

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
LABORATÓRIO DE QUÍMICA MEDICINAL E COMPUTACIONAL

Tiago Luiz Moda

**Modelagem *In Silico* de Propriedades Farmacocinéticas Para a
Avaliação de Candidatos a Novos Fármacos**

São Carlos

2011

Tiago Luiz Moda

**Modelagem *In Silico* de Propriedades Farmacocinéticas Para a
Avaliação de Candidatos a Novos Fármacos**

Tese apresentada ao Programa de Pós-Graduação do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciência.

Área de Concentração: Física Aplicada
Opção: Física Biomolecular

Orientador: Prof. Dr. Adriano Defini
Andricopulo

Versão Corrigida
(versão original disponível na unidade que aloja o programa)

São Carlos
2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Moda, Tiago Luiz.

Modelagem *In silico* de propriedades farmacocinéticas para a avaliação de candidatos a novos fármacos./ Tiago Luiz Moda; orientador Adriano Defini Andricopulo – versão corrigida -- São Carlos, 2011.

216 p.

Tese (Doutorado – Programa de Pós-Graduação em Física. Área de Concentração: Física Aplicada. Opção: Física Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2011.

1. Química medicinal. 2. Farmacocinética. 3. Modelagem molecular. 4. Base de dados. 5. Holograma QSAR. I. Andricopulo, Adriano Defini, orient. II. Título.

AGRADECIMENTOS

- ✓ Ao meu orientador, Prof. Dr. Adriano D. Andricopulo, pela orientação, confiança, amizade e pela ótima oportunidade de crescimento, tanto profissional quanto pessoal. Agradeço-o também por valorizar o trabalho sério e me mostrar como dedicação e compromisso conduzem a ciência da Química Medicinal.
- ✓ Aos alunos de iniciação científica os quais têm contribuição inestimável em alguns trabalhos realizados nessa tese de doutorado: Leonardo G. Torres, Alexandre E. Carrara e André C. da Silva.
- ✓ Ao Prof. Dr. Alexander Tropsha e seu grupo de pesquisa pela oportunidade dada e pelos valiosos ensinamentos sobre métodos QSAR durante o período de doutorado sanduíche nos EUA.
- ✓ Aos professores e funcionário do Instituto de Física de São Carlos da Universidade de São Paulo pela colaboração e excelente suporte para realização deste trabalho.
- ✓ Aos professores e amigos do Centro de Biotecnologia Molecular Estrutura – CEPID-FAPESP, que contribuíram positivamente para a realização deste trabalho.
- ✓ À FAPESP e CAPES pelo apoio financeiro.

The overall process from idea to product can therefore take from 9 to 16+ years to complete. Furthermore, historically only one in 20 compounds that start the development process ever become marketed drugs. This is not an enterprise for people with a low tolerance for failure or for those who need immediate gratification.

LAMATTINA, J. L. Drug truths - Dispelling the Myths About Pharma R&D. New Jersey: Wiley, 2009. p. 24-25.

RESUMO

MODA, T. L. **Modelagem *In Silico* de Propriedades Farmacocinéticas Para a Avaliação de Candidatos a Novos Fármacos.** 2011. 216 p. Tese (Doutorado em Ciência). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

Os processos farmacocinéticos de absorção, distribuição, metabolismo e excreção (ADME), têm sido identificados como as principais causas do insucesso de candidatos a fármacos em estágios avançados de desenvolvimento clínico. As metodologias modernas de modelagem *in silico* de propriedades farmacocinéticas estão integradas ao processo de planejamento de fármacos, sendo de extremo valor na identificação e seleção de novas entidades químicas candidatas a fármacos. Esta área emergente está atraindo grande atenção da indústria farmacêutica mundial, que tem integrado a otimização de múltiplas propriedades farmacodinâmicas e farmacocinéticas em todos os estágios de projetos de pesquisa e desenvolvimento (P&D). As propriedades farmacocinéticas podem ser estudadas através do uso de métodos *in silico* como o estudo das relações quantitativas entre a estrutura e atividade (QSAR), ou estrutura e propriedade (QSPR), entre outros. O trabalho desenvolvido nesta tese de doutorado teve como importante objetivo estudar as relações quantitativas entre a estrutura química e propriedades farmacocinéticas como absorção intestinal, metabolismo de primeira passagem mediado pelo CYP, permeabilidade da barreira hematoencefálica, bem como eventos de extremo interesse que as influenciem como a inibição da glicoproteína-P e solubilidade aquosa. Para a realização deste trabalho, conjuntos padrões de dados foram organizados para as propriedades farmacocinéticas contendo a informação qualificada sobre a estrutura química e a propriedade alvo correspondente. Os conjuntos de dados criados formaram as bases científicas para o desenvolvimento dos modelos preditivos empregando o método holograma QSAR (HQSAR). Os modelos finais de HQSAR gerados neste trabalho possuem elevada consistência interna e externa, apresentando bom poder de correlação e predição das propriedades alvo. Os modelos desenvolvidos, assim como os dados farmacocinéticos coletados,

foram disponibilizados para acesso livre através da internet na base de dados PK/DB (www.pkdb.ifsc.usp.br). Devido à simplicidade, robustez e consistência, estes modelos são guias úteis em Química Medicinal nos estágios iniciais do processo de descoberta e desenvolvimento de fármacos.

Palavras-chave: Química medicinal. Farmacocinética. Modelagem molecular. Base de dados. Holograma QSAR.

ABSTRACT

MODA, T. L. **Pharmacokinetic Properties *In Silico* Modeling for New Chemical Entities Evaluation**. 2011. 216 p. Tese (Doutorado em Ciência). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

The pharmacokinetic (PK) processes of absorption, distribution, metabolism and excretion (ADME), have been identified as one of the major causes of new chemical entities (NCEs) failure in early clinical trials. *In silico* models are receiving increased attention in recent years from the pharmaceutical industry, which is integrating a paradigm of multiple pharmacodynamic and pharmacokinetic properties optimization for the development of NCEs. ADME properties can be studied by *in silico* methods, such as quantitative structure–activity relationships (QSAR) or structure–property (QSPR), among other methodologies. The main goal of this PhD thesis was to study the quantitative relationships between chemical structure and pharmacokinetic properties, such as intestinal absorption, CYP mediated first pass metabolism, blood brain barrier permeability, as well as other important events that have influence on these PK properties, such as P-glycoprotein inhibition and water solubility. In the present work, standard data sets were organized encompassing the structural information and corresponding pharmacokinetic data. The standard data sets established the scientific basis for the development of predictive models using the hologram QSAR (HQSAR) method. The final HQSAR models possess high internal and external consistency with good correlative and predictive power for endpoint PK properties. All *in silico* models generated and standard data sets are freely available on the internet through the Database for Pharmacokinetic Properties (PK/DB - www.pkdb.ifsc.usp.br). Due to the simplicity, robustness and effectiveness, these models are useful guides in Medicinal Chemistry in the early stages of the drug discovery and development process.

Keywords: Medicinal chemistry. Pharmacokinetic. Molecular modeling. Database.
Hologram QSAR.

LISTA DE ABREVIATURAS

2 D	Bidimensional
3D	Tridimensional
5-HT	Receptores Serotonérgicos
A	Átomos
AD	Domínio de Aplicação
ADME	Absorção, Distribuição, Metabolismo e Excreção
B	Ligação
BBB	Barreira Hematoencefálica
C	Conectividade
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBME	Centro de Biotecnologia Molecular Estrutural
CCR	Índice de Classificações Corretas
CEPID	Centros de Pesquisa, Inovação e Difusão
Ch	Quiralidade
CRC	Algoritmo Cíclico da Verificação de Redundância
CYP	Citocromo P450 Oxidase
DA	Doador e Aceptor
DDI	Interação Fármaco-Fármaco
EC	<i>Enzyme Commission</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FC	Fase Clínica
FDA	Administração de Alimentos e Medicamentos dos Estados Unidos
GPCRs	Receptores Acoplados à Proteína G
H	Hidrogênio
HBA	Aceptores de Ligação Hidrogênio
HBD	Doadores de Ligação Hidrogênio
HIA	Absorção Intestinal Humana
HQSAR	Holograma QSAR

IFSC	Instituto de Física de São Carlos
kNN	Método do Vizinho mais próximo
$\log P$	Coefficiente de Partição de uma Substância em Sistema Octanol/Água
LQMC	Laboratório de Química Medicinal e Computacional
MDR	Resistência Múltipla a Fármacos
MM	Massa Molar
MSS	Planilha Molecular
NCE	Nova Entidade Química
PDEE	Programa de Doutorado no País com Estágio no Exterior
PDSP	<i>Psychoactive Drug Screening Program</i>
P&D	Pesquisa e Desenvolvimento
P-gp	Glicoproteína-P
PK/DB	Base de Dados para Propriedades Farmacocinéticas
PLS	Regressões Multivariadas por Mínimos Quadrados Parciais
PPB	Ligação às Proteínas Plasmáticas
PSA	Área de Superfície Polar
QSAR	Relações Quantitativas Entre a Estrutura e Atividade
QSPR	Relações Quantitativas Entre a Estrutura e Propriedade
SE	Esfera de Exclusão
SNC	Sistema Nervoso Central
SVM	Maquinas de Vetor Suporte
UNC	Universidade da Carolina do Norte
V_d	Volume de Distribuição
VolSurf	Volume & Superfície
VS	Triagem Virtual
WDI	<i>World Drug Index</i>
WOMBAT	<i>World of Molecular Bioactivity</i>

SUMÁRIO

Capítulo 1 Introdução	19
1 Introdução.....	21
1.1 Química Medicinal e a Inovação Farmacêutica.....	21
1.2 Processo de Descoberta e Desenvolvimento de Fármacos	23
1.3 Farmacodinâmica e Farmacocinética	25
1.4 Propriedades Farmacocinéticas.....	27
1.4.1 Absorção.....	27
1.4.2 Distribuição.....	30
1.4.2.1Ligação às Proteínas Plasmáticas.....	30
1.4.2.2Volume de Distribuição.....	32
1.4.3 Metabolismo e Excreção.....	33
1.4.4 Biodisponibilidade.....	38
1.4.5 Barreira Hematoencefálica	39
1.5 Solubilidade.....	40
1.6 Glicoproteína-P e os Processos de ADME.....	41
1.7 Modelagem <i>In Silico</i> de Propriedades Farmacocinéticas.....	43
1.7.1 Holograma QSAR.....	46
Capítulo 2 Objetivos	51
2 Objetivos.....	53
Capítulo 3 Materiais e Métodos	55
3 Materiais e Métodos.....	57
3.1 Conjunto Padrões de Dados.....	57
3.1.1 Conjunto de Dados de Absorção Intestinal Humana.....	58

3.1.2 Conjunto de Dados de Metabolismo de Primeira Passagem.....	58
3.1.3 Conjunto de Dados de Permeabilidade da Barreira Hematoencefálica.....	59
3.1.4 Conjunto de Dados de Inibição da Glicoproteína-P.....	60
3.1.5 Conjunto de Dados de Solubilidade Aquosa.....	60
3.2 Modelagem de QSAR.....	61

Capítulo 4 Resultados e Discussão.....

