baixa densidade de aceptores na cavidade indica que pode haver resíduos doadores de hidrogênios, mas a probabilidade de interação pode ser pequena.

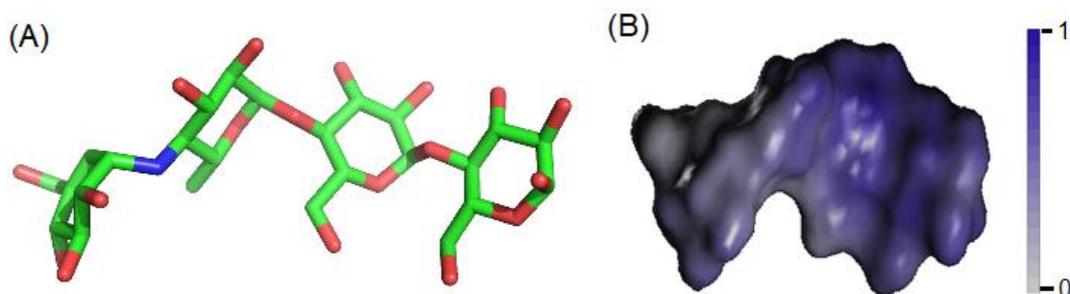


Figura 39: Molécula de acarbose com densidade deceptor de hidrogênio mostrada com superfície molécula. A estrutura mostra homogeneidade em toda sua estrutura em relação à densidade de aceptores de hidrogênio.

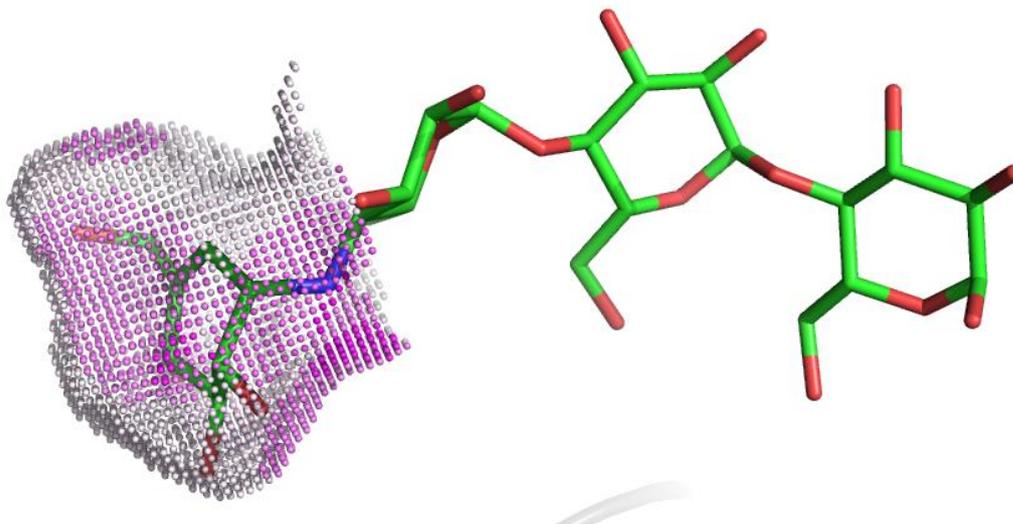


Figura 40: Cavidade da glucoamilase com densidade de **DOADORES** de hidrogênio. A densidade se distribui homogeneamente pela cavidade. Figura gerada pelo Pymol com o algoritmo agregado ao KVFinder.

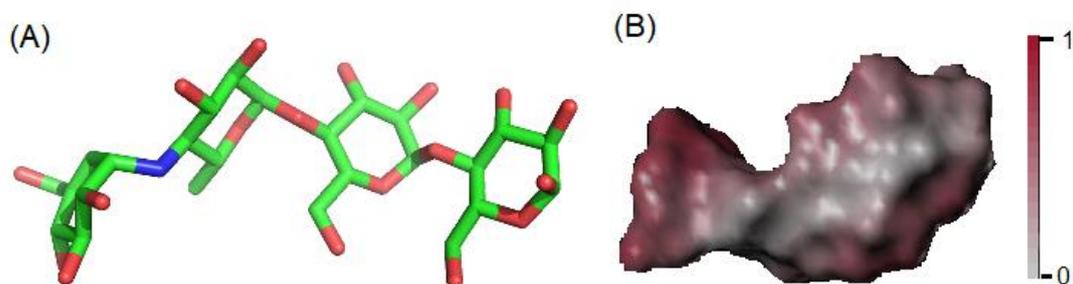


Figura 41: Molécula de acarbose (A) com densidade de doador de hidrogênio em sua superfície molecular(B).

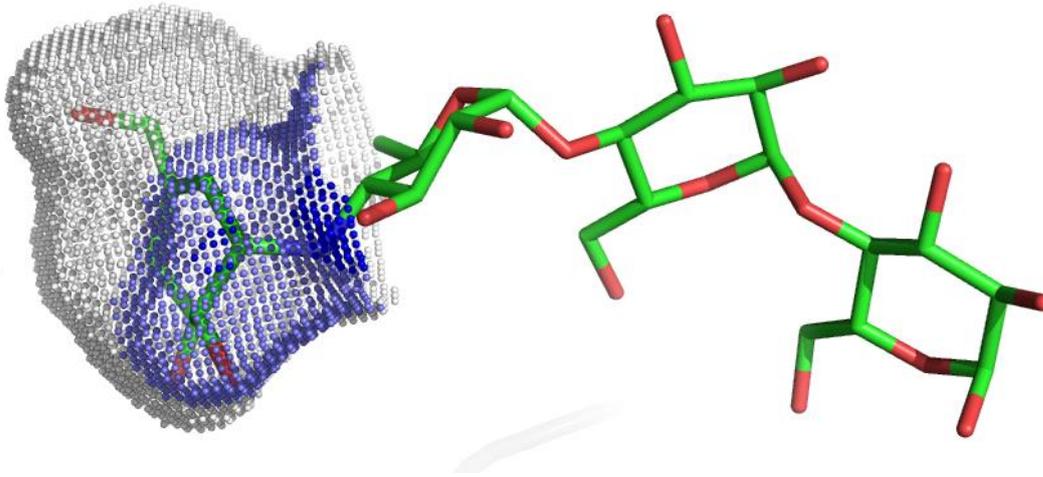


Figura 42: Densidade de aceptores gerada pelo KVFinder. A imagem demonstra que a densidade de ACEPTORES de hidrogênios da cavidade se localiza na extremidade.

A próxima comparação realizada está relacionada à cavidade e resíduos que a cercam. Neste caso, foram destacados locais com maior densidade de aceitor de hidrogênio na visualização. Estes locais estão presentes os resíduos Asp203, Tyr299, Asp327, Ile328, Trp406, Arg526, Asp542 e His600. As regiões onde existe maior densidade de aceptores de hidrogênio se localizam nas extremidades. A representação está ilustrada na Figura 43.

A densidade de doador de hidrogênio, representado pela cor rosa na Figura 44, possui maior potencial na extremidade da cavidade. Nestes locais estão os resíduos Asp203, Met444, Asp542, Arg526.

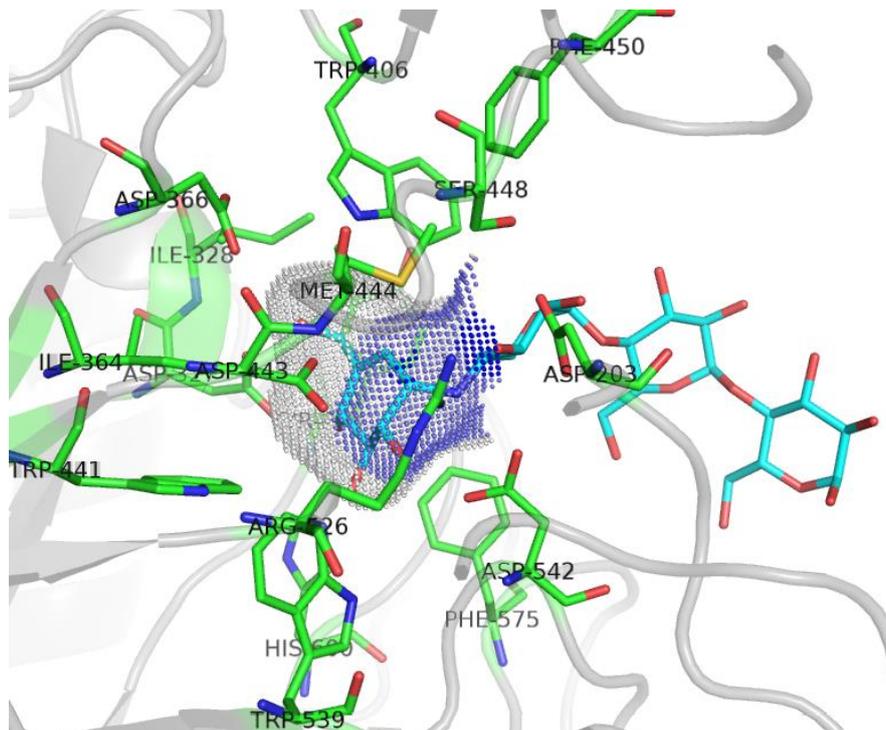


Figura 43: Relação cavidade-resíduo para densidade de aceptores de hidrogênio. Pode-se observar maior valor de densidade de aceptores na extremidade da cavidade próximo aos resíduos Asp203, Asp443, Arg526, Asp542.

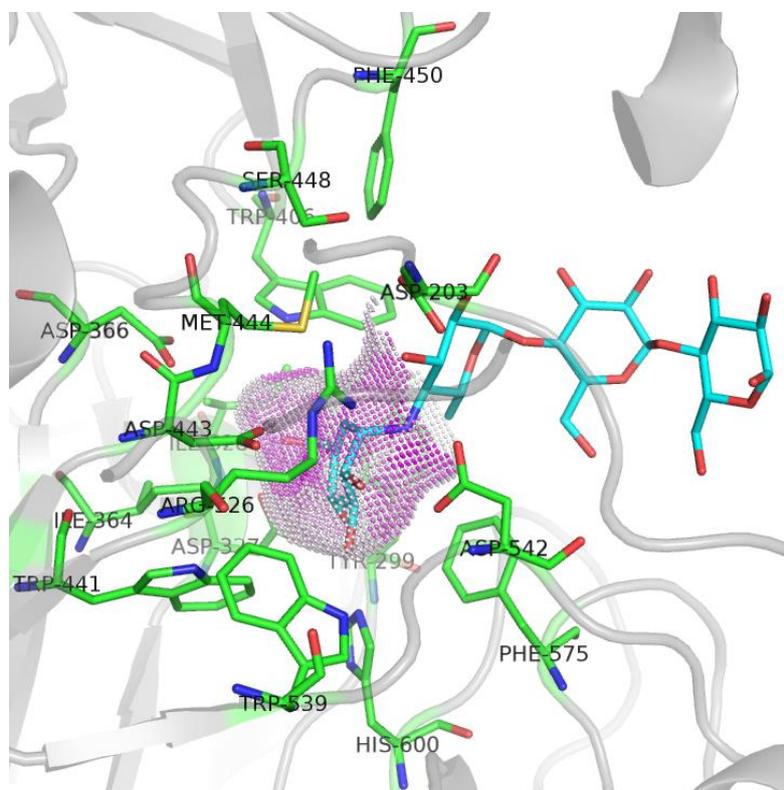


Figura 44: Região da cavidade onde se apresenta maior probabilidade de doador de hidrogênio. Esta região se concentra na extremidade da cavidade, onde há resíduos como Asp203, Met444, Asp542, Arg526.

5.3 Análise do programa em relação ao parâmetro: espaçamento entre as grades

Do ponto de vista quantitativo, devem-se sempre analisar com cautela as limitações que uma ferramenta computacional pode apresentar. O KVFinder com as rotinas de potenciais implementadas será, a partir de então, referido neste trabalho como KVFinder^{drug}. O programa utiliza uma matriz, em que o usuário pode determinar o valor do espaçamento de cada ponto da grade. Sabe-se que quanto menor este valor, maior é a acurácia dos resultados obtidos pelo programa. Em contraposição, determinar um valor muito pequeno do espaçamento de cada ponto da grade para a execução da ferramenta, pode representar um alto custo computacional.

Os testes que foram realizados analisaram as seguintes características: (i) área total (calculada pelo KVFinder^{drug}), (ii) volume, potenciais: (iii) lipofílico, (iv) densidade de acceptor e (v) doador de hidrogênio. Os testes foram feitos utilizando o computador Intel Core, com 8GB de memória e sistema operacional Scientific Linux 6.5 e usando os parâmetros 1,0, 0,8, 0,5 e 0,2 Å/pts para o intervalo entre os pontos da grade. O experimento ocorreu para o cálculo de todas as cavidades das proteínas e destacou sete macromoléculas: 2nyr(93), 2ocf(94), 2y00(94), 2hyy(95), 4cox(96), 4ey7(90) e 2y04(94), para saber de que forma os potenciais poderiam variar com os parâmetros do programa.

A primeira análise realizada foi baseada na área total da cavidade calculada pelo KVFinder^{drug}, e de acordo com a Figura 45 a diferença entre os valores para cada parâmetro é bastante relevante. Para a proteína 4cox o valor da área para 0,2 Å/pts é de quase 30.000Å² enquanto para 0,5 Å/pts este valor não ultrapassa de 5000 Å². Por outro lado, sabe-se que uma segunda análise delimitando o espaço de busca desejável pode ser realizado. Esta análise seria viável se tivesse um conjunto pequeno de proteínas a ser estudado, uma vez que, a preparação do programa KVFinder é realizado manualmente pelo usuário. A diferença de área para valores entre 0,5 Å/pts e 0,8Å/pts também pode ser considerada grande, a cavidade escolhida é aquela que possuía a molécula ligante. A partir de 0,8 Å/pts estes valores estabilizam, no entanto, a acurácia

dos resultados não seria tão expressiva para uma simples análise. A mesma situação ocorre quanto ao volume da cavidade, mas as diferenças são bem mais visíveis para as proteínas 2nyr e 2y04, sendo que para valores de 0,2Å/pts e 0,5Å/pts a diferença fica em torno de 6000Å³ e 5000Å³, respectivamente (Figura 46).

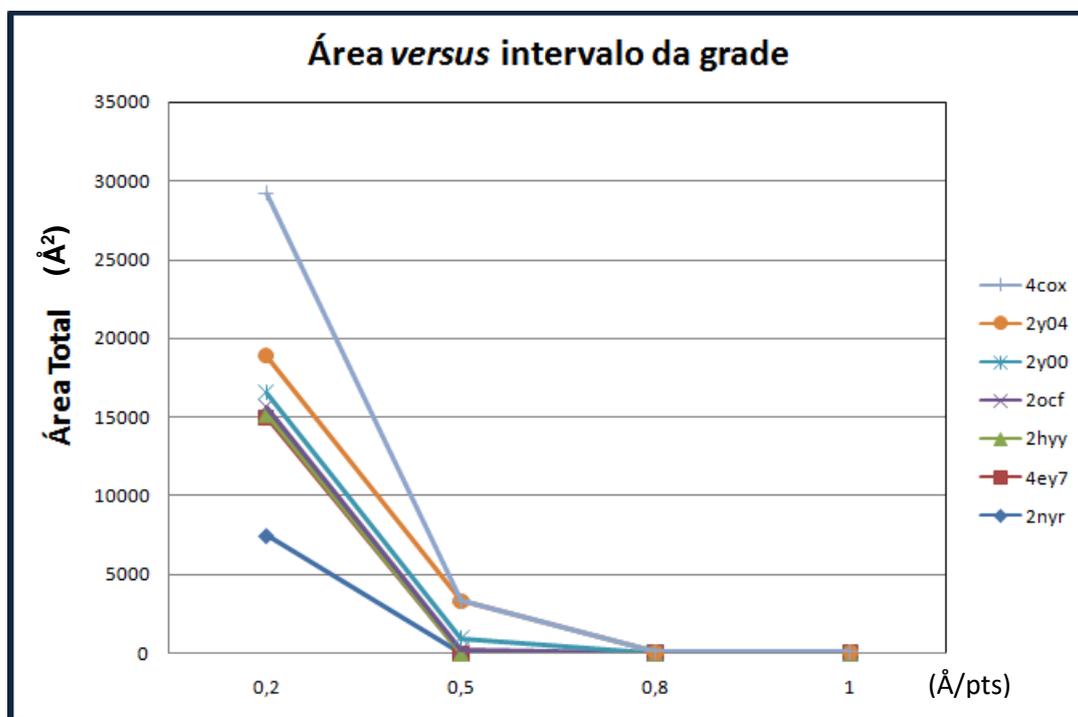


Figura 45: Variação da área total de sete cavidades para quatro valores de espaçamento da grade. O valor 0,2 apresenta maior valor de área total em relação aos outros valores. A discrepância é bem visível sendo maior para a proteína de código 2y00.

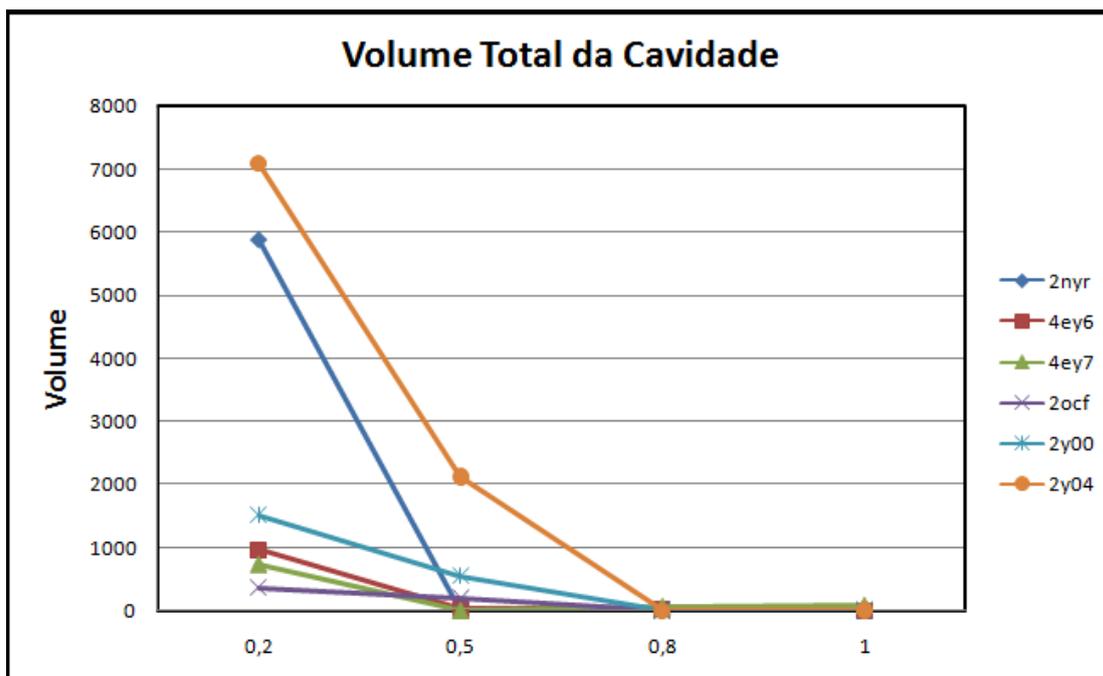


Figura 46: O volume da cavidade para seis proteínas diferentes. O volume segue o mesmo comportamento da área, sendo 2nyr e 2y04 as macromoléculas com maiores discrepâncias.

A segunda parte consiste na análise da variação dos potenciais: lipofílico e de densidade de hidrogênio. A análise do potencial lipofílico apresentou algumas particularidades (Figura 47). Por exemplo, a proteína 2ocf possui um maior potencial lipofílico quando a unidade da grade é $0,5\text{\AA}/\text{pts}$ em relação à de $0,2\text{\AA}/\text{pts}$. A proteína 4ey7 apresentou um potencial negativo para $0,2\text{\AA}/\text{pts}$, e demonstrou um leve aumento para os outros parâmetros. Por outro lado, a proteína 2y00 praticamente não obteve diferença nos resultados, relacionada aos parâmetros, $0,2\text{\AA}/\text{pts}$ e $0,5\text{\AA}/\text{pts}$. A proteína 2hyy mostrou maior discrepância entre os valores quando aumenta o espaçamento da grade. Para valores acima de $0,5\text{\AA}/\text{pts}$ os potenciais tendem a diminuir. Exceto para as proteínas 2y04 e 2y00, o volume encontrado é próximo de zero, o que indica que não é possível detectar uma cavidade a partir de $0,5\text{\AA}/\text{pts}$.

Na avaliação da proteína 2ocf quanto à hidrofobicidade, Figura 48, observa-se que os valores praticamente não se alteram quando as distâncias entre os pontos da grade variam entre $0,2\text{\AA}/\text{pts}$ e $0,5\text{\AA}/\text{pts}$. A proteína 4ey7 sofre pequena alteração quando aumentamos os parâmetros, sendo que a tendência

da cavidade é ser mais hidrofóbica. As demais proteínas tendem a sofrer sensíveis diferenças dos valores hidrofóbicos apresentando um comportamento similar entre si. A hidrofobicidade tende a diminuir com o aumento do intervalo da grade. A hidroflicidade, Figura 49, seguiu o mesmo comportamento de hidrofobicidade, sendo que quanto menor o valor de intervalo da grade mais negativa é a quantidade.

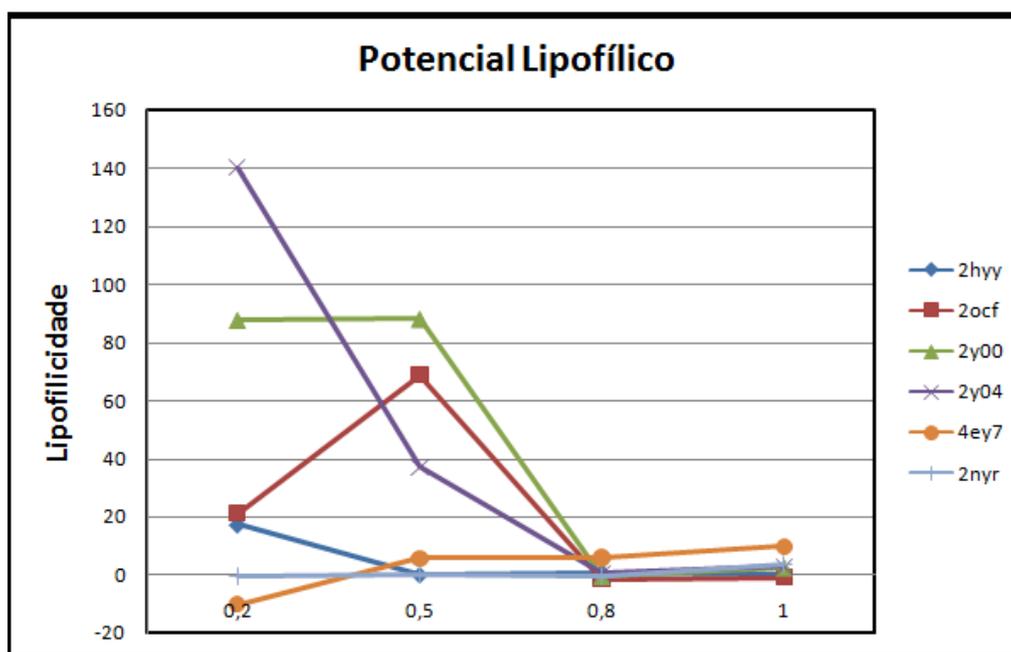


Figura 47: O potencial lipofílico total das cavidades de seis proteínas com o ligante. Os valores de lipoflicidade apresentam discrepâncias para os valores de intervalos de pontos da grade entre 0,2 e 0,5, no entanto, para 4ey7a cavidade varia de hidrofílica para hidrofóbica. Para 0,8 e 1,0 não há diferença significativa se aproximando do zero.

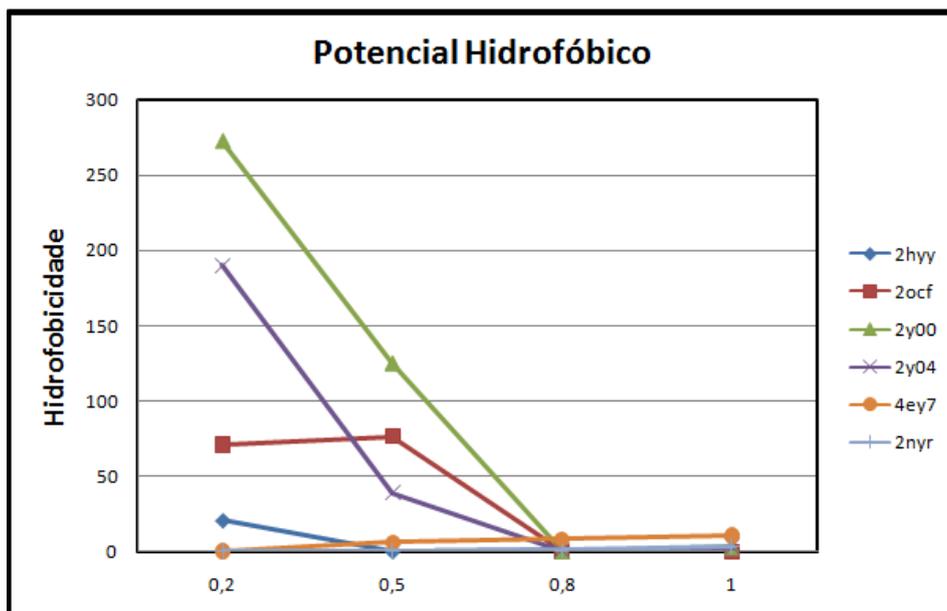


Figura 48: Hidrofobicidade da cavidade para cada valor da grade. A hidrofobicidade segue um comportamento semelhante à área e ao volume. A diferença pode ser grande como ocorre com 2y00 ou quase imperceptível como 2ocf. No geral estes valores tendem a diminuir.

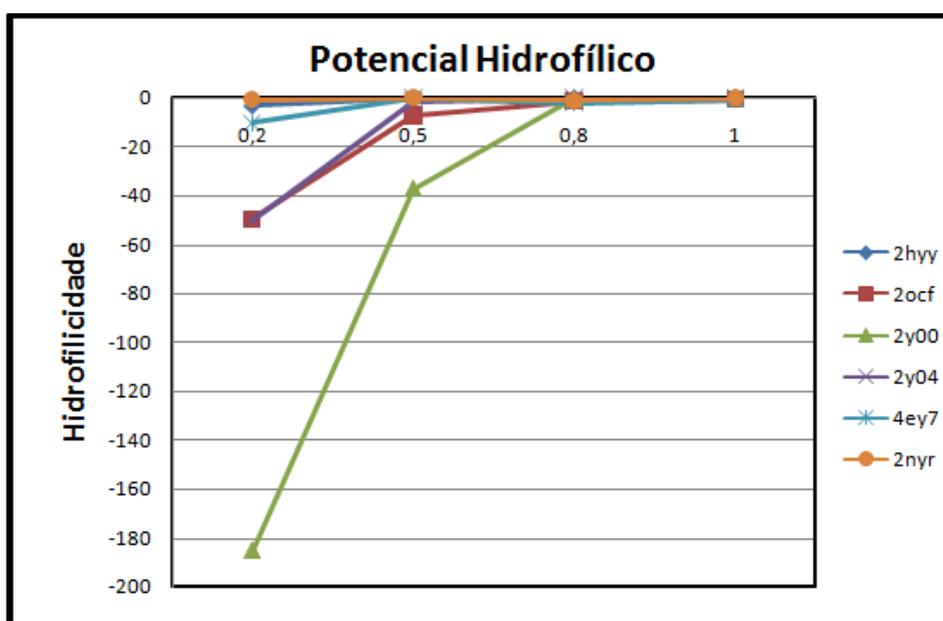


Figura 49: Hidrofilicidade para cada cavidade onde o ligante está presente. Valor de hidrofilicidade para todas as proteínas tendem a aumentar com o tamanho da grade, sendo a proteína 2y00 a que apresenta a maior diferença de hidrofilicidade.

5.3.1 Eficiência do programa para a execução de grande número de proteínas

Um dos problemas enfrentados na execução do trabalho foi relativo ao tempo de execução do programa. Como a ferramenta KVFinder^{drug} trabalha com um preenchimento de uma matriz booleana, encontrar todas as cavidades exigiu um tempo maior que o esperado. O trabalho foi realizado com a primeira versão do programa. O tempo de execução de cada macromolécula depende do seu tamanho, e aquelas que apresentaram maior número de átomos requisitaram mais tempo de execução do programa. O maior tempo de CPU registrado ficou em torno de 14h para a proteína 4O2B(97). O menor tempo registrado foi para a proteína 1FKF(98) com aproximadamente 35 minutos. Os algoritmos implementados resultaram em um tempo de execução bem menor que a busca de cavidades. Os maiores tempos registrados para o cálculo de densidade de hidrogênio e potencial lipofílico foram 1h e 13min e 1h e 02min, respectivamente. Enquanto o menor tempo de execução não passou de alguns segundos. As macromoléculas apresentadas foram 4OB2 (nitrila hidratada da *Pseudonocardia termofila* ligada ao ácido butaneborânico), 3TOP (subunidade C-terminal de glicoamilase-maltase humano complexo com acarbose)(99), 1PXX (ciclo-oxigenase ligada ao diclofenaco)(100), 2oth (complexo ternário da fosfolipase ligada A2 com indomethacin and nimesulide), 1PK2 (tecido tipo plasminogênio ativador complexado ao ácido 6-aminohexanóico)(101) e 1FKF (complexo imunossupressante-imunofilina). Todos os valores estão registrados na tabela 3.

Tabela 3: Tempo de execução do KVFinder^{drug} para proteínas.

Tempo de execução do KVFinder				
Proteína	Número de átomos	HB (h:min:s)	LP (h:min:s)	KVFinder (h:min:s)
4o2b	17332	1:13:46	1:02:21	12:14:41
3top	14264	0:36:56	0:31:29	10:19:35
1pxx	17894	0:14:38	0:28:51	9:45:54
2oth	944	0:00:29	0:00:18	0:47:03
1pk2	1330	0:00:07	0:00:05	0:35:04
1fkf	832	0:00:08	0:00:06	0:34:34

A Figura 50 mostra o gráfico do número de átomos em função do tempo de execução exigido para encontrar todas as cavidades de uma proteína. Os

algoritmos de Potencial Lipofílico e Densidade de Ligação de Hidrogênio levaram alguns minutos, sendo os maiores deles já mostrados na Tabela 3. Isso atesta que para um grande número de proteínas a utilização do programa se torna dispendiosa.