4 Resultados e Discussão.....	65
4.1 Absorção Intestinal Humana.....	65
4.1.1 Organização do Conjunto de Dados.....	65
4.1.2 Desenvolvimento de Modelos HQSAR-HIA.....	71
4.1.3 Validação dos Modelos de HQSAR-HIA.....	76
4.2 Metabolismo de Primeira Passagem.....	81
4.2.1 Organização do Conjunto de Dados.....	81
4.2.2 Desenvolvimento de Modelos de HQSAR-MET.....	84
4.2.3 Validação dos Modelos de HQSAR-MET.....	88
4.3 Permeabilidade da Barreira Hematoencefálica.....	92
4.3.1 Organização do Conjunto de Dados.....	92
4.3.2 Desenvolvimento de Modelos de HQSAR-BBB.....	95
4.3.3 Validação dos Modelos de HQSAR-BBB.....	98
4.4 Inibição da Glicoproteína-P.....	101
4.4.1 Organização do Conjunto de Dados.....	101
4.4.2 Desenvolvimento de Modelos de HQSAR-P-gp.....	105
4.4.3 Validação dos Modelos de HQSAR-P-gp.....	109
4.4.4 Contribuição dos Fragmentos.....	112
4.5 Solubilidade Aquosa.....	114
4.5.1 Organização do Conjunto de Dados.....	114
4.5.2 Desenvolvimento de Modelos de HQSAR-SOL.....	115
4.5.3 Validação dos Modelos de HQSAR-SOL.....	118

Capítulo 5 Base de Dados PK/DB	125
5 PK/DB: Base de Dados para Propriedades Farmacocinéticas.....	127
5.1 Apresentação.....	127
5.2 Caracterização.....	127
5.3 Estrutura e Aplicações.....	128
5.4 Predição de Propriedades Farmacocinéticas	129
Capítulo 6 Estágio no Exterior	131
6 Estágio de Doutorado no Exterior.....	133
6.1 Estágio.....	133
6.2 Base de Dados Complexas.....	133
6.2.1 GPCRs.....	133
6.3 Metodologia.....	136
6.3.1 Conjunto de Dados.....	136
6.3.2 Método do Vizinho Mais Próximo.....	137
6.3.3 Máquinas de Vetor Suporte.....	139
6.4 Cálculo dos Descritores.....	141
6.5 Índice de Classificações Corretas.....	142
6.6 Domínio de Aplicação.....	142
6.7 Modelos QSAR <i>k</i> NN e SVM.....	143
6.8 Triagem Virtual em Paralelo.....	146
6.9 Validação Experimental.....	148
Capítulo 7 Conclusões e Perspectivas	153
7 Conclusões e Perspectivas.....	155
Publicações	157

Referências	161
--------------------------	-----

Apêndices	181
------------------------	-----

Apêndice A.....	183
-----------------	-----

Apêndice B.....	185
-----------------	-----

Apêndice C.....	187
-----------------	-----

Apêndice D.....	189
-----------------	-----

Apêndice E.....	191
-----------------	-----

Apêndice F.....	193
-----------------	-----

Apêndice G.....	201
-----------------	-----

Anexo	209
--------------------	-----

Anexo A.....	211
--------------	-----

CAPÍTULO 1

Introdução

1 Introdução

1.1 Química Medicinal e Inovação Farmacêutica

A indústria farmacêutica vem passando por importantes transformações decorrentes, sobretudo, dos notáveis avanços científicos e tecnológicos fundamentados em novos paradigmas de natureza intrinsecamente multidisciplinar. A descoberta de novos fármacos de diversas classes terapêuticas tem proporcionado melhorias significativas na qualidade de vida das diversas populações no mundo¹⁻³. O papel relevante da química medicinal nos processos inovativos da gênese planejada de fármacos é preservado ao longo dos anos nas várias interfaces e conexões entre química e biologia. A química medicinal possui forte caráter multidisciplinar, abrangendo diversas áreas importantes, como química orgânica, bioquímica, farmacologia, informática, biologia molecular e estrutural, entre outras. Várias estratégias de química medicinal podem ser empregadas para a investigação do espaço químico-biológico, como a organização de bases de dados, aplicação filtros moleculares, emprego de triagens biológicas automatizadas em larga escala (do inglês, *high-throughput screening* – HTS) e triagem virtual (do inglês, *virtual screening* – VS), além das técnicas de planejamento baseado na estrutura do receptor (do inglês, *structure-based drug design* – SBDD) e de planejamento baseado na estrutura do ligante (do inglês, *ligand-based drug design* – LBDD). Em particular, os métodos em quimio e bioinformática têm sido amplamente empregados no processo de descoberta de novos fármacos, encontrando grande aplicação nas etapas de pesquisa básica, incluindo a identificação, seleção e otimização de moléculas candidatas a novas entidades químicas (do inglês, *new chemical entities* – NCEs)⁴. A integração de métodos computacionais avançados com o trabalho químico e biológico experimental é essencial na busca por novas moléculas bioativas com elevado potencial terapêutico e qualificadas em uma série complexa de propriedades farmacodinâmicas e farmacocinéticas⁵⁻⁶.

O processo de descoberta e desenvolvimento de NCEs é longo, complexo e envolve altos investimentos⁷. Desde a concepção até a introdução de um único fármaco no mercado farmacêutico, são estimados entre nove e dezesseis anos em P&D (pesquisa e desenvolvimento), com investimentos totais estimados no ano de 2009 da ordem de US\$ 500-940 milhões, podendo alcançar, em alguns casos, cifras superiores a US\$ 1 bilhão⁸⁻¹³.

Embora a indústria farmacêutica ocupe um lugar de destaque nos segmentos mais rentáveis do mercado mundial, o número de novos fármacos que alcançaram o mercado farmacêutico diminuiu consideravelmente na última década apesar dos crescentes investimentos em P&D (Figura 1.1)¹⁴. Este fato é agravado pelas perdas de patentes (perda de exclusividade de mercado) que as indústrias vêm sofrendo, reduzindo de forma significativa a fonte de receitas em função da competição com medicamentos genéricos. No ano de 2010, por exemplo, a Pfizer perdeu em países como Brasil, Canadá e Espanha a exclusividade sobre o hipolipemiante atorvastatina (Lípitor[®]), fármaco de maior sucesso de vendas da história da indústria farmacêutica com faturamento de US\$ 10,7 bilhões, somente em 2010¹⁵. Como medida gerencial de reestruturação, a Pfizer encerrará as atividades de P&D em Sandwich, Inglaterra, um de seus mais importantes centros de pesquisa e inovação. Alguns *blockbusters* (termo do inglês usado para designar os fármacos com vendas anuais superiores a US\$ 1 bilhão) desenvolvidos neste centro incluem a sildenafil (Viagra[®]) para o tratamento de disfunção erétil, o anti-hipertensivo anlodipino (Norvasc[®]) e o antifúngico fluconazol (Zoltec[®])¹⁵⁻¹⁶.

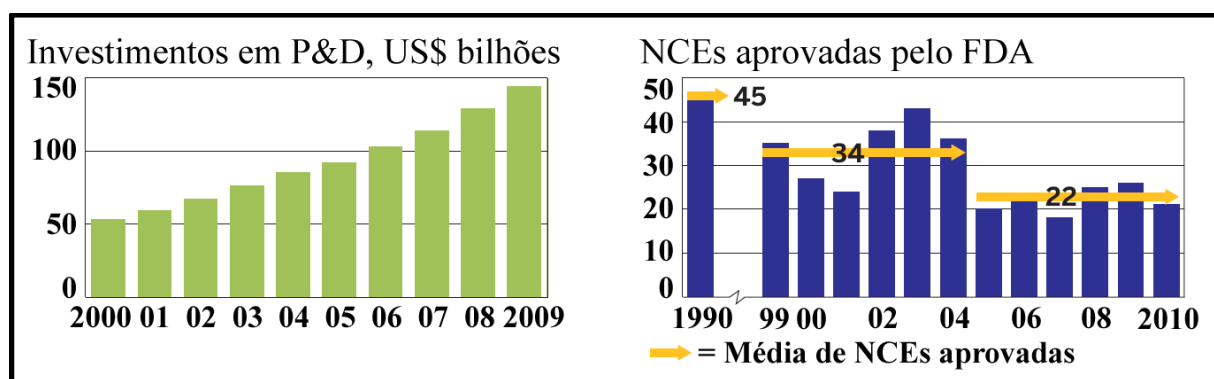


Figura 1.1 – Os investimentos em P&D de fármacos têm aumentado ao longo da última década, ao passo que um decréscimo no número de NCEs que chegam ao mercado tem sido observado¹⁴.

Estima-se que somente um em cada vinte compostos que passam pelo processo clínico de desenvolvimento chega ao mercado farmacêutico⁸. Somente em 2010, US\$ 63 bilhões foram investidos em P&D pelas indústrias farmacêuticas americanas, juntamente com os Institutos Nacionais de Saúde (do inglês, *National Institutes of Health* – NIH). Mesmo com o elevado nível de investimentos, apenas 22 NCEs foram aprovadas pelo FDA (Figura 1.1)^{14,17}. Para alcançar as exigências de produtividade impostas pelos elevados níveis de investimentos em P&D de fármacos, as maiores companhias farmacêuticas do mundo têm adotado como uma de suas principais estratégias o investimento em novas tecnologias e inovação para a descoberta de fármacos.

1.2 Processo de Descoberta e Desenvolvimento de Fármacos

O processo de descoberta e desenvolvimento de novos fármacos é dividido em duas grandes fases denominadas: pré-clínica (de pesquisa básica ou descoberta) e clínica (ou desenvolvimento)^{8,18}. De forma geral, a fase pré clínica envolve as seguintes etapas: (i) abordagem científica do problema médico; (ii) seleção de um alvo biológico relacionado à doença-alvo; (iii) desenvolvimento de estratégias multidisciplinares; (iv) estudos *in silico*; (v) síntese e caracterização de novas moléculas; (vi) ensaios *in vitro* e *in vivo*; (vii) avaliação toxicológica básica; (viii) síntese em escala industrial de princípios ativos; (ix) proteção da propriedade intelectual de NCEs e análise do mercado farmacêutico; (x) documentação e formulação da proposta de investigação de NCEs junto ao órgão responsável; entre outras. As fases clínicas (FC) envolvem as seguintes etapas: (i) FC I, na qual são realizados testes de dosagem dos princípios ativos em humanos saudáveis para a avaliação preliminar dos níveis de tolerância e toxidez; (ii) FC II, que envolve testes farmacológicos em um número reduzido de pacientes que apresentam a doença-alvo, para avaliação da dosagem, eficácia e segurança da composição; (iii) FC III, responsável por testes em um número significativo de pacientes em diferentes regiões

geográficas para avaliação das propriedades farmacocinéticas, além de eficácia e segurança; e, (iv) FC IV, que está relacionada com o monitoramento dos efeitos adversos e de longo prazo pós-aprovação do novo fármaco⁹.