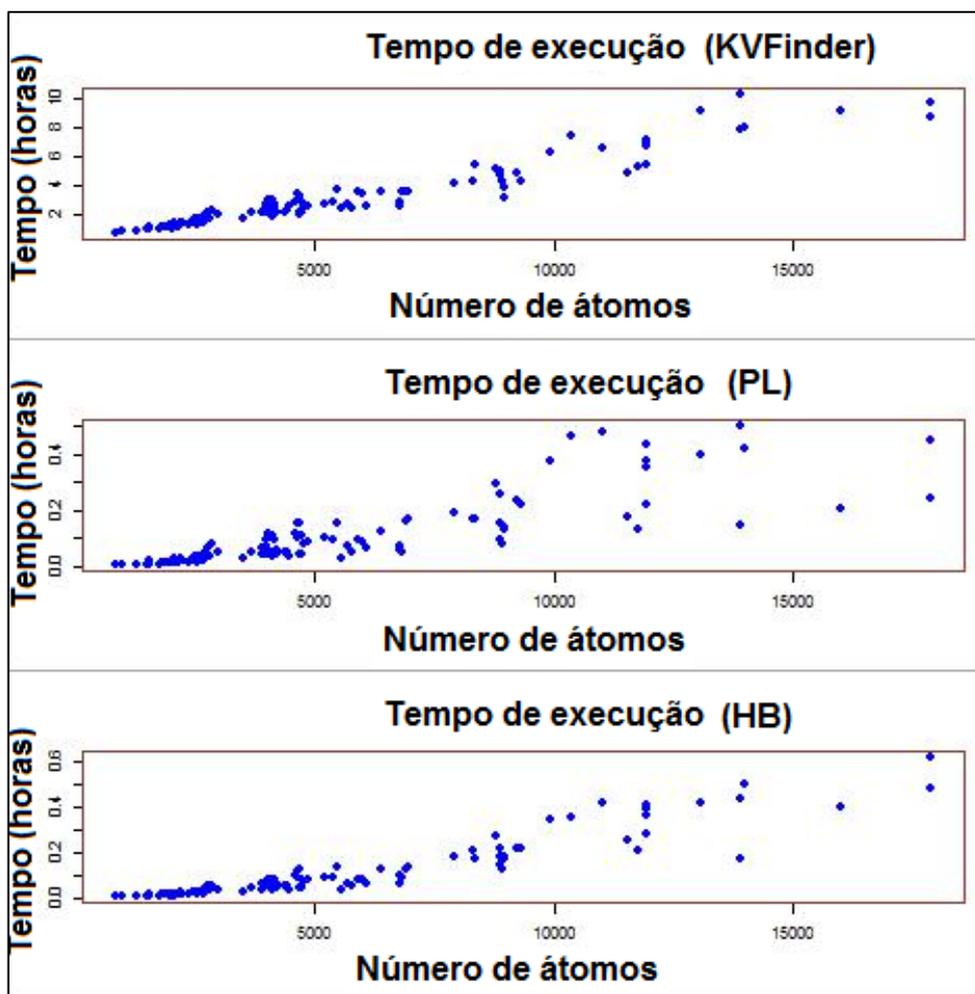


Figura 50: Tempo de execução do programa KVFinder em função do número de átomos das proteínas. Gráfico do tempo em horas para o KVFinder encontrar todas as cavidades, potencial lipofílico (PL) e densidade de hidrogênio (HB).

5.4 Análise de Drogabilidade

5.4.1 Seleção dos grupos drogáveis e não-drogáveis

O projeto visou aplicar o KVFinder^{drug} em dois grupos de proteínas, que são classificadas como drogáveis e não-drogáveis, escolhidas de acordo com que foi explicado na metodologia, seção 4.3. Os atributos utilizados para caracterizar a amostra dos dois conjuntos basearam-se no potencial isoelétrico (Figura 51), número de átomos (Figura 52) e peso molecular (Figura 53) existente nas proteínas. As comparações em relação ao peso molecular mostraram que a média das proteínas não-drogáveis é o dobro das proteínas drogáveis enquanto o potencial isoelétrico apresentou média semelhante entre os dois grupos. Estes aspectos foram escolhidos para que o conjunto controle não se distancie tanto do grupo de teste, ou seja, drogável. A média do potencial isoelétrico apresentou-se, para ambos os grupos, valor em torno de 6,65.

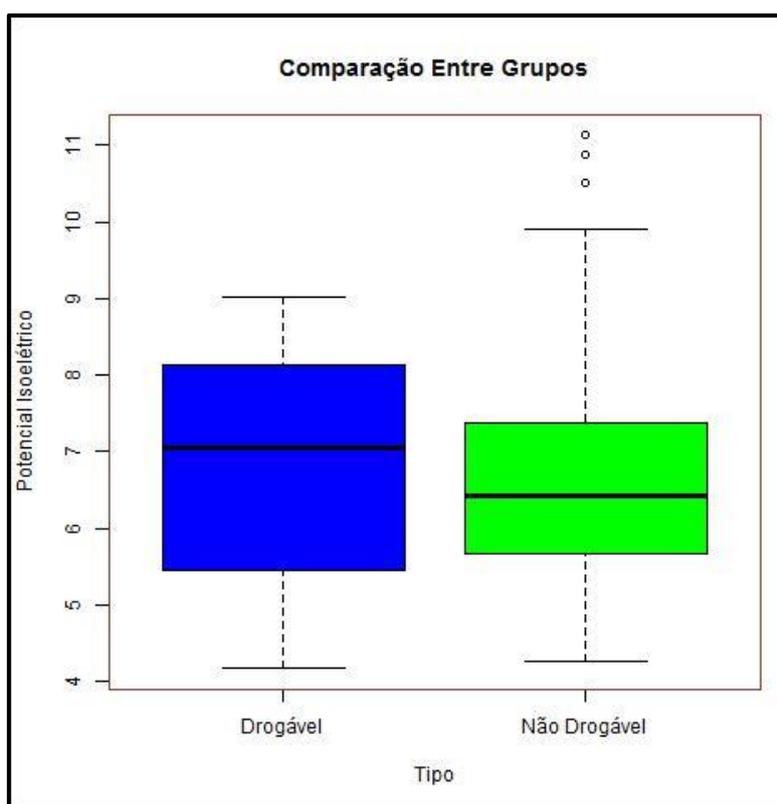


Figura 51: *Boxplot* do Potencial isoelétrico. Apesar dos dois grupos terem o valor de média próximo de 7, o grupo drogável (azul) é mais homogêneo.

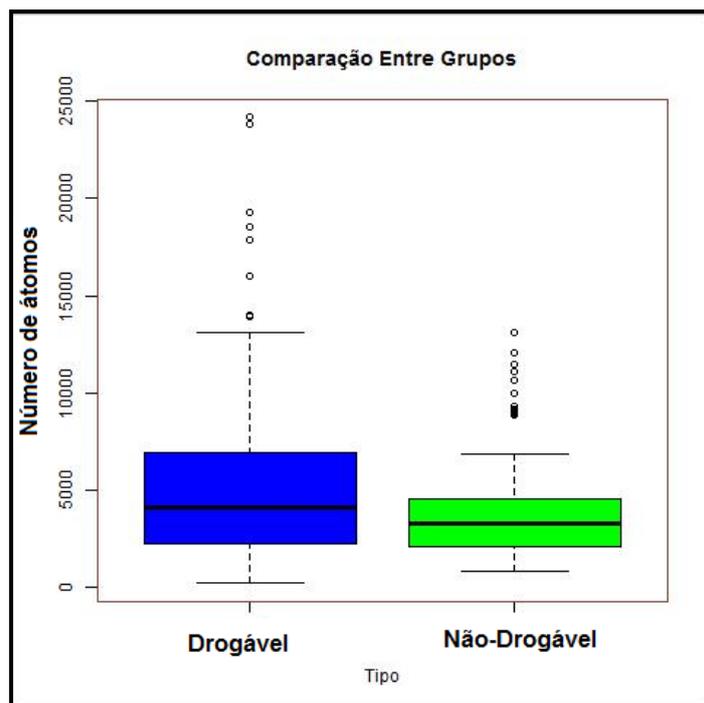


Figura 52: *Boxplot* em relação ao número de átomos das proteínas drogáveis e não drogáveis. O conjunto de proteínas drogáveis possuem maiores número de átomos e apresentam mais *outliers*.

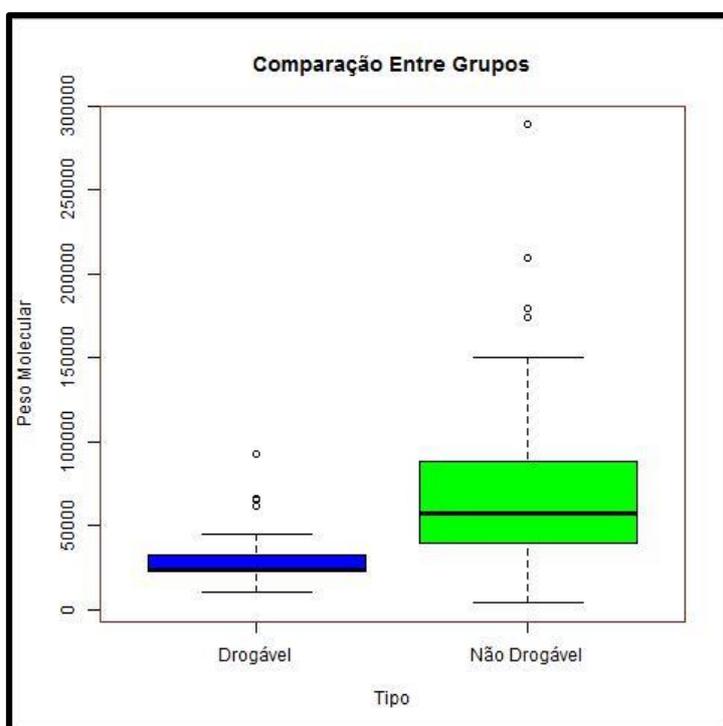


Figura 53: *Boxplot* do Peso molecular. As proteínas drogáveis possuem menor peso molecular que as proteínas não-drogáveis. As diferenças entre a mediana, o primeiro e o terceiro quartil é mais aparente para os grupos não-drogável.

5.4.2 Análise das características topológicas das cavidades

5.4.2.1 Área e volume da cavidade

As primeiras análises foram realizadas quanto às suas características topológicas, área e volume. Em relação à área, os valores são verificados quanto aos pontos que tangenciam a superfície molecular da cavidade e a área calculada pelo KVFinder.

De acordo com o *boxplot* da Figura 54 o conjunto de proteína drogável apresentou cavidades com maiores superfícies moleculares. As cavidades escolhidas para análise do conjunto não-drogável foram as que possuíam as maiores áreas da superfície molecular. Mesmo assim, o conjunto apresentou os menores valores de área de superfície que o grupo drogável.

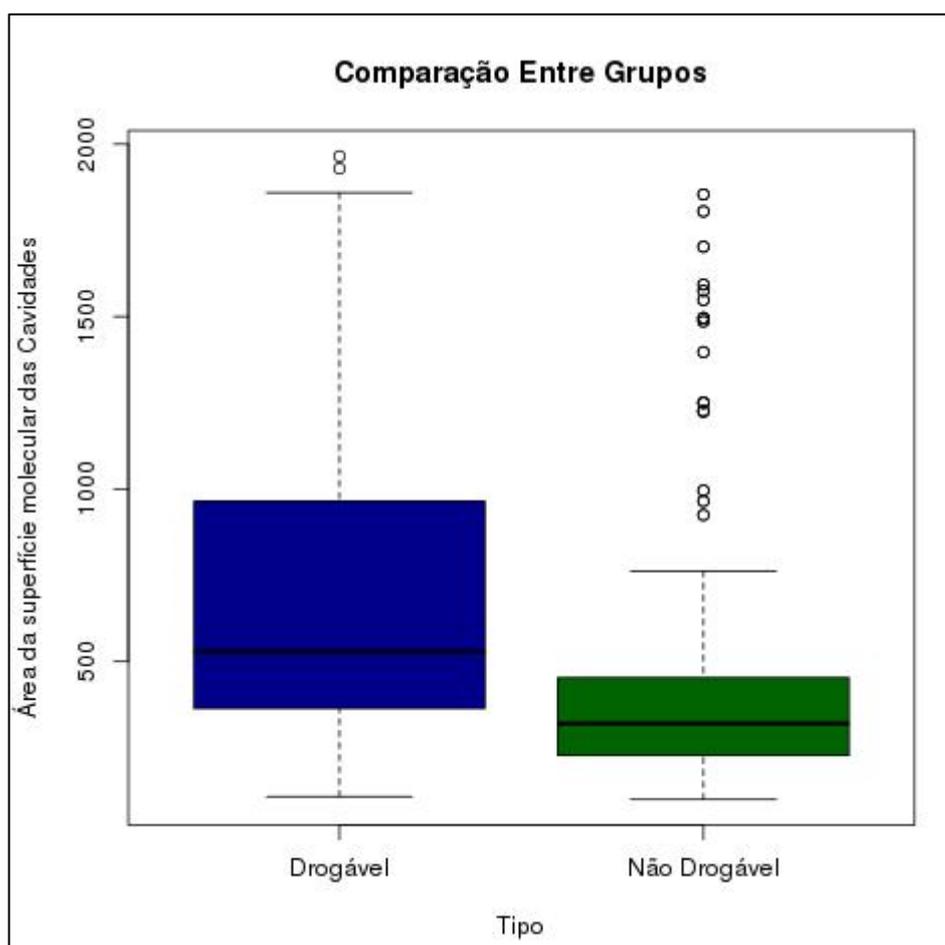


Figura 54: *Boxplot* relativo à área superficial das cavidades das duas amostras. A amostra drogável apresenta maior dispersão que a amostra não-drogável.

A média da área da superfície das cavidades não drogáveis foi de 1574\AA^2 , enquanto que para o grupo não-drogável este valor foi de 430\AA^2 . Na Figura 54 está representado o *boxplot* dos dois grupos de teste. Mesmo selecionando as maiores áreas das proteínas não-drogáveis, a amostra drogável apresenta o valor de área superficial maior e maior dispersão dos dados. Podemos supor desta forma que as cavidades que possuem maior área acessível ao solvente possuem maior capacidade de reconhecer um ligante com característica de fármaco. Entretanto, uma característica, utilizada em programas como SiteMap[®] e DogSite denominado *enclosure* (invólucro), pode possuir grande importância na análise da topologia da cavidade. O invólucro, já descrito na seção em que falamos sobre o servidor DogSite serve para definir o quão internalizado o fármaco pode se localizar na proteína.

Este descritor neste trabalho foi simplificado calculando a área da superfície molecular da cavidade da macromolécula e a área fechada que envolve todo o volume que é dada inicialmente pelo KVFinder.

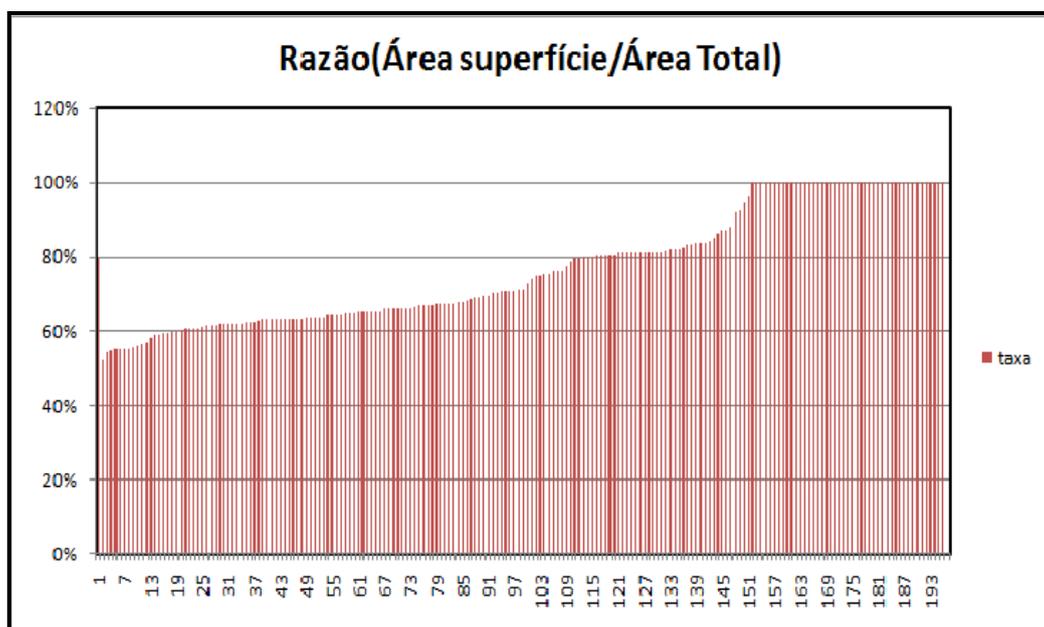


Figura 55: Gráfico da razão área superficial e área total de 193 proteínas drogáveis.

O que pode se observar no gráfico da Figura 55 é que as cavidades onde se encontram os fármacos têm sua área coberta em sua maioria acima de 60%. Este resultado concorda com valores encontrados na literatura (37,63,81). Além disso, é um importante descritor para determinar a drogabilidade de uma cavidade proteica proposta.

O volume para as duas amostras apresentou poucas diferenças. A média da amostra drogável mostrou um valor maior em relação à não-drogável. Pelo boxplot da Figura 66, podemos ver a variação de volume das cavidades das duas amostras. A amostra não-drogável apresenta maior dispersão e amostra drogável apresenta mais valores atípicos.

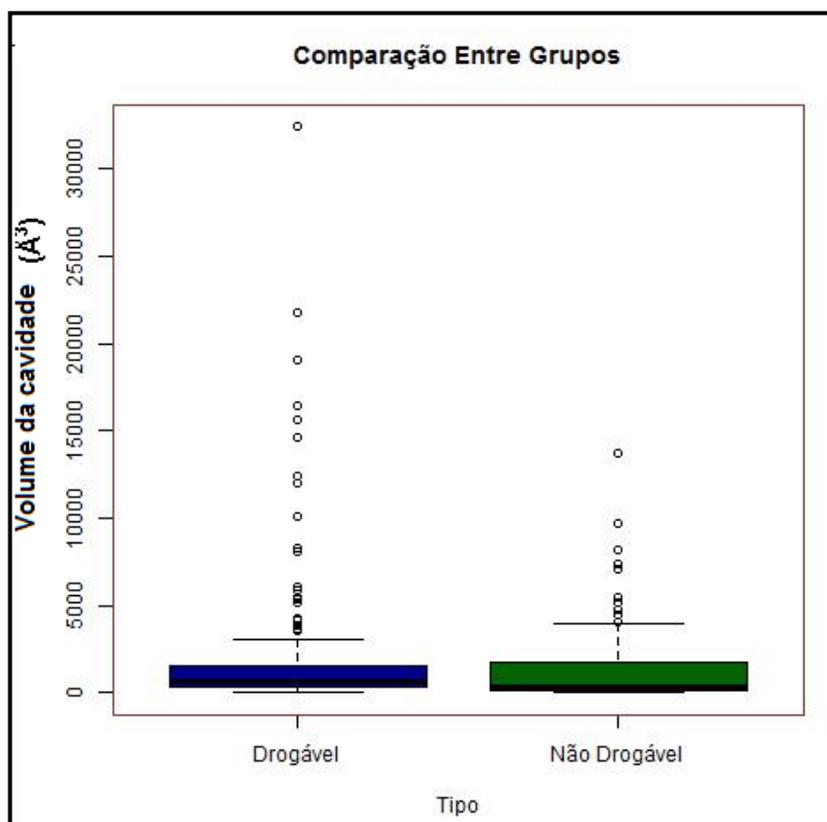


Figura 56: *Boxplot* do volume das cavidades dos grupos drogáveis e não-drogáveis. A amostra não-drogável possui menor volume de cavidade e uma dispersão maior.

5.4.3 Análise do Potencial Lipofílico

A primeira análise se baseou na distribuição do potencial lipofílico das duas amostras. O potencial lipofílico total foi calculado para as cavidades onde os fármacos estavam presentes e feita uma distribuição dos valores para saber como se apresenta a variação do potencial lipofílico total, a hidrofobicidade e hidrofiliabilidade das cavidades. A distribuição adotada foi a distribuição de Kernel que é uma forma não-paramétrica, ou seja, não assume que os dados tenham qualquer característica estrutural ou parâmetros para estimar a função de

densidade de probabilidade de uma variável aleatória. Esta estimativa está intimamente relacionada com o histograma.

O que se observa nos gráficos é que a variação da lipofilicidade é maior para as proteínas drogáveis ficando em torno de -5000 a 6000, enquanto que para o grupo não-drogáveis a variação ficando em torno de -1000 a 500. Para os grupos drogáveis os valores de hidrofobicidade como encontrado no gráfico tendem a ser mais frequentes que os valores de hidrofilicidade. Esta comparação é ressaltada no gráfico da Figura 57, observando os valores da escala encontrada dentro da elipse destacada em vermelho. Entre os valores de lipofilicidade 0 e -1000 a queda é mais abrupta, enquanto que, para os valores entre 0 e 1000 a queda é mais suave apresentando dois pequenos picos (seta em vermelho na Figura 57).

Para as proteínas não-drogáveis a relação é inversa, Figura 58. A queda mais abrupta ocorre para valores hidrofóbicos de cavidades de proteínas. Enquanto que para valores hidrofílicos, apesar de uma queda também abrupta entre 0 e -100 existe um aumento da frequência quando a hidrofilicidade atinge o valor -200 (seta vermelha), e a partir deste valor é apresentada uma curva mais suave.

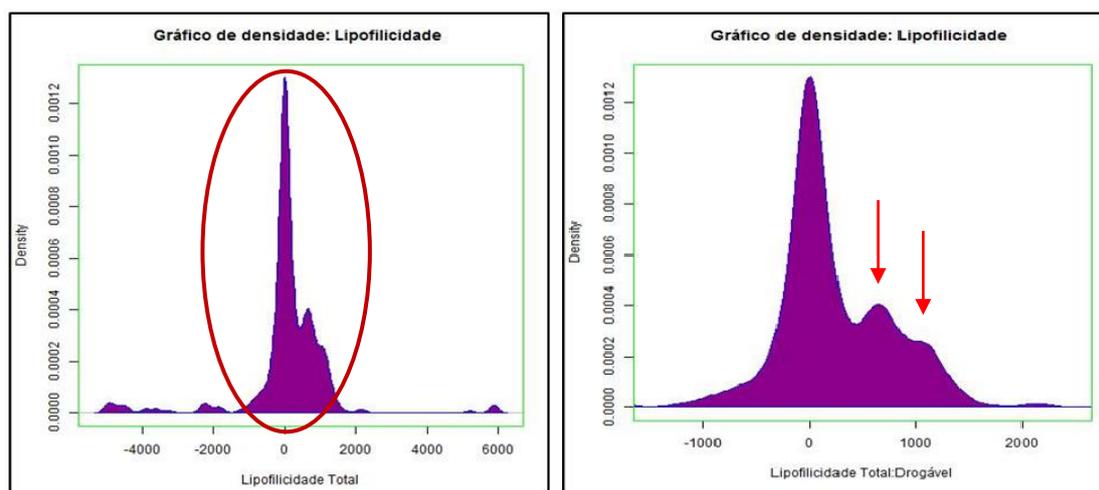


Figura 57: Gráficos da densidade de Kernel para o Potencial Lipofílico para as proteínas não drogáveis. No gráfico da esquerda, a região destacada em vermelho se concentra o maior número de cavidades. No gráfico da direita, a região com maior densidade foi ampliada. Pode-se ver que a maioria das cavidades tendem a ser hidrofóbicas. A região com valores positivos possui variação maior de densidade (setas em vermelho).

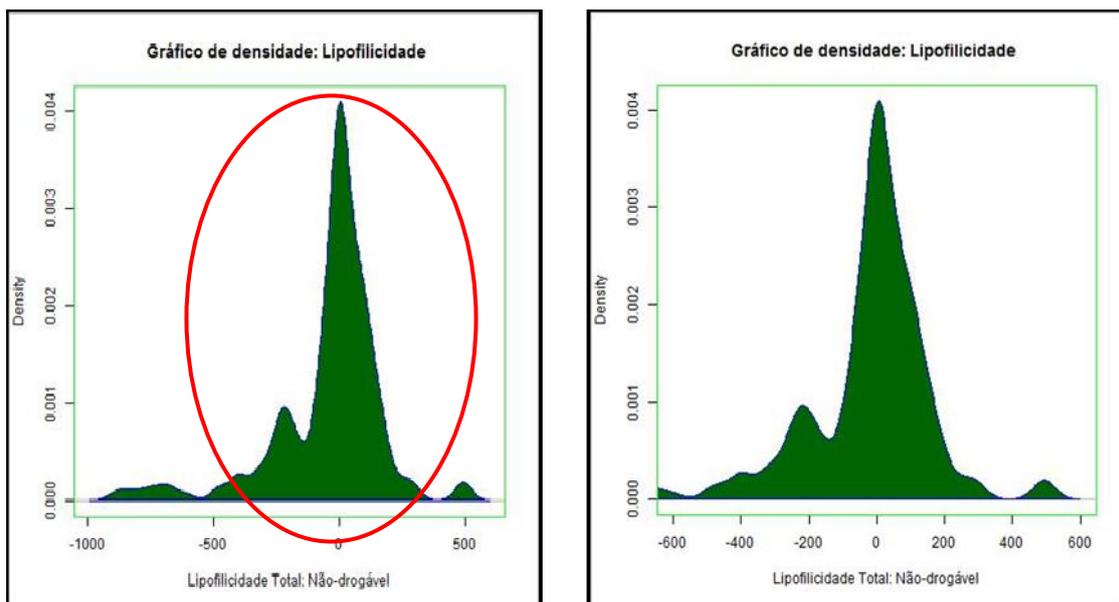


Figura 58: Gráficos do Potencial Lipofílico para o grupo não-drogável. A variação do potencial é menor que para o grupo drogável. O gráfico da esquerda a região onde há maior número de cavidades (círculo em vermelho) foi ampliada e apresentada no gráfico da direita. Os grupos não-drogáveis tendem a apresentar cavidades mais hidrofílicas.

De acordo com dados na literatura, as cavidades dos alvos drogáveis tendem a aparecer mais hidrofóbicas que hidrofílicas. Este fato, considerando a afinidade entre um ligante e uma macromolécula, concorda com uma das regras de Lipinski sobre a lipofilicidade do fármaco. As drogas com biodisponibilidade oral devem ser passíveis de atravessar as bicamadas lipídicas e atingir os alvos desejáveis. Para uma melhor visualização dos resultados apresentamos um *boxplot*, Figura 59, para observar a distribuição empírica dos dois grupos. Observamos que no conjunto drogável a dispersão dos dados é maior assim como o potencial lipofílico. O grupo não-drogável possui menor dispersão e menor valor de potencial lipofílico. Alguns valores atípicos desta amostra mostram a tendência de um menor potencial lipofílico.

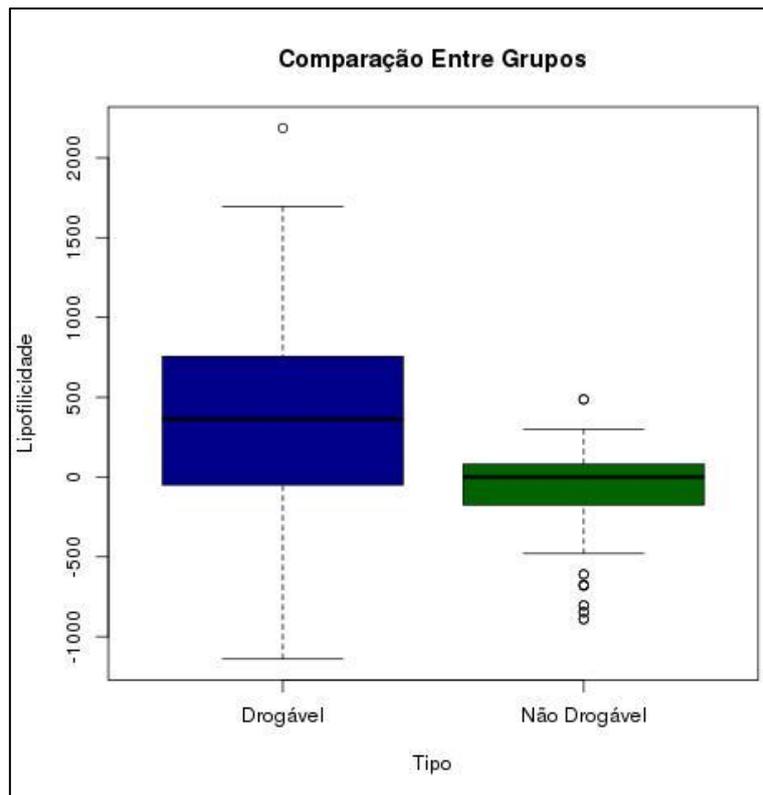


Figura 59: Lipofilicidade apresentada pelas amostras drogáveis e não-drogáveis.

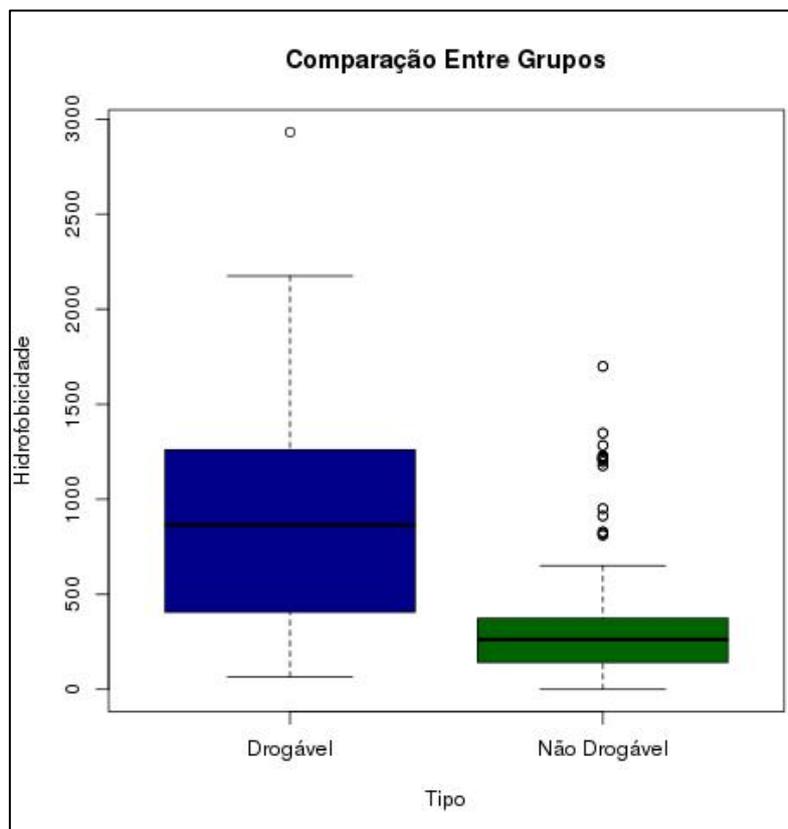


Figura 60: Comparação entre grupos drogáveis e não-drogáveis quanto à hidrofobicidade. O grupo drogável apresenta maior hidrofobicidade, enquanto que o grupo não-drogável apresenta menor dispersão e valores atípicos em sua amostra.