O elevado risco do processo de descoberta de fármacos é evidenciado nos cerca de 80% de insucessos de NCEs que atingem as fases de desenvolvimento clínico. Das NCEs eliminadas, 50% não apresentarem propriedades farmacocinéticas apropriadas, além de alta toxicidade, conforme ilustrado na Figura 1.2¹⁹.

As altas taxas de insucesso no processo levaram a um crescente interesse por novas estratégias e métodos capazes de contribuir efetivamente para o processo de otimização de propriedades de candidatos a NCEs, tendo como finalidade a redução do tempo e dos altos custos envolvidos no processo²⁰⁻²¹. Para continuar a produzir NCEs de forma competitiva e sustentável, a indústria farmacêutica vem aumentando a participação de soluções integradas no seu plano de pesquisa básica. Sobretudo, modelos *in silico* (termo usado para estudos gerados com o auxílio de computador) estão sendo empregados de maneira cada vez mais ampla e integrada nos estágios iniciais de P&D para seleção e otimização de moléculas com maior potencial de desenvolvimento. Estes modelos têm o propósito de selecionar um número menor de moléculas a serem avaliadas em ensaios *in vitro* e *in vivo*, reduzindo custos e tempo, e maximizando as chances de identificação de NCEs qualificadas²²⁻²³. Neste contexto, diversas metodologias vêm sendo empregadas para o planejamento de NCEs que apresentem perfil farmacodinâmico e farmacocinético otimizado²⁴⁻²⁶.

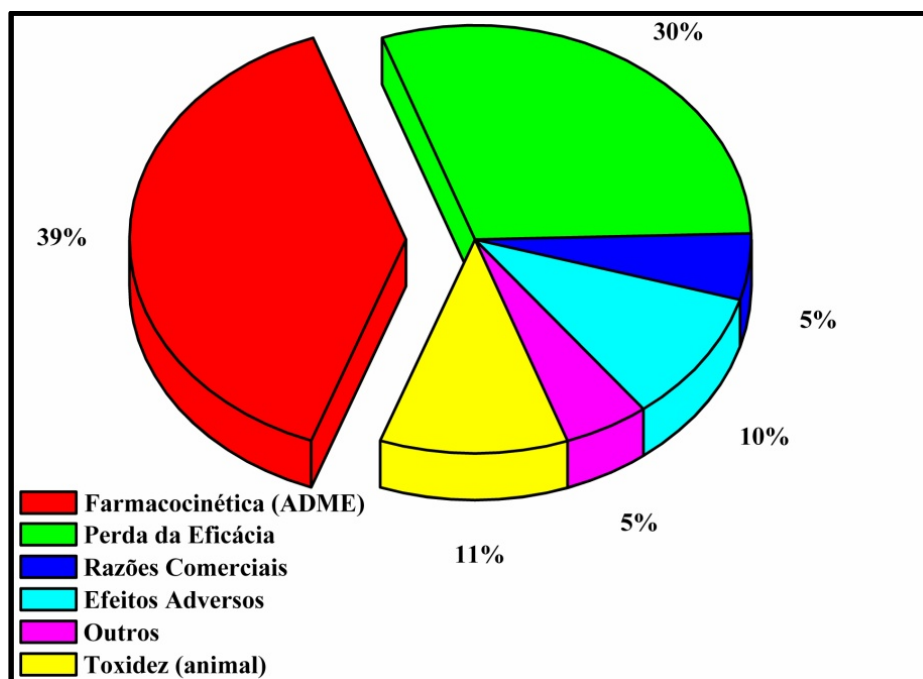


Figura 1.2 – Principais razões de insucesso de NCEs nas fases clínicas de desenvolvimento¹⁹.

1.3 Farmacodinâmica e Farmacocinética

Por muitas décadas, o requisito principal para que uma molécula alcançasse o processo de desenvolvimento de NCEs era a fase farmacodinâmica, ou seja, os candidatos deveriam possuir propriedades otimizadas frente ao alvo biológico, como potência, afinidade e seletividade²⁷. Desta forma, o foco predominante recaía nos estudos das interações intermoleculares responsáveis pelo processo de reconhecimento molecular²⁸. Contudo, para que os efeitos farmacológicos observados *in vitro* resultem na ação terapêutica dos fármacos, *in vivo*, é necessário que os princípios ativos possam estar disponíveis por determinado período de tempo nas vizinhanças dos sítios alvos de ação. Os estudos dos diversos caminhos que o princípio ativo deve percorrer no complexo compartimento biológico do corpo humano, desde a sua administração até atingir a circulação sistêmica para exercer o efeito terapêutico desejado, fazem parte da chamada fase farmacocinética, que determina assim o tempo e o curso do princípio ativo pelo corpo humano¹⁹. Esta fase abrange propriedades conhecidas como ADME, que se referem aos processos farmacocinéticos de absorção (A), distribuição (D),

metabolismo (M) e excreção (E). A concentração de fármaco nos tecidos e o tempo de meia-vida são fatores fundamentais que determinam a dosagem e o regime de administração dos medicamentos para garantir os níveis de concentração do fármaco na faixa terapêutica desejada por um período apropriado. Concentrações abaixo da faixa podem afetar a eficácia do tratamento, ao passo que concentrações acima podem levar a toxidez, colocando em risco saúde do paciente (Figura 1.3). É fundamental a integração de processos de otimização de múltiplas propriedades farmacodinâmicas e farmacocinéticas nos estágios iniciais do planejamento de NCEs³⁰⁻³¹.

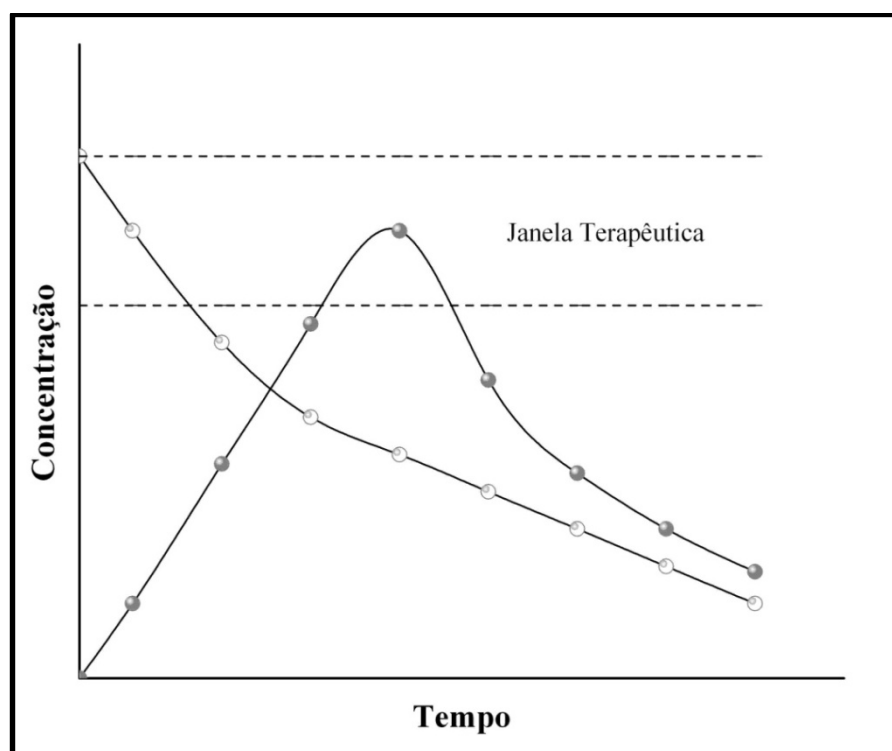


Figura 1.3 – Concentração plasmática vs. tempo: (○) administração intravenosa; (●) administração oral.

Muitos esforços têm sido concentrados em um campo científico considerado importante: a modelagem *in silico* de propriedades de ADME³²⁻³³. As propriedades farmacocinéticas podem ser estudadas através de métodos como os estudos das relações quantitativas entre a estrutura e atividade (do inglês, *quantitative structure-activity relationships* – QSAR), e os estudos das relações quantitativas entre a estrutura e propriedade (do inglês, *quantitative structure-property relationships* – QSPR), entre outros^{19,31}.

1.4 Propriedades Farmacocinéticas

Por questões de mercado, conveniência e segurança, a administração por via oral é a preferencial³⁴⁻³⁵. O tratamento por esta via implica a introdução de um fármaco no organismo (administração), de modo que possa alcançar a circulação sanguínea (absorção) e dirigir-se para o sítio específico de ação (distribuição). O princípio ativo pode ser também biotransformado em outras substâncias químicas (metabolismo), mas deve, por fim, abandonar o organismo (excreção), principalmente pela urina e fezes.

1.4.1 Absorção

Após a administração, o fármaco deve ser dissolvido e solubilizado no trato gastrointestinal para que possa ser absorvido no estômago ou através do intestino¹⁹. A absorção ocorre por dois mecanismos principais. No primeiro, chamado de mecanismo transcelular, ocorre a simples difusão passiva através das membranas e citoplasmas celulares devido à diferença de concentração entre o sítio de absorção e o sangue. No segundo mecanismo de difusão passiva, chamado de paracelular, ocorre a passagem entre as junções intercelulares. Esta via apresenta como limitação o tamanho das substâncias. A absorção pode ocorrer também por transporte ativo, que requer gasto de energia e é mediado por proteínas transportadoras ou carregadores (Figura 1.4). Esse tipo de absorção apresenta algumas características típicas, tais como: (i) o transporte ocorre geralmente contra o gradiente de concentração, (ii) o mecanismo de transporte pode levar a altas concentrações de fármaco (efeito de saturação), (iii) o processo de absorção pode ser afetado pela competição entre duas ou mais estruturas similares³⁶⁻³⁷.

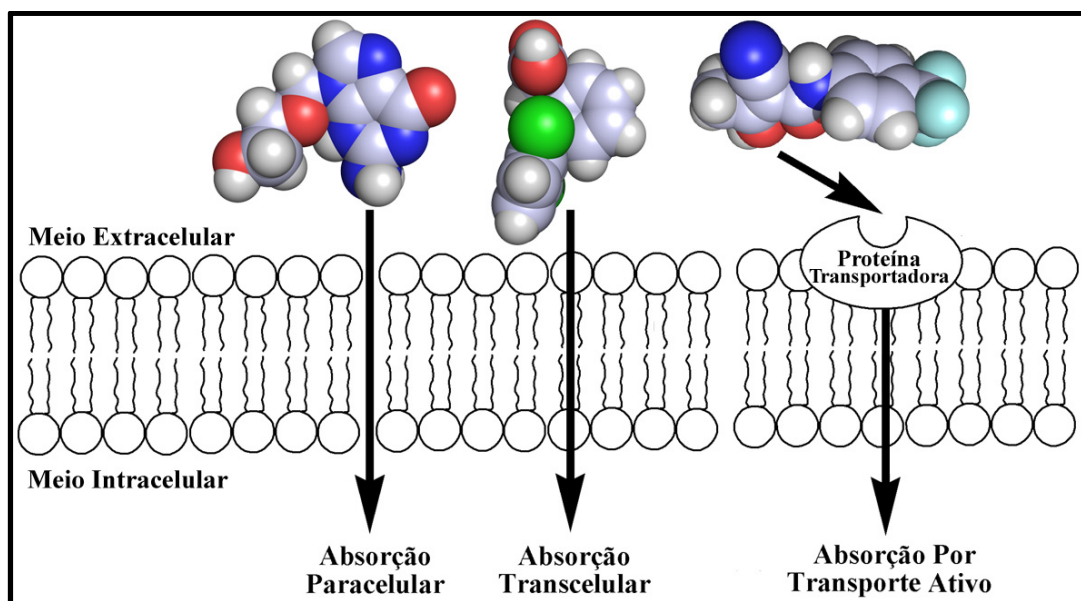


Figura 1.4 – Principais processos de absorção dos fármacos no intestino humano³⁷.