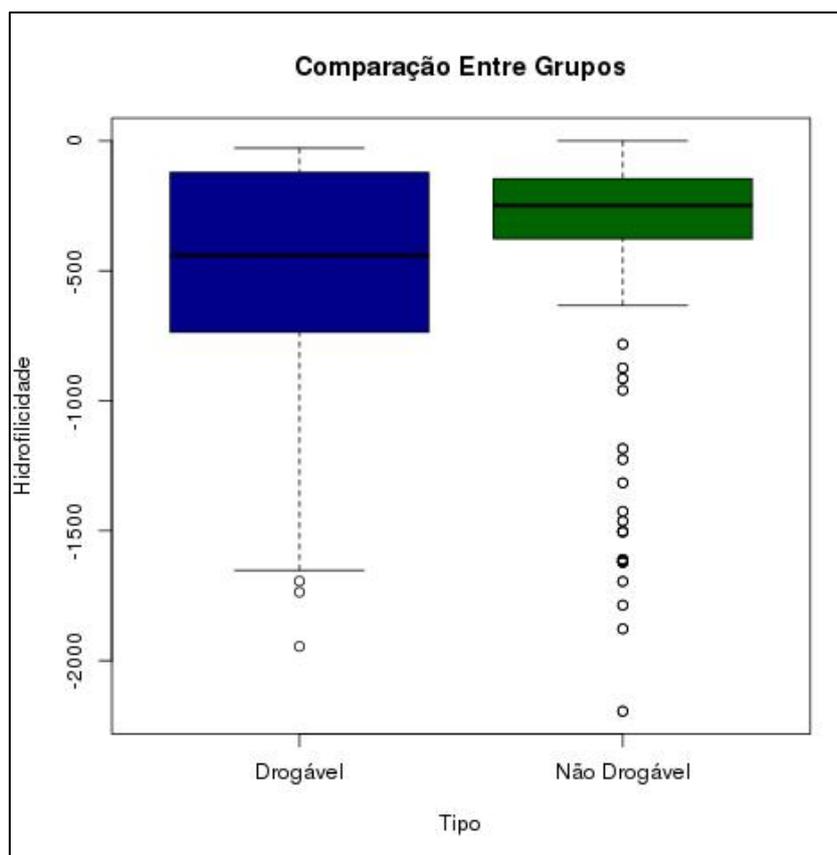


Figura 61: Relação entre a hidrofiliidade de cada grupo. O grupo drogável apresenta maior dispersão, no entanto, menor hidrofiliidade que o grupo não-drogável.

A mesma análise foi realizada utilizando os valores dos potenciais hidrofóbicos e hidrofílicos separadamente. Os padrões encontrados são semelhantes, assim como os valores discrepantes para as proteínas não-drogáveis e uma menor variação para cavidades hidrofóbicas e hidrofílicas dos alvos.

Neste caso, estamos analisando um conjunto grande de cavidades com áreas e volumes diferentes. Para uma avaliação mais precisa da influência da lipofiliidade na drogabilidade de um alvo combinamos esta quantidade ao valor de área, para se ter uma proporção de quão lipofílica pode ser uma cavidade. A rotina adicionada ao KVFinder calcula o valor da área da superfície molecular que se apresenta hidrofóbica ou hidrofílica. Estes valores foram apresentados através de dois *boxplot* que mostram uma comparação entre as áreas hidrofílicas e hidrofóbicas, Figuras 62 e 63.

Podemos observar a dispersão dos valores das áreas hidrofóbicas e hidrofílicas. Isso se deve ao tamanho e forma da cavidade. Devido o número da amostra ser pequeno (199 proteínas), os conjuntos de testes variam consideravelmente para os descritores das cavidades.

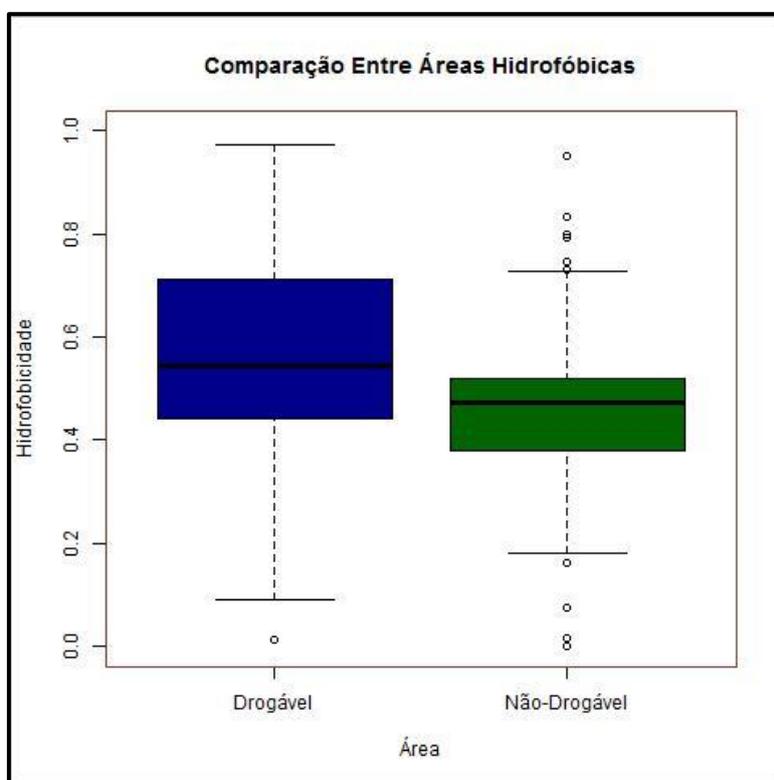


Figura 62: *Boxplot* da área hidrofóbica entre as duas amostras.

Na Figura 62 é visível que as proteínas drogáveis apresentam uma área hidrofóbica maior que as proteínas não drogáveis. A média fica em torno de 0,55, ou seja, mais de 50% da área das cavidades onde os fármacos se ligam são hidrofóbicas enquanto que os valores escolhidos das cavidades das proteínas não drogáveis a média se apresenta em torno de 44%. Na literatura(82), as cavidades das proteínas drogáveis apresentam uma área hidrofóbica maior que 60%. Esta discrepância quantos aos resultados aqui apresentados pode estar relacionada ao número de proteínas utilizado para amostragem que é bem inferior ao número de proteínas que foram publicadas, aproximadamente 103.400 macromoléculas.

O mesmo procedimento foi realizado para os percentuais de área hidrofílica, Figura 63, onde a dispersão dos dados é maior para as proteínas drogáveis e a média apresentada é menor. As proteínas não-drogáveis apresentam maior

número de áreas hidrofílicas indicando que esta característica é mais acentuada para este conjunto.

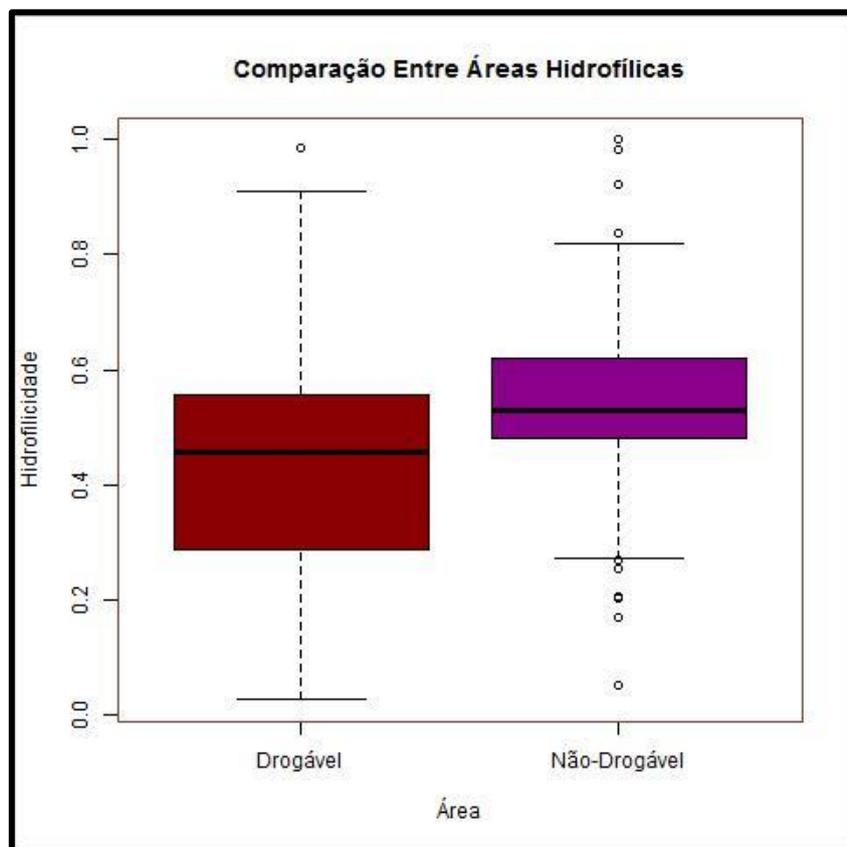


Figura 63: Relação entre as áreas hidrofílicas de cada amostra.

O comportamento hidrofóbico e hidrofílico das cavidades em função da área foi analisado e o resultado está representado na Figura 64 e Figura 65. Podemos observar uma dependência linear dos dois descritores em relação à área da superfície molecular. As cavidades com área entre 100Å^2 e 2000Å^2 foram utilizadas para o cálculo da regressão linear para a hidrofobicidade e hidrofiliabilidade. Essa restrição foi feita para evitar valores atípicos.

Utilizando o software R foi possível realizar uma regressão linear e obter os valores de uma curva para os gráficos que representam as quantidades relativas à lipofiliabilidade. O valor do R^2 encontrado para o ajuste da hidrofobicidade das cavidades foi 0,59, enquanto que para a hidrofiliabilidade foi de 0,63. Estes valores não são satisfatórios para o ajuste linear destas características.

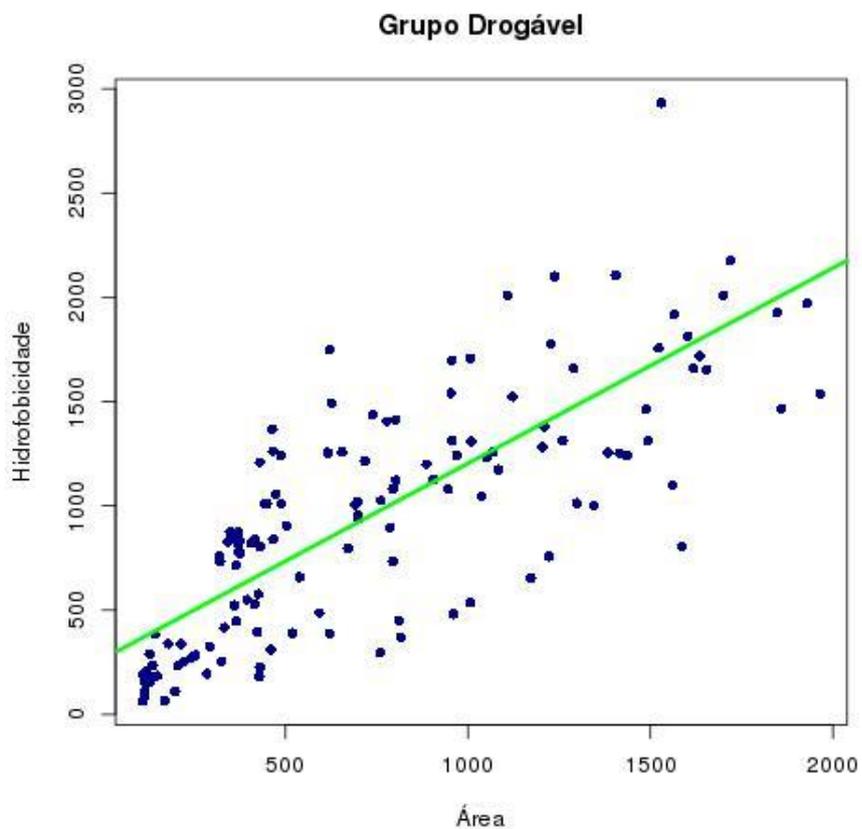


Figura 64: Relação entre o valor de hidrofobicidade e área total da cavidade. A reta em verde apresenta o cálculo da regressão linear calculada através do software R.

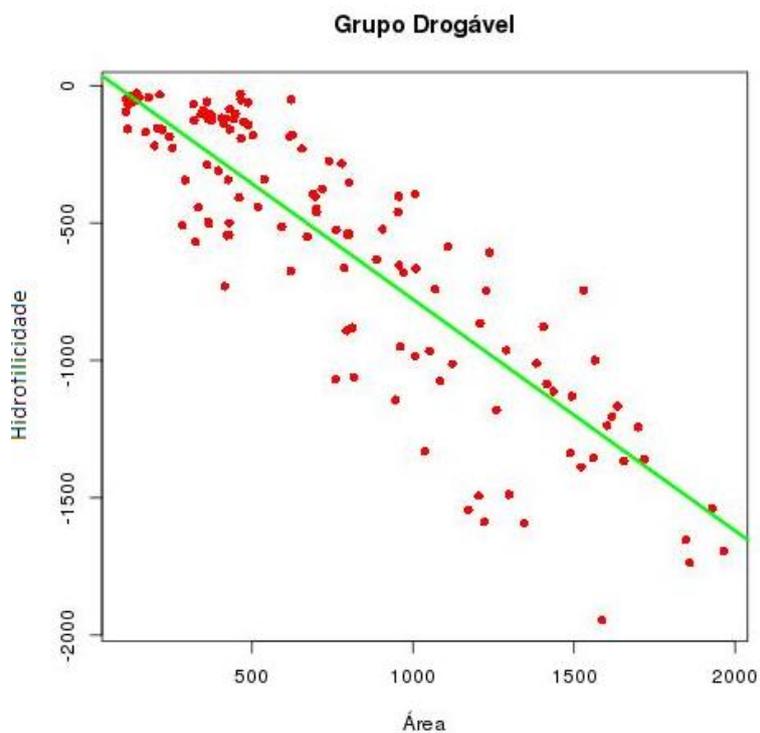


Figura 65: Regressão linear dos valores de hidrofiliidade calculados em relação a área total da cavidade.

A segunda análise está relacionada à área dos descritores de lipofilicidade que ocupa em relação à área total de cada cavidade. As disposições dos pontos também apresentam uma relação de linearização dos descritores e área total. Assim sendo, foi realizado novamente um ajuste linear destes descritores. Os valores encontrados para R^2 foram 0,86 e 0,87 para os ajustes das áreas hidrofóbica e hidrofílicas, respectivamente. Estes valores indicam que as áreas de hidrofobicidade e hidrofilicidade podem ser um bom preditor de cavidades drogáveis.

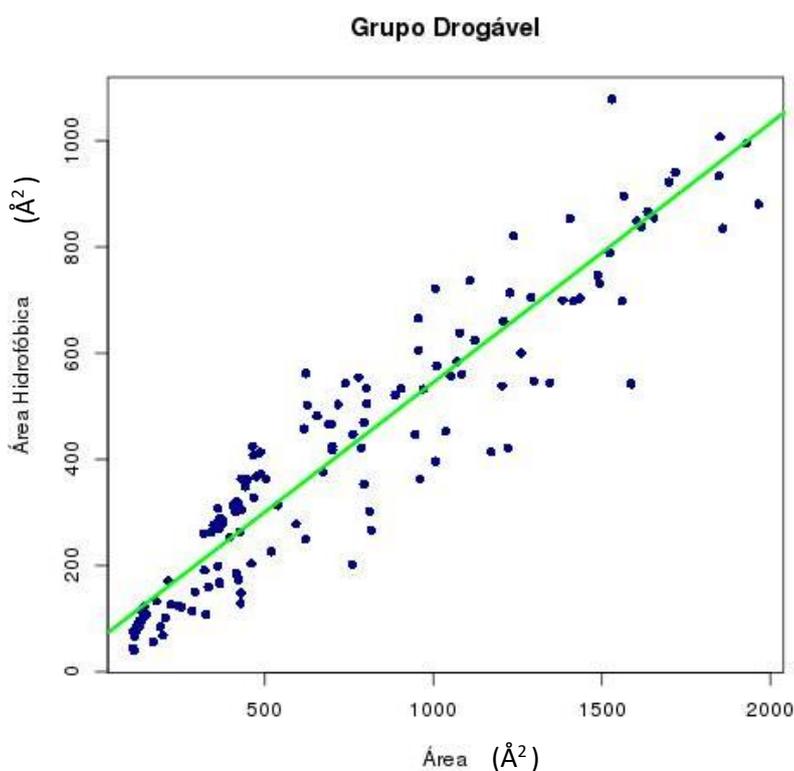


Figura 66: Regressão linear do descritor "área hidrofóbica" em relação a área superficial total

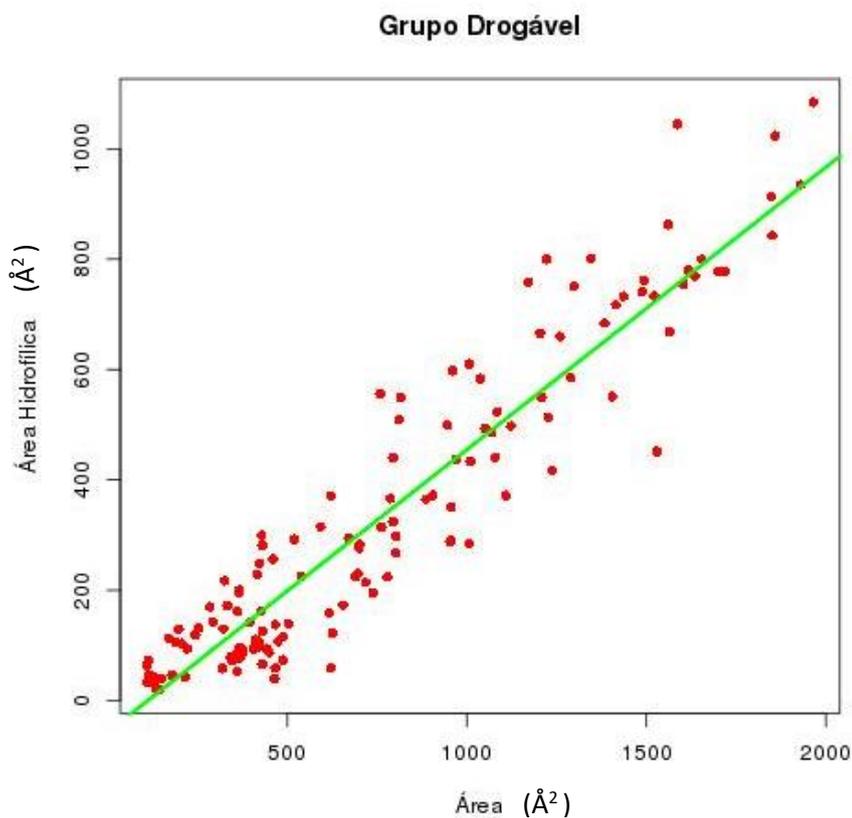


Figura 67: Regressão linear da área hidrofílica em relação à área superficial total.

Com o software R encontramos a razão entre a área hidrofóbica o vínculo desta quantidade com a área total. A Figura 68, o primeiro gráfico inclui todas as cavidades drogáveis, enquanto o segundo gráfico somente as cavidades com área menor que 5000Å^2 foram analisadas. O que se vê é que esta quantidade é maior que 40% para a maioria das cavidades.

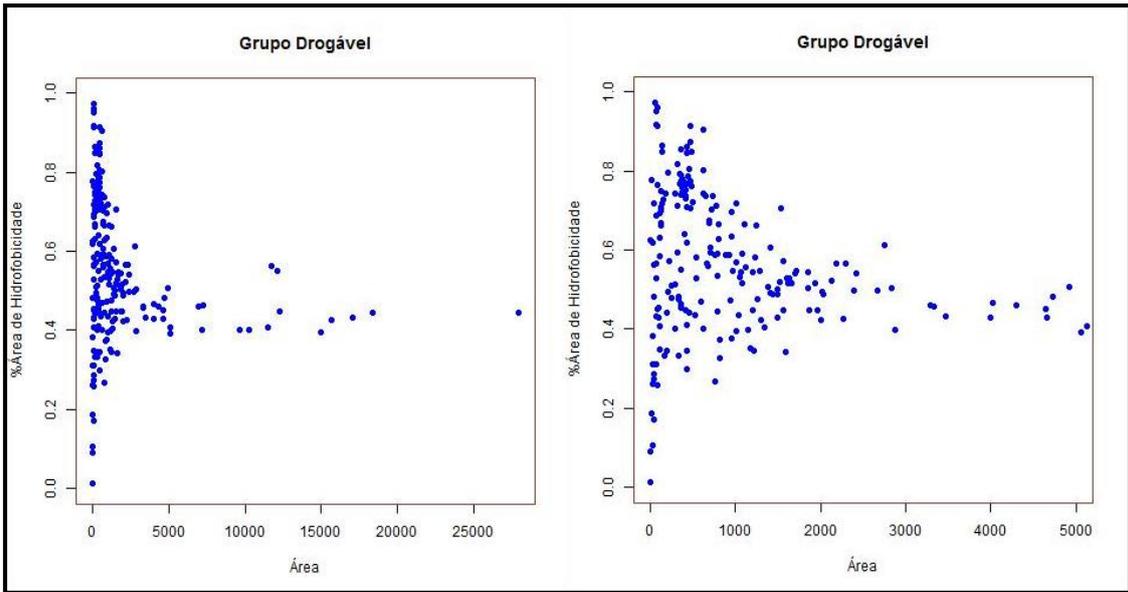


Figura 68: Gráfico da razão da área hidrofóbica das amostras drogáveis em relação à área total da cavidade.

5.4.4 Análise da Densidade de Ligação de Hidrogênio

A densidade de ligação de hidrogênio para doadores e aceptores não apresentou muitas diferenças em relação aos dois conjuntos, Figura 70. A densidade de doadores de hidrogênio apresenta-se um pouco maior para o grupo não drogável. A dispersão de densidade de doares e aceptores é grande para as duas amostras.

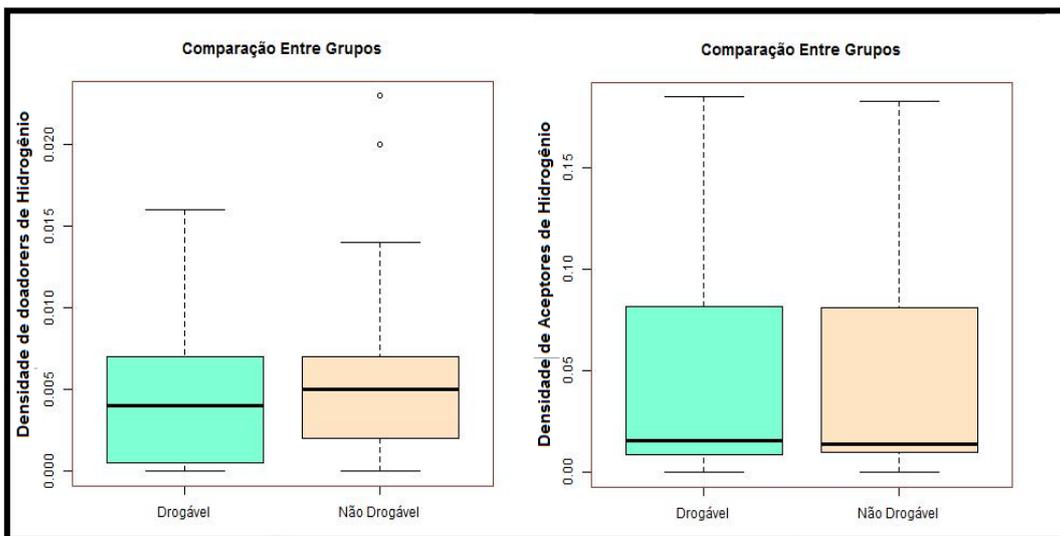


Figura 69: Boxplot para densidade de aceptores e doadores de hidrogênio entre os dois grupos. Ambos grupos apresentam um perfil semelhante, no entanto, pode-se inferir

pelo gráfico da esquerda que o grupo não-drogável apresenta média de densidade de doadores de hidrogênios um pouco superior.

As rotinas adicionadas que dão informações sobre as ligações de hidrogênio também fornecem o número de doadores e aceptores com proximidade de 4Å de cada cavidade, onde para cada quantidade realizamos uma distribuição de Kernel, Figura 71. O conjunto drogável possui a maior frequência de cavidades com número de aceptores presentes. Para o conjunto não-drogável é possível ver que a maior frequência de acceptor possui número absoluto menor que 100, apesar da curva ter uma largura menor e um decaimento mais brusco.

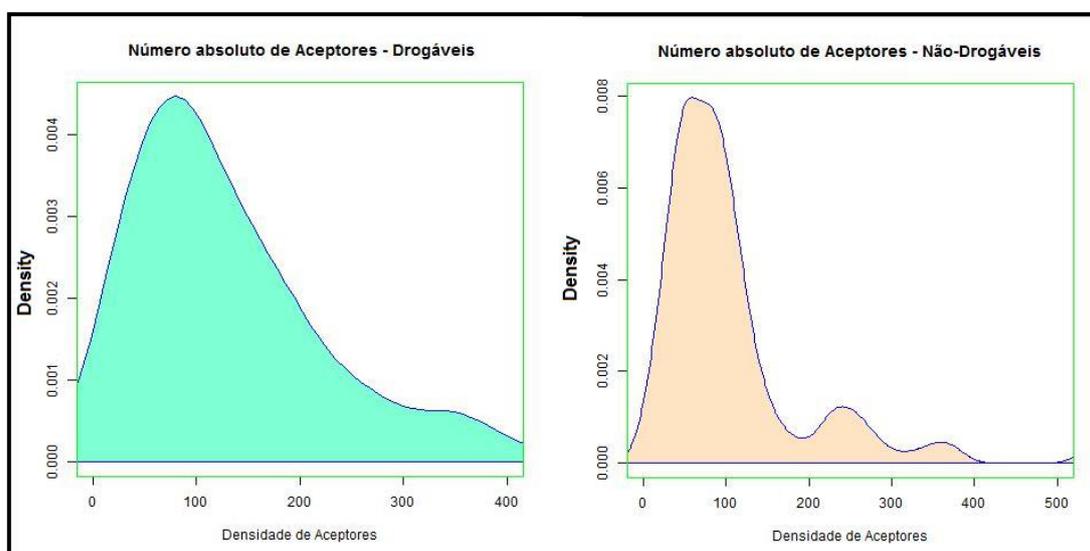


Figura 70: Distribuição de Kernel para número absoluto de aceptores para os conjuntos drogável e não-drogável.

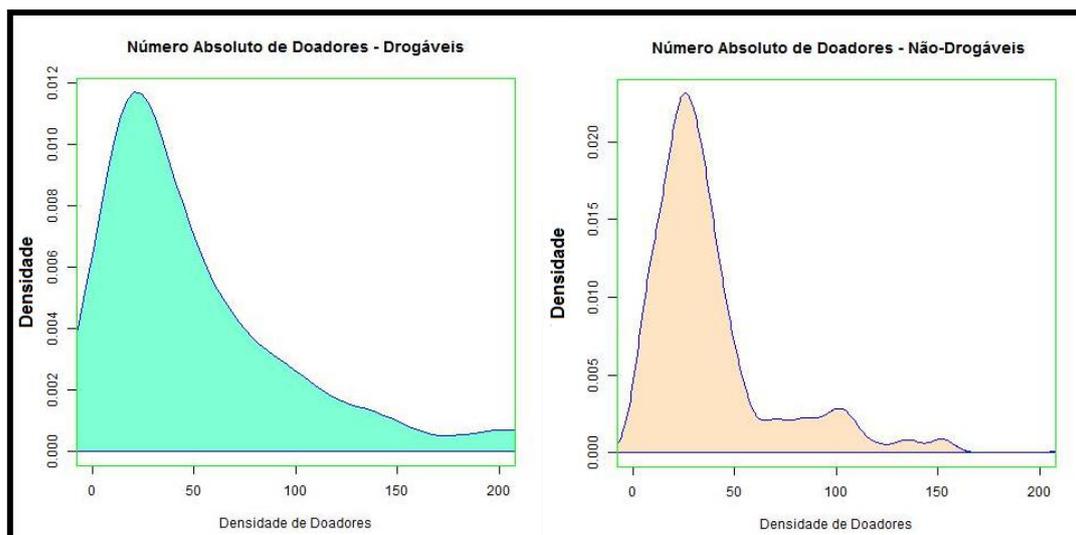


Figura 71: Densidade de Kernel para número absoluto de doadores de hidrogênio para os dois grupos. A curva que representa o grupo drogável se mostra mais suave e possui maior número de doadores.

Na Figura 72 vemos que o número de doadores de hidrogênio para a maioria das cavidades do conjunto não-drogável possui quantidade menor que 50. O comportamento da distribuição de aceptores e doadores para os dois grupos é semelhante. Isto pode ser visto também através dos gráficos da Figura 81 onde há o número absoluto de doadores e aceptores de hidrogênio.

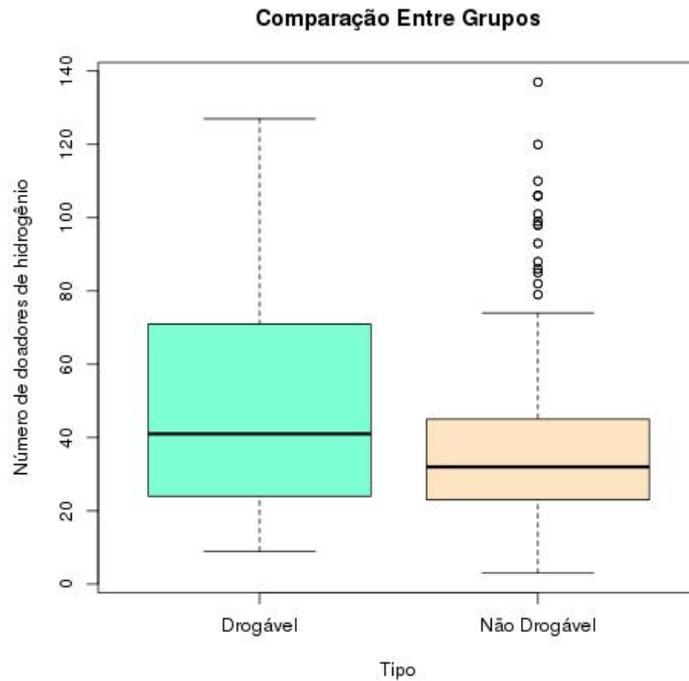


Figura 72: *Boxplot* do número absoluto de doadores de hidrogênio para as duas amostras. A amostra não-drogável apresenta menor dispersão e menor média de doadores de hidrogênio.

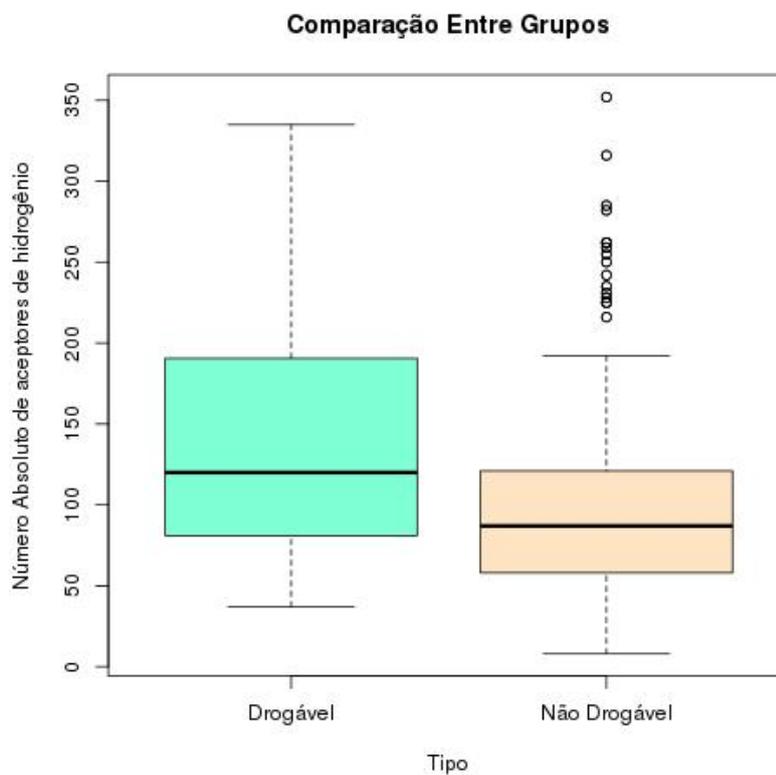


Figura 73: Comparação entre número absoluto de aceptores de hidrogênio. A amostra drogável apresenta maior dispersão e maior média do número absoluto de aceptores de hidrogênio.