Entre as principais propriedades moleculares que influenciam a absorção de fármacos se destacam solubilidade e lipofilia. O primeiro passo para um fármaco ser absorvido é a desintegração da formulação ingerida, seguida pela dissolução do princípio ativo. Em geral, uma baixa solubilidade (ou inadequada taxa de dissolução) está relacionada à alta lipofilia. Entretanto, moléculas muito hidrofílicas apresentam problemas de permeabilidade e absorção³⁷⁻³⁸. A lipofilia, que pode ser expressa pela medida do coeficiente de partição ($\log P$, partição de uma substância entre as fases aquosa e orgânica comumente representada por sistema octanol/água), representa a passagem do meio extracelular para o meio intracelular através da bicamada lipídica¹⁹. O balanço ideal entre solubilidade e lipofilia é, portanto, um ponto fundamental no planejamento de NCEs.

Diversos estudos foram realizados para investigar o perfil dos fármacos disponíveis no mercado em relação à absorção e permeabilidade. Um destes estudos, desenvolvido por Lipinski e colaboradores, conhecido como a “regra dos 5” (do inglês, “*rule of 5*”), merece destaque especial³⁹.

Este estudo foi baseado em um amplo conjunto de NCEs coletadas do *World Drug Index* (WDI) e estabeleceu de forma pioneira limites simples para absorção e permeabilidade de fármacos. Segundo a “regra dos 5”, existe uma grande probabilidade de compostos candidatos a fármacos apresentarem baixa absorção e

permeabilidade quando dois ou mais dos quatro parâmetros apresentados na Tabela 1.1 (destacados em cinza) são encontrados, por exemplo massa molar maior que 500 e $\log P$ maior que 5. Com o uso da regra de Lipinski é possível a seleção e planejamento de compostos com perfil farmacocinético mais apropriado.

Tabela 1.1 – Limites de propriedades usados para definir parâmetros farmacocinéticos inapropriados⁴³

	Propriedade	NCE	Lider	Fragmento
Regra dos 5	MM	> 500	> 400	> 300
	$\log P$	> 5	> 4	> 3
	HBA	> 10	> 8	> 3
	HBD	> 5	> 4	> 3
Propriedade Adicional	PSA (Å^2)	> 140	> 120	> 90

***Composto Lider**, compostos com propriedades biológicas desejadas, mas não otimizadas. **MM**, massa molar (g/mol). **$\log P$** , coeficiente de partição octanol/água. **HBA**, número de aceptores de ligação hidrogênio. **HBD**, número de doadores de ligação hidrogênio. **PSA**, área de superfície polar (do inglês, *polar surface area*).

A partir destes resultados, diversos outros estudos foram realizados. A caracterização de propriedades moleculares em diferentes coleções de compostos levou a uma série de observações importantes para compostos chamados líderes, os quais tem potencial de desenvolvimento para tornar-se NCEs através de incrementos estruturais e otimização de propriedades biológicas. Esses não poderiam ter massa molar maior que 400 ou $\log P$ maior que 4, entre outras propriedades (Tabela 1.1), para serem considerados na etapas de P&D de fármacos. Também foram observados parâmetros restritivos aos fragmentos ligantes de proteínas, que podem ser otimizados através do aumento de massa molar tornando-se compostos líderes. No caso dos fragmentos esses não poderiam ter massa molar maior que 300 ou $\log P$ maior que 3, entre outras propriedades (Tabela 1.1).

Assim como para novas coleções de compostos, novas propriedades moleculares também foram estudadas além da “regra dos 5” para determinação de limites. Por exemplo, a área de superfície polar (do inglês, *polar surface area* – PSA) foi descoberta como um fator limitante a ser considerado na análise da permeabilidade celular de NCEs, compostos líderes e fragmentos⁴⁰⁻⁴³.

Apesar de sua grande importância, os filtros moleculares devem ser empregados com cautela, pois as restrições moleculares impostas podem levar a não seleção de moléculas promissoras para o desenvolvimento de fármacos de algumas classes, tais como antibióticos, fungicidas e glicosídeos cardioativos, administrados por via oral e absorvidos por transporte ativo.

1.4.2 Distribuição

Após serem absorvidos, os fármacos alcançam a circulação sistêmica e são rapidamente distribuídos devido à alta diferença de concentração que há entre o sítio de absorção e o sangue e tecidos. Porém, a distribuição não é uniforme nos vários tecidos⁴⁴. O termo distribuição se refere à transferência reversível do fármaco de um local a outro do organismo. Os padrões de distribuição de um fármaco refletem alguns fatores fisiológicos, bem como suas propriedades físico-químicas e farmacocinéticas⁴⁵.

Uma das principais características do processo de distribuição é a extensão da ligação dos fármacos às proteínas plasmáticas. Esta é uma das propriedades que afetam a concentração dos fármacos na circulação e, por consequência, tem papel direto na magnitude da ação farmacológica (efeito terapêutico).

1.4.2.1 Ligação às Proteínas Plasmáticas

A ligação às proteínas plasmáticas (do inglês, *plasma protein binding* – PPB) é uma propriedade farmacocinética fundamental para a ação dos fármacos. Ao alcançarem a via sistêmica, os fármacos podem se ligar, em diferentes proporções, às proteínas plasmáticas. Essa ligação é uma medida da afinidade do fármaco pelas proteínas do plasma, especialmente pela albumina e α_1 -glicoproteína ácida⁴⁶. A PPB

possui papel central no transporte de fármacos no organismo humano, afetando a distribuição, a excreção e os seus efeitos farmacológicos e toxicológicos⁴⁷.

As interações dos fármacos com as proteínas plasmáticas ocorrem por processos rápidos e reversíveis. À medida que o fármaco não ligado se difunde dos capilares para os tecidos, mais fármaco ligado deve se dissociar da proteína até o retorno do equilíbrio na circulação sanguínea, caracterizado por concentrações relativamente constantes da forma ligada e não ligada. Dependendo da extensão da ligação, as formas podem variar de 1% a 99,9%. O complexo fármaco-proteína age como um reservatório na corrente sanguínea, retardando a chegada de fármacos aos sítios de eliminação e disponibilizando-os no local de ação⁴⁸⁻⁴⁹. Quando a ligação às proteínas é forte (fração livre $< 0,1$), pode ocorrer uma diminuição na intensidade máxima de ação de uma dose de fármaco, alterando assim sua resposta clínica. Por outro lado, quando a ligação é fraca, pode ocorrer um aumento na intensidade de ação do fármaco. Em grande parte, fármacos que apresentam PPB $> 90\%$ possuem baixa toxidez⁵⁰.

Outro aspecto importante se refere às interações fármaco-fármaco (do inglês, *drug drug interaction* – DDI), representadas pela competição de diferentes fármacos pelos sítios de PPB. Nestes casos, os fármacos de menor afinidade relativa são deslocados aumentando a sua fração livre no plasma. Pode também ocorrer a modificação da estrutura terciária das proteínas plasmáticas após a ligação do fármaco, alterando assim a afinidade da proteína por outros xenobióticos. A diminuição de 1% da PPB de um fármaco que se liga 99% às proteínas plasmáticas, terá duplicada a fração livre do mesmo, podendo levar a variações significativas na magnitude da ação farmacológica. Nos casos em que ocorre um aumento da concentração livre de um dos fármacos em um nível superior ao limite da faixa terapêutica (Figura 1.3), devido a DDI, podem ser observadas manifestações de efeitos adversos e sinais de toxicidade aguda.

1.4.2.2 Volume de Distribuição

O volume de distribuição (V_d) é um parâmetro farmacocinético importante a ser considerado quando o receptor do fármaco não se encontra na via sistêmica, mas em outros tecidos. O V_d depende de vários fatores como a taxa de fluxo sanguíneo em diferentes tecidos, lipofilia do fármaco, pH e afinidade por diferentes tecidos. Esse parâmetro é usado para quantificar a distribuição de um fármaco pelo corpo após administração oral ou parenteral, e é frequentemente proporcional a massa total do organismo.

O V_d é definido como o volume teórico no qual uma determinada quantidade de princípio ativo precisaria ser uniformemente distribuído para produzir a concentração sanguínea observada. Essa quantidade é determinada a partir da dose inicial administrada por via intravenosa comparada com a concentração de fármaco encontrada na corrente sanguínea, após o equilíbrio entre o sangue e outros tecidos ter ocorrido⁵¹. A Tabela 1.2 mostra alguns valores estimados de V_d que informam os tecidos por onde os fármacos devem estar distribuídos.

Tabela 1.2 – Estimativas relacionadas ao volume de distribuição para adultos sem anomalias quanto ao peso ou estatura⁵¹

V_d (L)	Porcentagem em relação ao peso corporal (%)	Tecidos abrangidos
5	4.3	Distribuição somente pela via sistêmica
10-20	15-27	Distribuição pelo fluido extracelular
25-30	35-42	Distribuição pelo fluido intracelular
40	60	Todos os fluidos corporais

1.4.3 Metabolismo e Excreção

O processo de eliminação inicia no momento da absorção do princípio ativo, que pode ser metabolizado ou sofrer efluxo do organismo de forma inalterada. O metabolismo é um processo químico através do qual o organismo altera os xenobióticos, incluindo os fármacos, preparando-os para a eliminação. Embora o fígado seja o principal responsável pelo metabolismo dos fármacos, esse processo pode ocorrer também nos pulmões, intestino, barreira hematoencefálica e rins⁵².

Após serem absorvidos no intestino, os fármacos entram na via hepática até o fígado onde podem sofrer metabolismo de primeira passagem (Figura 1.5). As biotransformações nas estruturas químicas dos princípios ativos têm o objetivo de aumentar a solubilidade dos metabólitos, facilitando o processo de excreção primário. Existem várias vias de eliminação, como suor, lágrimas, respiração e fezes, sendo a principal delas, a urina⁵¹.