Pode se observar que o conjunto drogável apresenta maior dispersão e maior número absoluto de doadores e aceptores de hidrogênio. O grupo não-drogável apresenta alguns valores atípicos. Esse comportamento pode ser explicado pela maneira como cada quantidade é obtida, a densidade de aceptores e doadores pode não englobar nos cálculos, como mostrado na metodologia, todo o volume do átomo, mas somente parte dele, enquanto que ao computar os números de doadores e receptores é realizada uma soma sobre os valores absolutos de cada quantidade.

A análise sobre a densidade de doadores e aceptores em função da área, Figura 75, não apresentou nenhum padrão que pudesse ser ajustado por uma função. Pode-se ver que os valores de densidade não apresentam nenhuma dependência singular com a área. Uma área superficial grande ou pequena pode apresentar valores aleatórios. Na figura da densidade de aceptores o que podemos ver é uma “quantização” dos valores. Grupos que apresentam um pouco acima de zero e grupos que apresentam valores próximos a 0,1. Esta afirmação vale somente para estes grupos testes. Para o grupo drogável esta densidade é bem distribuída assim como na densidade de doadores de hidrogênio.

A relação do número absoluto de aceptores e doadores apresenta um gráfico linear em função da área sendo que para o grupo drogável esta linearidade é melhor definida, Figura 76. O grupo não-drogável apresenta qualitativamente uma curva sutil nos seus valores principalmente para o grupo de doadores de hidrogênio.

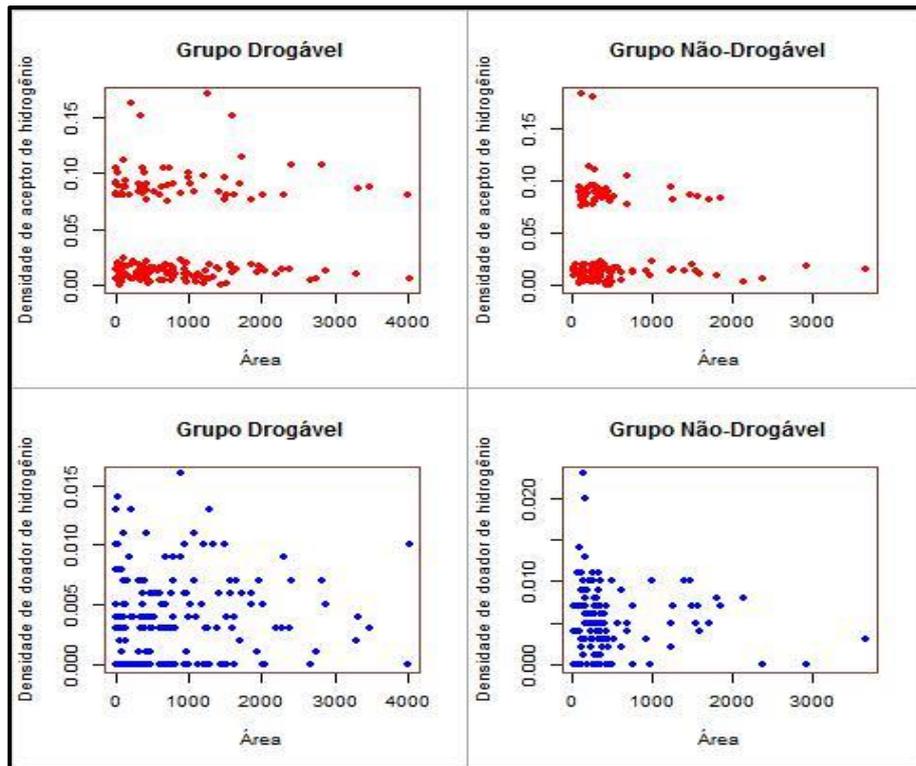


Figura 74: Densidade de aceptores e doadores de hidrogênio para os grupos drogáveis e não drogáveis. As características dos gráficos mostram que não é possível ajustar uma função que descreve uma dependência entre os descritores.

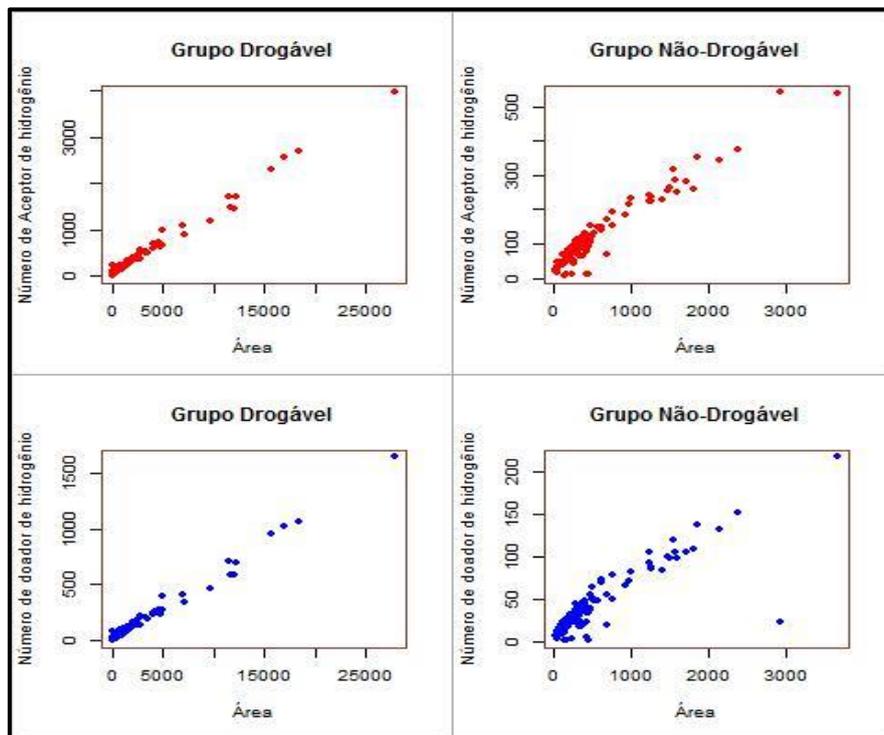


Figura 75: Relação de linearidade para o número de aceptores e doadores de hidrogênio para o grupo drogável.

A próxima análise foi realizada quanto à área que os valores de densidade de aceptores e doadores de hidrogênio apresentam para os dois grupos, Figura 77. Pode-se afirmar que as proteínas drogáveis possuem maior média de aceptores e doadores de hidrogênio e maior área ocupada por esses dois descritores. Podemos ratificar, com os resultados encontrados, a importância da interação entre aceptores e doadores de hidrogênio para o reconhecimento do fármaco pela proteína.

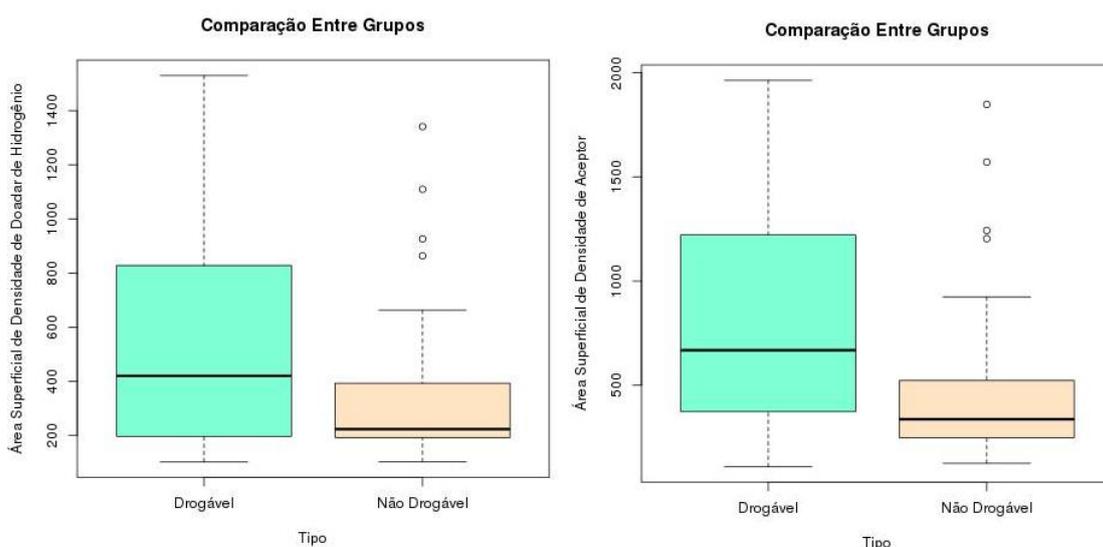


Figura 76: Relação entre a área de doadores e aceptores de hidrogênio para grupos drogáveis e não drogáveis. O grupo drogável apresentou para as duas amostras maior dispersão e maior área de doadores e aceptores de hidrogênio.

Posteriormente, foi realizada uma análise em relação à área de densidade de ligação de hidrogênio e a área superficial total. Os gráficos das Figuras 78 e 79 demonstraram um perfil linear significativo para o grupo drogável. A regressão linear destes valores foi realizada utilizando a ferramenta R, o R^2 encontrado para os dois ajustes mostrou-se significativo. Para a área sob influência de doadores de hidrogênio o ajuste indicou um R^2 de 0,96, enquanto para a área de influência dos aceptores de hidrogênio este valor é de 0,99. Considerando o número de proteínas obtido no banco de dados, os descritores de densidade de hidrogênio foram relevantes para caracterizar a drogabilidade das cavidades analisadas.

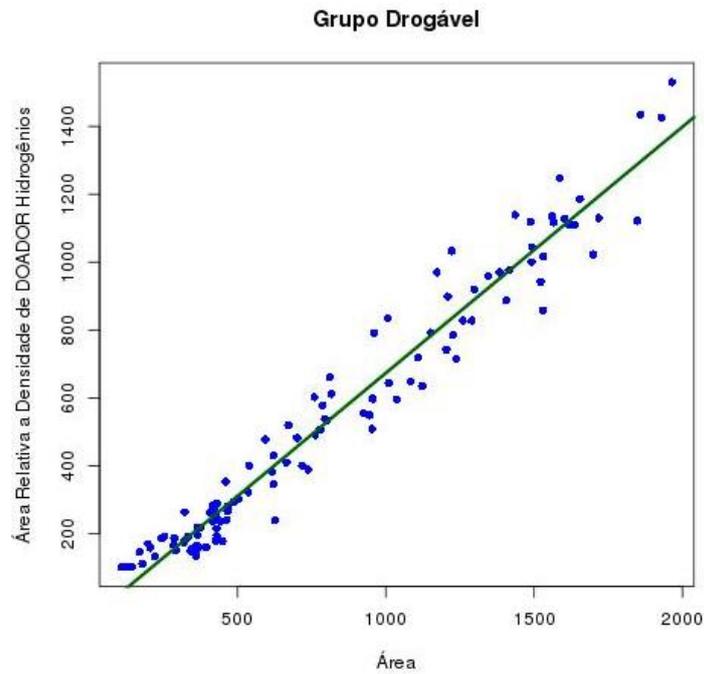


Figura 77: Regressão linear obtida para a área de influência dos doadores e aceptores de hidrogênio. O R^2 obtido foi de 0,96.

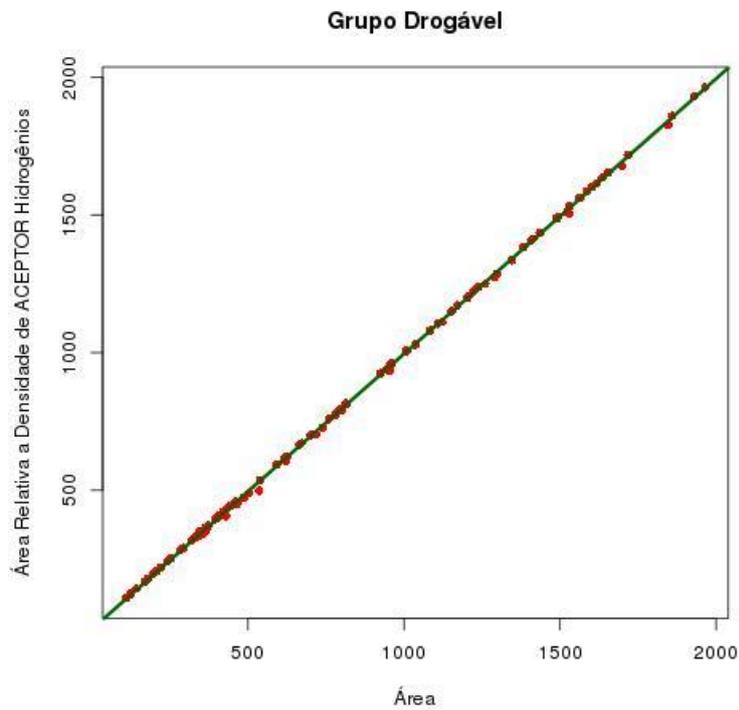


Figura 78: Regressão linear obtida para a área de influência dos doadores e aceptores de hidrogênio. O R^2 obtido foi de 0,99.

6 Conclusão

Podemos concluir, na primeira parte do trabalho realizado sobre o potencial lipofílico e densidade de ligação de hidrogênio, que as rotinas inseridas no KVFinder geram uma descrição aceitável das propriedades das cavidades. Os três complexos proteína-ligante utilizados para o teste demonstraram visualmente semelhanças com as cavidades que foram calculadas pelo MOLCAD do SYBYL. O potencial lipofílico demonstrou complementariedade química com o ligante com características de fármacos. O cálculo da densidade de aceptores e doadores de hidrogênio também demonstrou complementariedade com grupos aceptores e regiões com características de doadores de hidrogênio, e vice-versa. Estes resultados foram demonstrados através da estrutura química da molécula e a região onde se encontrava. As mesmas conclusões com relação à densidade de ligação de hidrogênio podem ser observadas quanto à relação cavidade-resíduo da proteína. As regiões com mais alto potencial de doadores e aceptores de hidrogênio foram aquelas que apresentaram resíduos que possuem maior probabilidade de realizar estas interações e principalmente seu posicionamento quanto às cavidades.

As relações topológicas da cavidade, volume e área apresentaram grandes diferenças em relação ao tamanho da grade. A superfície molecular é muito sensível à escala utilizada na grade e apresenta melhores resultados com o valor definido como $0,2\text{\AA}/\text{pts}$. O potencial lipofílico diminui sensivelmente quando aumenta a distância entre os pontos da grade.

Seguindo a análise do KVFinder podemos concluir que o custo computacional para a busca de todas as cavidades da proteína se torna muito grande quando se aumenta o número de átomos, sendo esta característica uma limitação para a utilização da ferramenta no uso de todo um proteoma. Os algoritmos utilizados para cálculo de potencial demonstraram um tempo bem menor de execução.