O metabolismo de fármacos ocorre em duas fases distintas bem caracterizadas. A primeira fase (fase 1) engloba reações catalisadas pela superfamília de isoformas do citocromo P450 (abreviado CYP; ou com menor frequência CYP450). Nesta fase ocorrem, predominantemente, reações de oxidação, redução, hidroxilação e hidrólise, levando a metabólitos mais polares. Neste processo, os metabólitos gerados podem ser inativos ou ativos. Os metabólitos ativos podem apresentar maior ou menor potência relativa, ou ainda, toxidez variada em comparação ao princípio ativo. Os pró-fármacos são um caso especial, onde o fármaco é administrado em uma forma inativa (pró-fármaco) que sofrerá biotransformação para liberar o princípio ativo (fármaco), o qual exercerá a ação farmacológica através da modulação do alvo apropriado. O processo de biotransformação pode ocorrer durante a absorção pela membrana intestinal, durante a passagem pelo fígado (metabolismo de primeira passagem) ou no metabolismo sistêmico (Figura 1.5)^{50,52-53}. A segunda fase (fase 2) compreende reações de conjugação, onde os produtos formados na fase 1 são associados covalentemente com substâncias como a glutatona, ácido glicurônico, sulfato ou certos aminoácidos, formando conjugados mais hidrossolúveis, geralmente inativos, que são eliminados

preferencialmente pela urina ou, então, excretados pela bile e eliminados nas fezes^{28,38,54-55}. O modo de distribuição e eliminação de metabólitos pode diferir daquele dos fármacos originais, como resultado da adição de grupos funcionais que modificam as propriedades moleculares e físico-químicas dos metabólitos. A sigla CYP é utilizada de forma genérica para identificar uma superfamília relacionada de enzimas oxidativas (EC 1.14). Na fase 1, estas enzimas têm como função principal a incorporação de um átomo de oxigênio no substrato orgânico. Sete das 57 isoformas humanas conhecidas são responsáveis pelo metabolismo de mais de 90% dos fármacos em uso clínico. Estas isoformas são: CYP1A2, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1 e CYP3A4⁵⁶. Dentre estas, as três principais, CYP3A4, CYP2D6 e CYP2C9 (Figura 1.6), são responsáveis pelo metabolismo de cerca de 70% dos fármacos, sendo a CYP3A4, a forma mais abundante no fígado e responsável pelo metabolismo de aproximadamente 50% dos fármacos^{36,57-58}. Embora com papel menos importante no metabolismo de fármacos, as diferentes isoformas do CYP também estão presentes em outros órgãos do corpo, como pulmões, rins, intestino e barreira hematoencefálica¹⁹.

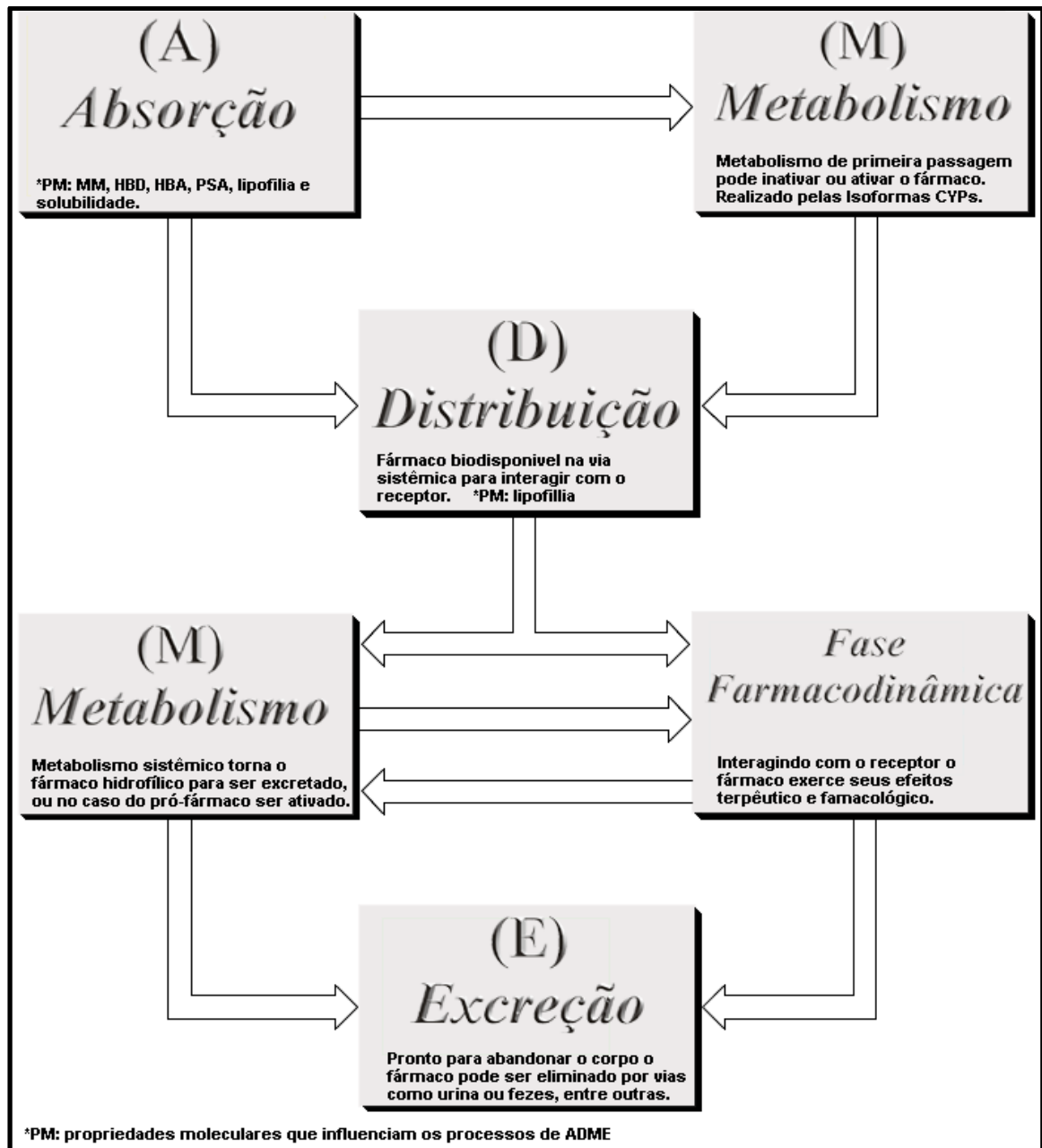


Figura 1.5 – Caracterização das propriedades de ADME na passagem do fármaco pelo organismo humano.

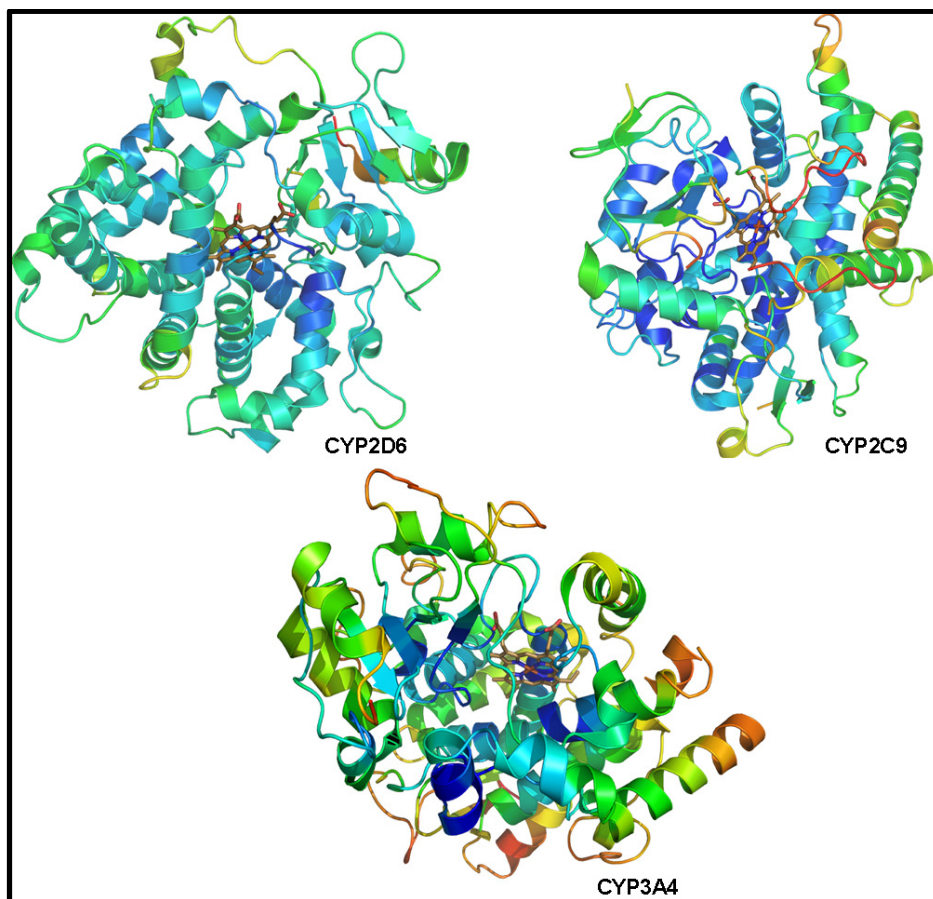


Figura 1.6 – Estrutura monomérica das isoformas CYP2D6, CYP2C9 e CYP3A4 com o grupo heme ao centro (PDB IDs 2F9Q, 1R9O e 1W0E, respectivamente).

Uma grande diversidade de pequenas moléculas (massa molar entre 150 e 1.250 g/mol) se liga à cavidade de interação da isoforma CYP3A4, cujo volume considerável permite a sua ocupação de várias maneiras⁵⁹. Diversos fármacos são metabolizados pela CYP3A4, como a atorvastatina (Lípitor[®]), cimetidina (Tagamet[®]), diazepam (Valium[®]), lovastatina (Mevacor[®]), omeprazol (Losec[®]), ritonavir (Norvir[®]), sildenafil (Viagra[®]) e triazolam (Halcion[®])⁵⁴. A atorvastatina é um exemplo de fármaco extensivamente metabolizado, sendo que apenas 14% da dose inicialmente administrada por via oral (100%) alcançam a circulação sistêmica na forma intacta. A outra fração (86%) é biotransformada através do metabolismo de primeira passagem em metabólitos ativos ou inativos. Seus metabólitos ativos, derivados orto- e para-hidroxilados e outros produtos de beta-oxidação, são responsáveis por aproximadamente 70% da atividade inibitória circulante (Figura 1.7)⁶⁰.

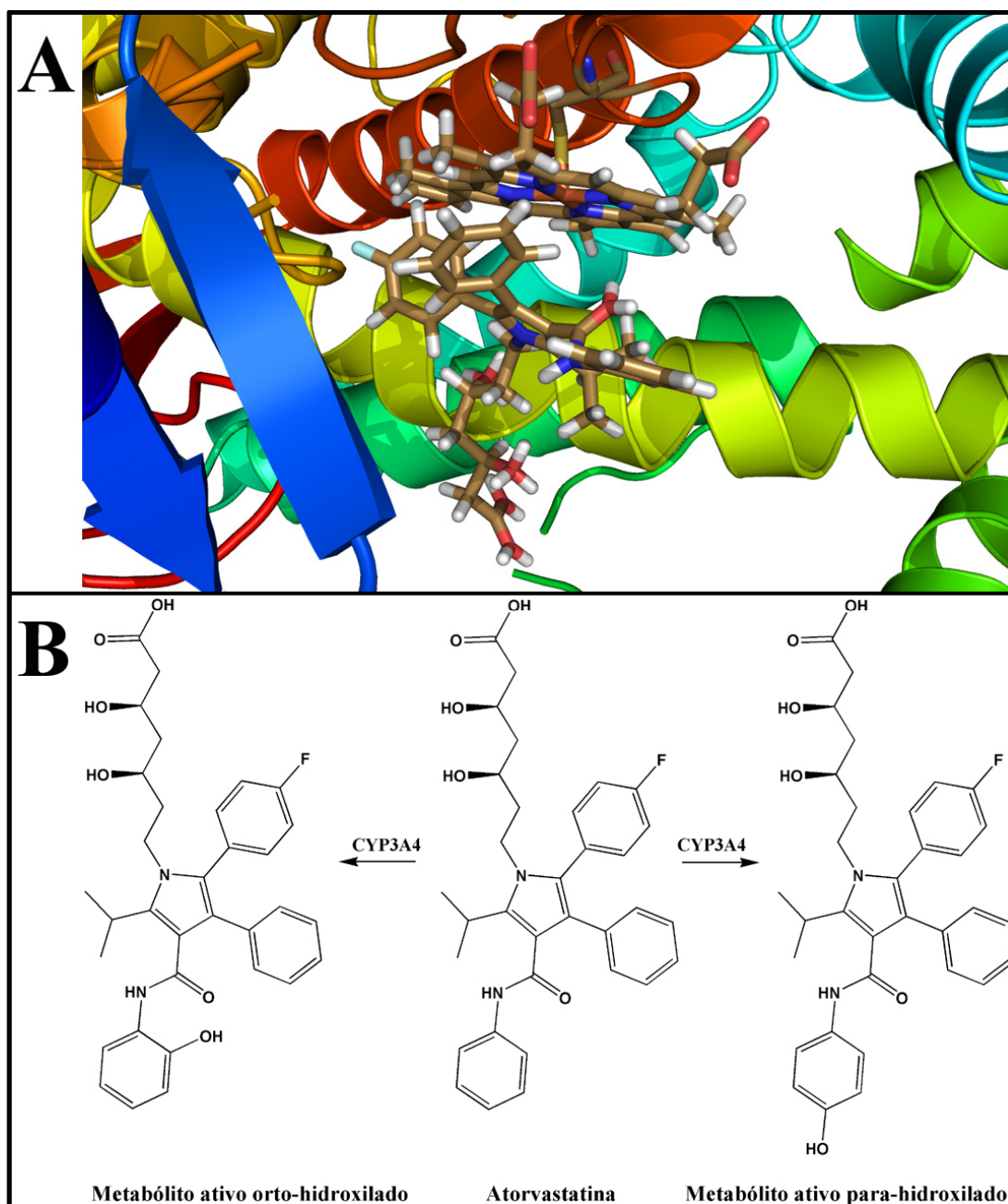


Figura 1.7 – (A) Modelo de interação da atorvastatina com o grupo heme da CYP3A4 (PDB ID 2J0D). (B) Geração de metabólitos ativos da atorvastatina, derivados orto e para-hidroxiados.

Os fármacos interagem com as isoformas CYP de três formas principais: (i) como substratos, sendo susceptíveis as reações de oxidação, hidroxilação, redução e hidrólise, levando a metabólitos mais facilmente excretados pelo organismo; (ii) como inibidores, impedindo o metabolismo de substratos; ou (iii) como indutores, estimulando o metabolismo de substratos⁵⁴. O metabolismo tem importante espaço nas

pesquisas da área farmacêutica e tem passado por sensíveis avanços científicos e tecnológicos⁶¹⁻⁶⁴.

1.4.4 Biodisponibilidade

A biodisponibilidade oral (do inglês, *oral bioavailability*) é uma das propriedades de ADME mais importantes. A biodisponibilidade absoluta em humanos é uma medida da fração do fármaco que alcança a via sistêmica em comparação com a dose administrada por via oral, sendo expressa em valores percentuais (%). Quando um fármaco é administrado por via intravenosa, sua biodisponibilidade é considerada 100% (Figura 1.3)^{50,52}.

A indústria farmacêutica tem como objetivo principal o desenvolvimento de NCEs que possam ser administradas por via oral, portanto, a biodisponibilidade torna-se um parâmetro de grande valor nas estratégias de P&D de fármacos. A intensidade da biodisponibilidade oral varia consideravelmente de acordo com a diversidade química e a classe terapêutica, sendo afetada por diversos processos que têm início na administração do fármaco, tais como desintegração e dissolução da formulação, solubilidade do princípio ativo, absorção intestinal e metabolismo de primeira passagem mediado pela CYP (Figura 1.5)^{19,34-35}.

1.4.5 Barreira Hematoencefálica

A permeabilidade da barreira hematoencefálica (do inglês, *blood brain barrier* – BBB) é uma propriedade fundamental para os fármacos que atuam no sistema nervoso central (SNC), no tratamento de doenças como epilepsia, mal de Alzheimer e esquizofrenia. Por outro lado, encontra também importância no desenvolvimento de fármacos que não atuam no SNC, pois os mesmo devem apresentar limitada capacidade de transpor a BBB, minimizando assim, possíveis efeitos adversos⁶⁵.

A BBB é composta por células endoteliais unidas nos capilares cerebrais (Figura 1.8) que tem como objetivo proteger o cérebro de substâncias químicas presentes no sangue. Sua característica semipermeável permite que algumas substâncias a atravessem e outras não. Os capilares cobrem aproximadamente 95% da área total da barreira entre o sangue e o cérebro, atingindo um comprimento estimado de 650 km e área superficial entre 10 e 20 m². Toda esta complexa rede de capilares torna o cérebro pouco acessível à entrada de moléculas polares, as quais requerem mecanismos especiais envolvendo proteínas transportadoras. Moléculas com área de superfície polar < 90 Å² e massa molar < 450 apresentam maior probabilidade de transpor a BBB⁶⁵⁻⁶⁷.

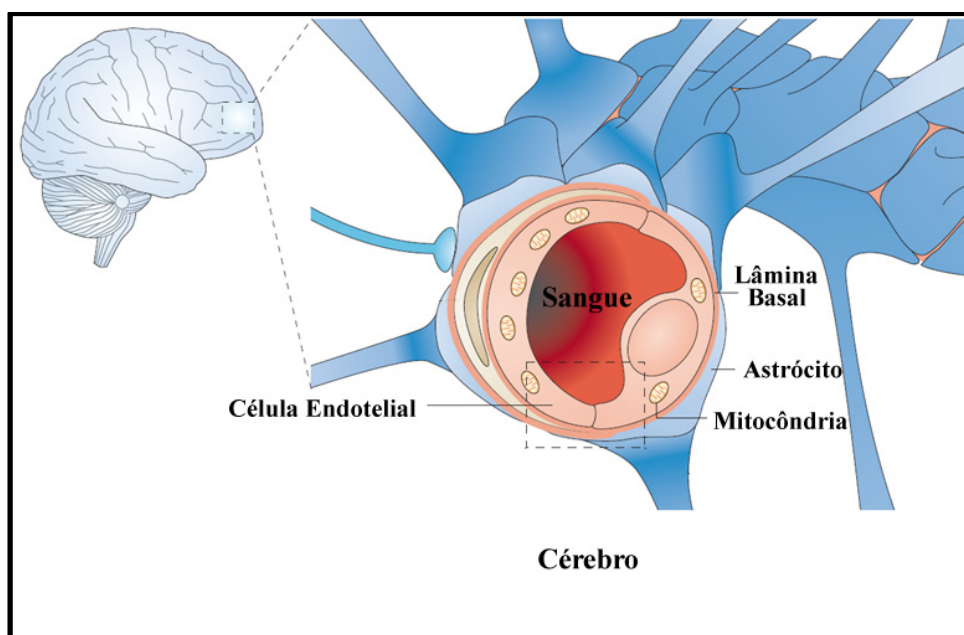


Figura 1.8 – Componentes da barreira hematoencefálica⁶⁶.

1.5 Solubilidade

As propriedades físico-químicas afetam diretamente o perfil farmacocinético dos fármacos e têm contribuído na otimização de NCEs. A solubilidade é uma das mais importantes, possuindo papel central na relação com outras propriedades físico-químicas (Figura 1.9)⁴³⁻⁶⁸.

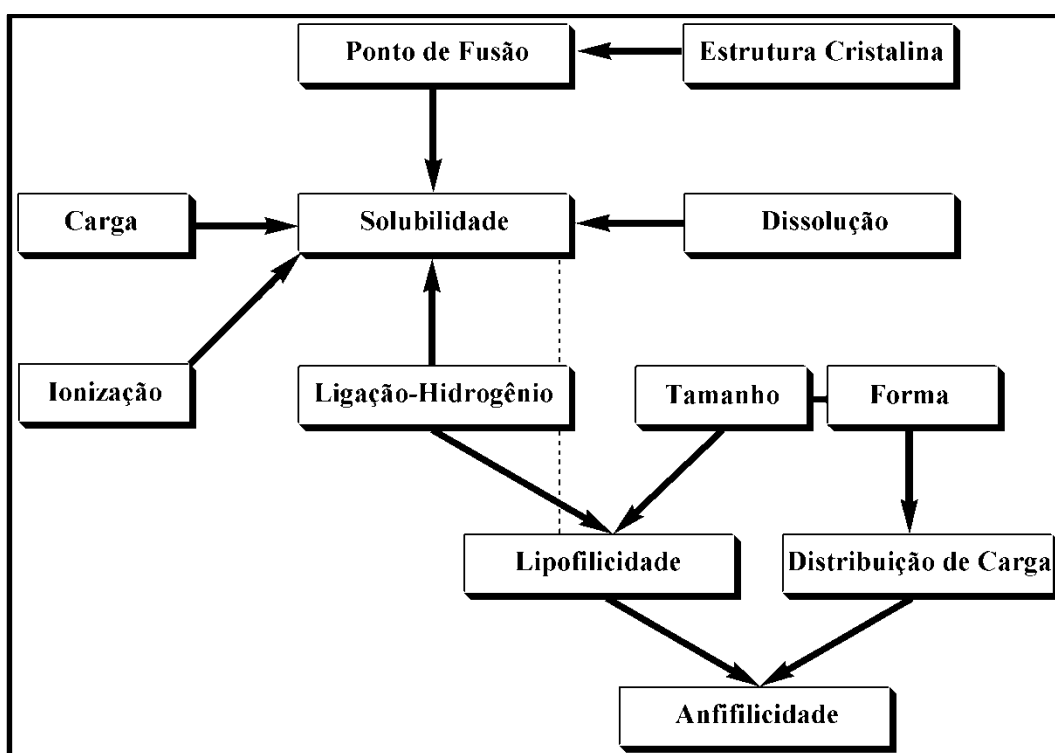


Figura 1.9 – Relações entre propriedades físico-químicas (solubilidade com papel central)⁴³.