Quanto à drogabilidade, concluímos que para o conjunto de proteínas drogáveis utilizado na análise o valor a quantidade da área interna à proteína (razão entre a superfície molecular e a área que engloba todo o volume, incluindo

a tampa) é maior ou próximo de 60%, concordando com valores da literatura. Em relação à área e ao volume os grupos drogáveis apresentaram maior valor de área superficial da molécula assim como a média do volume encontrado foi bem maior. A relação hidrofobicidade vs área mostrou linearidade assim como a relação hidrofiliicidade vs área. Os valores de lipofiliicidade foram maiores para os grupos drogáveis, enquanto que os grupos não drogáveis demonstraram ter cavidades mais hidrofílicas. Em relação à área os resultados foram semelhantes para os dois grupos, sendo possível também realizar uma regressão linear para o conjunto drogável encontrando valores promissores para uma futura análise mais sistemática da drogabilidade.

A densidade de hidrogênio tanto de aceptores como de doadores não demonstrou nenhuma diferença significativa. Entretanto quando analisado o número absoluto de doadores e aceptores foi possível observar o crescimento linear em relação à área sendo possível uma regressão, assim como a análise da área que indicou alto valor para R^2 .

7 Referências

1. Drews J. Drug Discovery: A Historical Perspective. Vol. 287, Science. 2000. p. 1960–4.
2. Goodman & Gilman - As Bases Farmacológicas da Terapêutica.pdf.
3. Rubin RP. A Brief History of Great Discoveries in Pharmacology : In Celebration of the Centennial Anniversary of the Founding of the American Society of Pharmacology and Experimental Therapeutics. 2007;59(4):289–359.
4. Drews J. Drug Discovery: A Historical Perspective. Science (80-) [Internet]. 2000 Mar 17 [cited 2014 Sep 4];287(5460):1960–4. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.287.5460.1960>
5. Maehle AH, Prüll CR, Halliwell RF. The emergence of the drug receptor theory. Nat Rev Drug Discov. 2002;1(8):637–41.
6. Fleming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B.influenzae. Br J Exp Pathol. 1929;10(1923):226–36.
7. Bosch F, Rosich L. The contributions of paul ehrlich to pharmacology: A tribute on the occasion of the centenary of his nobel prize. Vol. 82, Pharmacology. 2008. p. 171–9.
8. Watson JD, Crick FHC. Molecular structure of nucleic acids. Nature [Internet]. 1953;171:737–8. Available from: <http://www.nature.com/physics/looking-back/crick/%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/13054692>
9. Corey RB, Pauling L. Molecular models of amino acids, peptides, and proteins. Rev Sci Instrum. 1953;24:621–7.
10. RAMACHANDRAN GN, RAMAKRISHNAN C, SASISEKHARAN V. Stereochemistry of polypeptide chain configurations. J Mol Biol. 1963;7:95–9.
11. Dayhoff MO, Barker WC, Hunt LT. [47] Establishing homologies in protein sequences. Methods Enzymol. 1983;91:524–45.
12. Warshel A. Consistent Force Field Calculations. II. Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes. J Chem Phys [Internet].

- 1970;53:582. Available from:
<http://scitation.aip.org/content/aip/journal/jcp/53/2/10.1063/1.1674031>
13. Levitt M, Chothia C. Structural patterns in globular proteins. Vol. 261, Nature. 1976. p. 552–8.
 14. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291:1304–51.
 15. Verli H et al. Bioinformática: da Biologia à flexibilidade Molecular. 11-21 p.
 16. Initiative AGH. Fact sheet. 2007;(February).
 17. Goodford PJ, St-Louis J, Wootton R. The interaction of human haemoglobin with allosteric effectors as a model for drug-receptor interactions. Br J Pharmacol. 1980;68:741–8.
 18. Blundell T, Dodson G, Hodgkin D, Mercola D. Insulin: The Structure in the Crystal and its Reflection in Chemistry and Biology by. Adv Protein Chem. 1972;26.
 19. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. Biochemistry [Internet]. 1991 Oct;30(43):10363–70. Available from:
<http://pubs.acs.org/doi/abs/10.1021/bi00107a001>
 20. Congreve M, Murray CW, Blundell TL. Keynote review : Structural biology and drug discovery. 2005;10(13).
 21. Blundell TL. Structure-based drug design. Nature. 1996;384:23–6.
 22. Miller M, Schneider J, Sathyanarayana BK, Toth M V, Marshall GR, Clawson L, et al. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science. 1989;246:1149–52.
 23. Kubinyi H. Strategies and recent technologies in drug discovery. Vol. 50, Pharmazie. 1995. p. 647–62.
 24. Franc I, Lipinski A, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. 1997;23.
 25. Patel D V., Gordon EM. Applications of small-molecule combinatorial chemistry to drug discovery. Drug Discov Today. 1996;1:134–44.
 26. BAUM RM. Combinatorial Approaches Provide Fresh Leads for Medicinal Chemistry. Chem Eng News. 1994;72:20–6.
 27. Ferreira RS, Oliva G, Física I De, et al. Revisão. 2011;34(10):1770–8.

28. Andrews PR, Craik DJ, Martin JL. Functional group contributions to drug-receptor interactions. *J Med Chem*. 1984;27:1648–57.
29. Brink T Ten, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *J Chem Inf Model*. 2009;49:1535–46.
30. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45:177–82.
31. Terstappen GC, Reggiani A. In silico research in drug discovery. *Trends Pharmacol Sci* [Internet]. 2001 Jan;22(1):23–6. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0165614700015844>
32. Laskowski RA. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph*. 1995;13:323–30.
33. Hendlich M, Rippmann F, Barnickel G. LIGSITE : Automatic and efficient detection of potential small molecule- binding sites in proteins. 1998;3263(98):359–63.
34. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci*. 1998;7:1884–97.
35. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*. 1985;28:849–57.
36. An J, Totrov M, Abagyan R. Comprehensive identification of “druggable” protein ligand binding sites. *Genome Inform*. 2004;15:31–41.
37. Oliveira SHP, Ferraz F a N, Honorato R V, Xavier-Neto J, Sobreira TJP, de Oliveira PSL. KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics* [Internet]. 2014 Jan [cited 2014 Sep 1];15:197. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4071799&tool=pmcentrez&rendertype=abstract>
38. Laurie ATR, Jackson RM. Structural bioinformatics Q-SiteFinder : an energy-based method for the prediction of protein – ligand binding sites. 2005;21(9):1908–16.
39. Guilloux V Le, Schmidtke P, Tuffery P. Fpocket : An open source platform for ligand pocket detection. 2009;11:1–11.

40. Masuya M, Doi J. Detection and geometric modeling of molecular surfaces and cavities using digital mathematical morphological operations. *J Mol Graph*. 1995;13:331–6.
41. Lopes PS. Desenvolvimento de um algoritmo para identificação e caracterização de cavidades em regiões específicas de estruturas tridimensionais de proteínas . 2011;
42. An J, Totrov M, Abagyan R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics*. 2005;4:752–61.
43. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics [Internet]*. 2012 Aug 1 [cited 2014 Dec 15];28(15):2074–5. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/22628523>
44. Hansch C. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Chem Res*. 1969;2:232–9.
45. Gaillard P, Carrupt P a, Testa B, Boudon a. Molecular lipophilicity potential, a tool in 3D QSAR: method and applications. *J Comput Aided Mol Des [Internet]*. 1994 Apr;8(2):83–96. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/7914913>
46. Wildman S, Crippen G. Prediction of Physicochemical Parameters by Atomic Contributions. *J Chem Inf Model [Internet]*. 1999;39:868–73. Available from:
<http://pubs.acs.org/doi/abs/10.1021/ci990307l%5Cnhttp://pubs.acs.org/cgi-bin/doilookup/?10.1021/ci990307l>
47. Heiden W, Moeckel G, Brickmann J. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J Comput Aided Mol Des [Internet]*. 1993 Oct;7(5):503–14. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/8294943>
48. Ghose K. Atomic Physicochemical Parameters for Three-Dimensional Structure - Directed Quantitative Structure- Activity Relationships I . Partition Coefficients as a Measure of. 1986;7(4):565–77.
49. Hansch C, Rockwell SD, Jow PY, Leo A, Steller EE. Substituent constants for correlation analysis. *J Med Chem*. 1977;20:304–6.

50. Nys GG, Rekker RF. Concept of hydrophobic fragmental constants (f-values). II. Extension of its applicability to the calculation of lipophilicities of aromatic and heteroaromatic structures. *Eur J Med Chem.* 1974;9:361–75.
51. Rekker RF. The hydrophobic fragmental constant. *Pharmacochem Libr.* 1977;2:231–8.
52. Ghose AK, Pritchett A, Crippen GM. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity : Modeling Hydrophobic Interactions Relationships 1. 1988;9(1):80–90.
53. Audry E, Dubost J-P, Colleter J-C, Dallet P. A new approach to structure-activity relations: the “molecular lipophilicity potential.” *Eur J Med Chem.* 1986;21:71–2.
54. Furet P, Sele a., Cohen NC. 3D molecular lipophilicity potential profiles: a new tool in molecular modeling. *J Mol Graph [Internet].* 1988 Dec;6(4):182–9. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0263785598800015>
55. Fauchère JL. A quantitative structure-activity relationship study of the inhibitory action of a series of enkephalin-like peptides in the guinea pig ileum and mouse vas deferens bioassays. *J Med Chem.* 1982;25:1428–31.
56. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain . *J Chem Inf Comput Sci.* 1989;29:163–72.
57. Koji B. The Nature of Hydrogen Bond : New linsights Into Old Theories. 2008;692–708.
58. Kollman PA, Allen LC. THE THEORY OF THE HYDROGEN BOND ”. 1972;1971(2):1966–70.
59. Perrin CL, Nielson JB. “Strong” hydrogen bonds in chemistry and biology. *Annu Rev Phys Chem.* 1997;48:511–44.
60. Base S. Sybyl base.
61. French AS. Computer Simulation of Space-Filling Molecular Models. *IEEE*

- Trans Comput. 1977;C-26.
62. Zachmann CD, Kast SM, Brickmann J. Quantification and visualization of molecular surface flexibility. *J Mol Graph*. 1995;13:89–97.
 63. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* [Internet]. 2010 Aug 12 [cited 2014 Nov 2];53(15):5858–67. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20684613>
 64. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1:727–30.
 65. Radusky L, Defelipe L a, Lanzarotti E, Luque J, Barril X, Marti M a, et al. TuberQ: a Mycobacterium tuberculosis protein druggability database. *Database (Oxford)* [Internet]. 2014 Jan [cited 2015 May 20];2014:bau035. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4014675&tool=pmcentrez&rendertype=abstract>
 66. Barril X, Luque FJ. Molecular simulation methods in drug discovery: a prospective outlook. *J Comput Aided Mol Des* [Internet]. 2012 Jan [cited 2014 Jul 16];26(1):81–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22160626>
 67. Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, Chen YZ. Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev*. 2006;58:259–79.
 68. Hubbard TJP, Ailey B, Brenner SE, Murzin AG, Chothia C. SCOP: A structural classification of proteins database. Vol. 27, *Nucleic Acids Research*. 1999. p. 254–6.
 69. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Vol. 42, *Nucleic Acids Research*. 2014.
 70. Overington JP, Al-lazikani B, Hopkins AL. How many drug targets are there ? *2006*;5(December):993–6.
 71. Vitkup D, Melamud E, Moulton J, Sander C. Completeness in structural genomics. *Nat Struct Biol*. 2001;8:559–66.
 72. Han LY, Zheng CJ, Xie B, Jia J, Ma XH, Zhu F, et al. Support vector machines approach for predicting druggable proteins: recent progress in

- its exploration and investigation of its usefulness. Vol. 12, Drug Discovery Today. 2007. p. 304–13.
73. Hann MM, Leach AR, Harper G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J Chem Inf Comput Sci.* 2001;41:856–64.
 74. Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov.* 2011;10:507–19.
 75. Barril X. Druggability predictions: Methods, limitations, and applications. *Wiley Interdiscip Rev Comput Mol Sci.* 2013;3:327–38.
 76. Brady GP, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des.* 2000;14:383–401.
 77. Coleman RG, Sharp KA. Protein pockets: Inventory, shape, and comparison. *J Chem Inf Model.* 2010;50:589–603.
 78. Hajduk PJ, Huth JR, Tse C. Predicting protein druggability REVIEWS. 2005;10(23).
 79. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. 2007;25(1):71–5.
 80. Halgren TA. Identifying and Characterizing Binding Sites and Assessing Druggability. 2009;377–89.
 81. Perola E, Herman L, Weiss J. Development of a Rule-Based Method for the Assessment of Protein Druggability. 2012;
 82. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model [Internet].* 2012 Feb 27;52(2):360–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22148551>
 83. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 2007;35(SUPPL.2).
 84. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. Vol. 26, *Journal of Computational Chemistry.* 2005. p. 1668–88.
 85. Geng Y, Bush M, Mosyak L, Wang F, Fan QR. Structural mechanism of

- ligand activation in human GABA(B) receptor. *Nature* [Internet]. 2013 Dec 12 [cited 2015 Apr 21];504(7479):254–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3865065&tool=pmcentrez&rendertype=abstract>
86. Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, et al. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature*. 1989;342:299–302.
 87. Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL, Rose DR. Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *J Mol Biol* [Internet]. 2008 Jan 18 [cited 2015 May 29];375(3):782–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18036614>
 88. Froestl W. Chemistry and pharmacology of GABAB receptor ligands. Vol. 58, *Advances in Pharmacology*. 2010. 19-62 p.
 89. Pliska V, Testa B VDW. Lipophilicity in Drug Action and Toxicology. Vol. 134, *The Journal of chemical physics*. 1996. 439 p.
 90. Cheung J, Rudolph MJ, Burshteyn F, Cassidy MS, Gary EN, Love J, et al. Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *J Med Chem*. 2012;55(22):10282–6.
 91. Kim EE, Bakei CT, Dwyer MD, Murcko MA, Rac BG, Tung RD, et al. Crystal Structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J Am Chem Soc*. 1995;117(3):1181–2.
 92. Sim L, Quezada-Calvillo R, Sterchi EE, Nichols BL, Rose DR. Human Intestinal Maltase-Glucoamylase: Crystal Structure of the N-Terminal Catalytic Subunit and Basis of Inhibition and Substrate Specificity. *J Mol Biol*. 2008;375(3):782–92.
 93. Koide A, Abbatiello S, Rothgery L, Koide S. Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc Natl Acad Sci U S A*. 2002;99(3):1253–8.
 94. Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, Leslie AGW, et al. The structural basis for agonist and partial agonist action on a

- β 1-adrenergic receptor. *Nature*. 2011;469(7329):241–5.
95. Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, et al. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr Sect D Biol Crystallogr* [Internet]. 2007;63(1):80–93. Available from: <http://scripts.iucr.org/cgi-bin/paper?S0907444906047287>
 96. Kurumbail RG, Stevens AM, Gierse JK, McDonald JJ, Stegeman RA, Pak JY, et al. Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* [Internet]. 1996;384(6610):644–8. Available from: <http://www.nature.com/doi/10.1038/384644a0>
 97. Prota AE, Danel F, Bachmann F, Bargsten K, Buey RM, Pohlmann J, et al. The novel microtubule-destabilizing drug BAL27862 binds to the colchicine site of tubulin with distinct effects on microtubule organization. *J Mol Biol*. 2014;426(8):1848–60.
 98. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* (80-). 1991;252(5007):839–42.
 99. Ren L, Qin X, Cao X, Wang L, Bai F, Bai G, et al. Structural insight into substrate specificity of human intestinal maltase-glucoamylase. *Protein Cell* [Internet]. 2011;2(10):827–36. Available from: <http://link.springer.com/10.1007/s13238-011-1105-3>
 100. Rowlinson SW, Kiefer JR, Prusakiewicz JJ, Pawlitz JL, Kozak KR, Kalgutkar AS, et al. A Novel Mechanism of Cyclooxygenase-2 Inhibition Involving Interactions with Ser-530 and Tyr-385. *J Biol Chem*. 2003;278(46):45763–9.
 101. Byeon IJL, Llinás M. Solution structure of the tissue-type plasminogen activator kringle 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic drug. *J Mol Biol*. 1991;222(4):1035–51.