O processo de absorção inicia com a desintegração e dissolução do fármaco^{34,45}. Uma baixa solubilidade prejudica a absorção e permeabilidade do princípio ativo, sendo, portanto, fundamental o estudo desta propriedade nos estágios iniciais do processo de pesquisa básica. A biodisponibilidade oral dos fármacos é afetada diretamente pela combinação dos processos de solubilidade, dissolução, absorção e permeabilidade. Em geral, baixa solubilidade está relacionada com alta lipofilia que, por sua vez, está relacionada ao metabolismo. Fármacos lipofílicos são mais susceptíveis ao metabolismo mediado pelo CYP, adquirindo maior caráter hidrofílico^{45,69}.

1.6 Glicoproteína-P e os Processos de ADME

As propriedades de ADME de candidatos a fármacos são influenciadas por fatores intrínsecos, como propriedades moleculares e físico-químicas, ou fatores extrínsecos, como composição de membranas biológicas, proteínas metabolizadoras, proteínas transportadoras, entre outros. Um desses fatores é a glicoproteína-P (P-gp), que é uma proteína transportadora ATP-dependente⁷⁰. A P-gp funciona como uma bomba de efluxo (Figura 1.10), retirando substratos para fora das células com o importante papel biológico de proteger o organismo contra substâncias nocivas. A expressão da P-gp é detectada em vários tecidos humanos, como o epitelial do fígado, rins, intestino delgado e grosso e nas células endoteliais dos capilares do cérebro⁷⁰⁻⁷¹. O efluxo promovido pela P-gp interfere nos processos de ADME, desde a absorção intestinal, retornando os compostos em fase de absorção para o lúmen intestinal, até a permeabilidade da barreira hematoencefálica, impedindo desta forma que as substâncias atinjam o sítio de ação no SNC⁷²⁻⁷⁴.

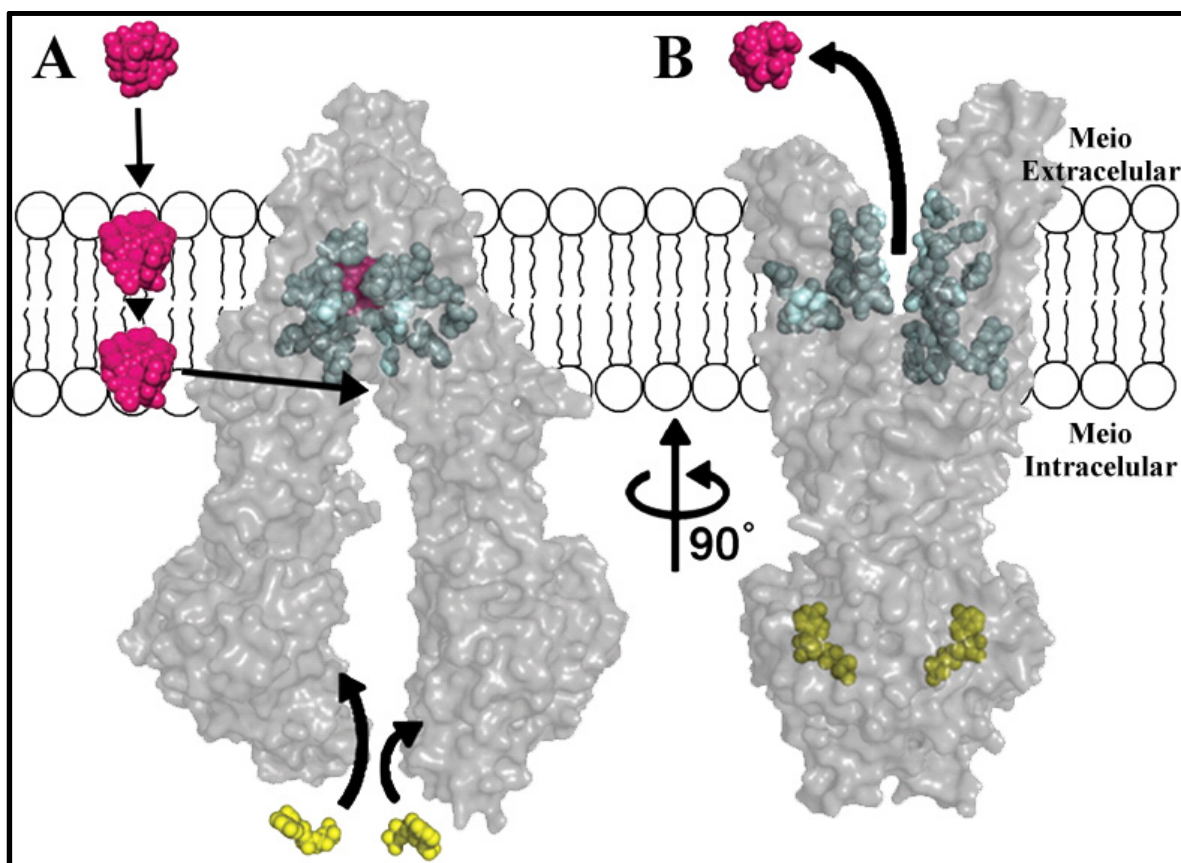


Figura 1.10 – Modelo de transporte de substrato pela P-gp. (A) O substrato (magenta) permeia pela bicamada lipídica entrando no sítio de ligação aberto da P-gp. (B) ATP (amarelo) liga-se a P-gp causando mudanças conformacionais substanciais e expondo o substrato para a parte extracelular, ocorrendo assim o processo de efluxo⁷⁸.

Proteínas transportadoras (em especial a P-gp) são super expressadas em tumores de sistemas, como o de colo, rins, carcinoma adrenal, pulmões e no carcinoma gástrico. A P-gp reconhece e transporta compostos químicos com grande diversidade estrutural, em um fenômeno conhecido como resistência múltipla a fármacos (do inglês, *multidrug resistance* – MDR), responsável por grande parte dos insucessos de quimioterápicos no tratamento do câncer⁷⁵⁻⁷⁶. A inibição da P-gp pode levar a melhores índices de absorção intestinal, biodisponibilidade e permeabilidade da barreira hematoencefálica, além de reduzir os índices de MDR em casos de ação quimioterápica contra tumores. Inibidores da P-gp, quando administrados em conjunto com fármacos de classes terapêuticas variadas, possuem papel importante na melhoria dos índices farmacocinéticos e de MDR⁷⁷⁻⁷⁸. Um caso típico verifica-se no estudo clínico do GF 120918, que apresenta alta atividade inibitória da P-gp sendo capaz de

umentar em mais de quatro vezes o nível plasmático do topotecan (Hycamtin[®])⁷⁹. A identificação de moduladores seletivos da P-gp é de grande valor no desenvolvimento de NCEs na terapia do câncer.

1.7 Modelagem *In Silico* de Propriedades Farmacocinéticas

Existe grande interesse em química medicinal no desenvolvimento de modelos *in silico* de ADME que possam auxiliar na seleção de candidatos promissores. A indústria farmacêutica busca intensamente a integração efetiva de modelos de ADME *in silico*, *in vitro* e *in vivo*, criando um moderno paradigma em P&D de fármacos (Figura 1.11)^{24-26,31,80}.

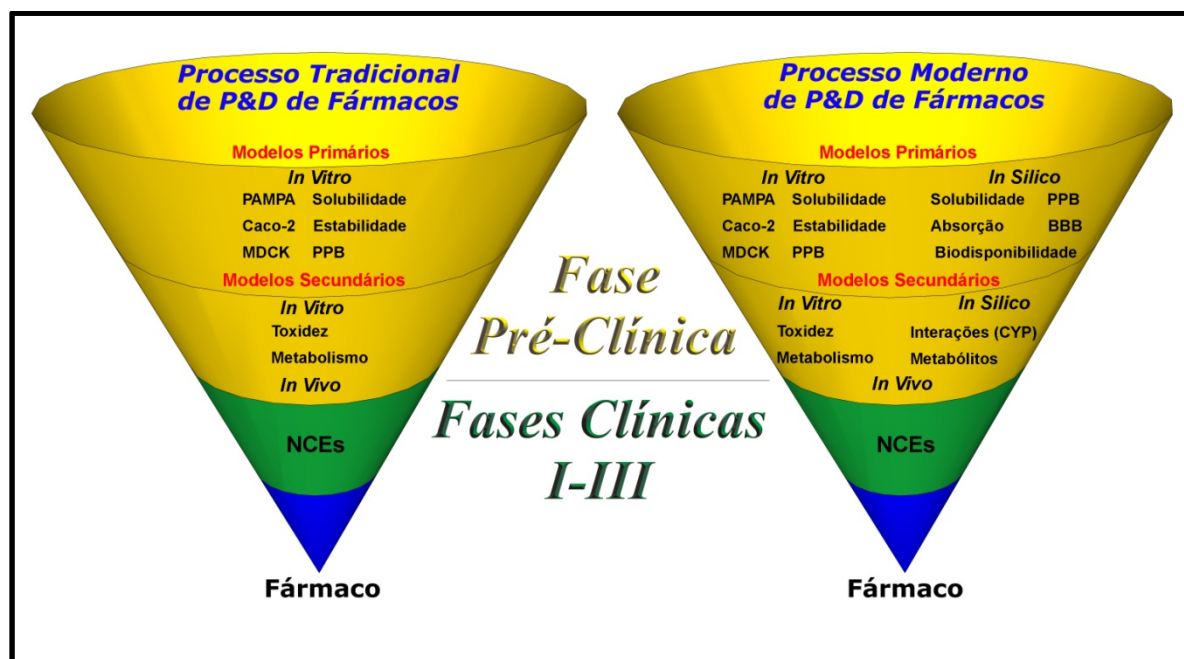


Figura 1.11 – Integração de modelos *in silico*, *in vitro* e *in vivo* de propriedades de ADME no processo de P&D de fármacos.

O emprego de uma variedade de tecnologias de larga escala, principalmente por parte da indústria farmacêutica, possibilitou a geração de uma quantidade significativa de dados de propriedades farmacocinéticas^{5,81-83}. Tendo em vista a necessidade do gerenciamento e organização das informações, diversas bases de dados comerciais, ou

de acesso livre, como por exemplo, a base PK/DB (*Database for Pharmacokinetic Properties*), foram criadas para estudos de propriedades de ADME⁸⁴⁻⁸⁶. A partir destas iniciativas, a modelagem *in silico* tem evoluído constantemente, com progressos sensíveis nas áreas de correlação de dados, predição de propriedades, simplificação e automatização de processos e tarefas, entre outros aspectos relevantes (Figura 1.12).

Algoritmos descritivos e técnicas matemáticas de correlação são ferramentas importantes no entendimento e predição de interações que unem química e biologia. As técnicas de QSAR/QSPR estão entre as mais relevantes na avaliação do perfil do espaço químico-biológico de pequenas moléculas⁸⁷. Além disso, os modelos *in silico* apresentam a clara vantagem de poderem ser usados em todas as etapas do processo de P&D de fármacos, atendendo a demanda de larga escala da indústria farmacêutica, simplificando os custos e o tempo do processo^{19,80-81}.

Em particular, métodos de QSAR têm sido empregados na geração de modelos para diversas propriedades farmacocinéticas, tais como biodisponibilidade oral e ligação às proteínas plasmáticas (Figura 1.12)^{47,61,88-93}. A tecnologia de QSAR, com sua rica história de cinco décadas, continua a se renovar e reinventar, incorporando novos paradigmas e aplicações importantes na área de planejamento de fármacos. QSAR é uma técnica que emprega descritores baseados em estruturas moleculares e utiliza algoritmos computacionais para relacionar os descritores obtidos com o valor da propriedade alvo de interesse⁹⁴. Os métodos de QSAR buscam identificar e quantificar as relações predominantes no complexo campo da modelagem, representadas pela estrutura química e atividade biológica, ou ainda, em um paradigma mais moderno, pela estrutura química e propriedade farmacocinética (QSAR-ADME)⁹⁵⁻⁹⁷.

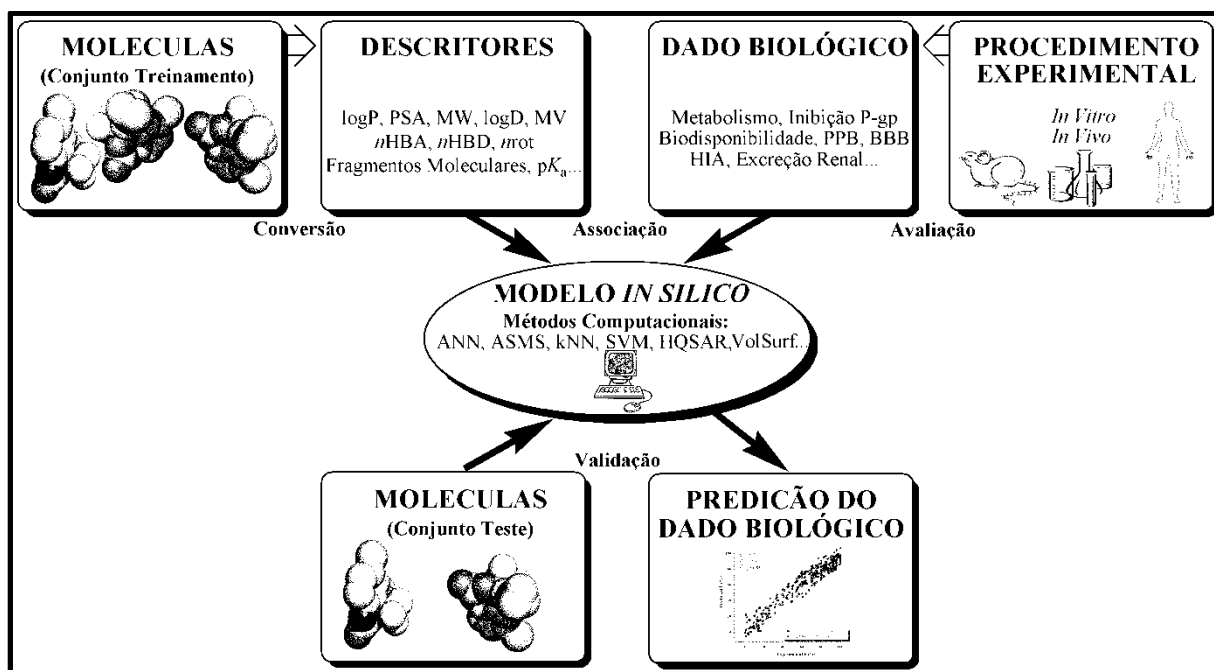


Figura 1.12 – Estrutura e função de modelos *in silico*.

No campo da modelagem de propriedades farmacocinéticas, o emprego de métodos de QSAR é especialmente atrativo devido ao grande volume de dados de ADME disponível na literatura. Embora vários progressos notáveis tenham sido alcançados na área de correlação de dados, o maior desafio da tecnologia de QSAR-ADME é a criação de modelos capazes de prever propriedades farmacocinéticas de NCEs ainda não sintetizadas^{90,31}.

Neste trabalho de tese de doutorado foi empregado de forma inédita um método de fragmentos moleculares (holograma QSAR, HQSAR) para a geração de modelos preditivos de QSAR-ADME para diversas propriedades, tais como absorção intestinal humana, metabolismo de primeira passagem, permeabilidade da barreira hematoencefálica, inibição da glicoproteína-P e solubilidade em água. Os modelos desenvolvidos, assim como os dados farmacocinéticos coletados, foram disponibilizados para acesso livre através da internet na base de dados PK/DB (www.pkdb.ifsc.usp.br). Entre outros serviços, a base oferece a possibilidade de busca de compostos por estrutura, subestrutura ou nome, de forma simples e rápida. Uma plataforma de predição de propriedades para novos compostos também foi disponibilizada, utilizando os modelos *in silico* desenvolvidos nesta tese.

1.7.1 Holograma QSAR

O método holograma QSAR emprega hologramas moleculares originários da fragmentação molecular 2D. Os hologramas definem a combinação de diferentes tipos de fragmentos que dão origem aos descritores. Os descritores moleculares codificam todos os fragmentos possíveis: lineares, ramificados e sobrepostos. Informações 3D adicionais, como hibridização e quiralidade, são também codificadas nos hologramas moleculares. Os modelos de QSAR são desenvolvidos empregando-se o método de regressões multivariadas por mínimos quadrados parciais (do inglês, *Partial Least Squares* – PLS) e técnicas de validação cruzada.

HQSAR é uma técnica na qual não há necessidade da geração de conformações bioativas e o alinhamento molecular tridimensional. Somente a estrutura 2D e o valor da propriedade alvo são requeridos para a aplicação deste método^{47,88}. Durante o processo de modelagem de HQSAR, vários parâmetros que afetam diretamente a geração, e por conseqüência, a qualidade dos modelos, podem ser controlados. Uma representação do processo envolvido na geração de hologramas moleculares no método HQSAR é apresentada na Figura 1.13.

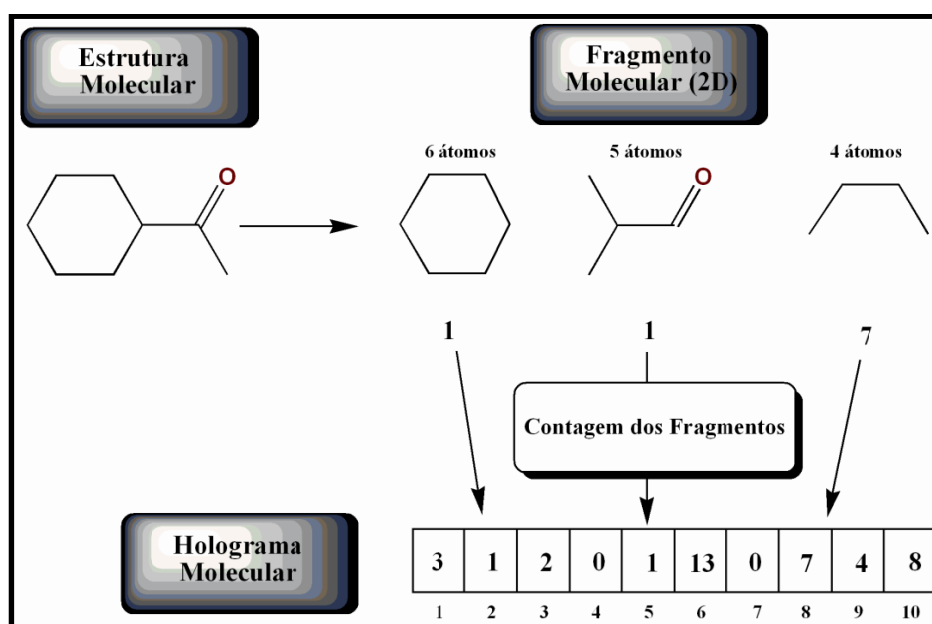


Figura 1.13 – Esquema ilustrativo do processo de obtenção de hologramas moleculares.

Entre os parâmetros mais importantes estão: comprimento do holograma, distinção de fragmentos e tamanho dos fragmentos. O comprimento do holograma determina o padrão de ocupação específico dos fragmentos. O tamanho dos fragmentos pode ser controlado através do tamanho mínimo (M) e máximo (N) de átomos impostos aos fragmentos a serem gerados. A maneira como os fragmentos são distribuídos nas caixas do holograma tem influência direta na qualidade dos modelos.

A distinção de fragmentos é um parâmetro ajustável de fundamental importância. Ele define o tipo de informação molecular 2D que será incluída na criação dos descritores. A Tabela 1.3 apresenta os diferentes tipos de distinção de fragmentos que podem ser considerados no método HQSAR. Os fragmentos gerados são relacionados com a propriedade alvo através do método PLS. Dessa forma, o método de HQSAR identifica padrões moleculares (fragmentos, subestruturas) relevantes para a propriedade alvo em estudo. Além disso, a incorporação progressiva de informações sobre cada fragmento e cada um de seus constituintes, codifica informações sobre a natureza tridimensional das interações presentes^{47,88}.

Tabela 1.3 – Parâmetros de distinção de fragmentos no método HQSAR

Distinção de Fragmentos	Definição	Representação Molecular
Átomo (A)	Permite que os fragmentos sejam distinguidos com base nos tipos de átomos	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo.</p>
Ligação (B)	Permite que os fragmentos sejam distinguidos com base nos tipos de ligações: simples, duplas, triplas e aromáticas	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo. Há uma ligação dupla entre o N e o C da esquerda, uma ligação tripla entre o N e o C da direita, e uma ligação simples entre o N e o C da esquerda.</p>
Conectividade (C)	Permite que os fragmentos sejam diferenciados baseando-se no estado de hibridização dos átomos contidos no fragmento	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo. Cada átomo de carbono tem linhas representando ligações para outros átomos fora do fragmento, indicando diferentes estados de hibridização.</p>
Hidrogênio (H)	A distinção dos fragmentos é feita com base na presença e número de átomos de hidrogênio	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo. Cada átomo de carbono tem um ou dois átomos de hidrogênio (rotulados 'H') conectados a ele.</p>
Quiralidade (Ch)	Permite que os fragmentos sejam distinguidos com base na presença de centros estereogênicos	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo. Cada átomo de carbono tem um ou dois átomos de hidrogênio (rotulados 'H') conectados a ele. Um dos átomos de hidrogênio no carbono da esquerda está representado com uma linha em cunha, indicando quiralidade.</p>
Doador e Aceptor (DA)	Esta opção permite que os fragmentos sejam distinguidos com base na presença de átomos aceptores ou doadores de ligações hidrogênio	<p>Um diagrama de um fragmento molecular com cinco átomos. Um átomo centralizado é rotulado 'N'. Quatro outros átomos, rotulados 'C', estão conectados ao átomo de nitrogênio: um está diretamente acima, um à esquerda, um à direita e um diretamente abaixo. Cada átomo de carbono tem um ou dois átomos de hidrogênio (rotulados 'H') conectados a ele. Um dos átomos de hidrogênio no carbono da esquerda está representado com uma linha em cunha, indicando quiralidade. Um asterisco (*) está colocado ao lado do átomo de nitrogênio, indicando que ele é um átomo doador/aceptor de ligação de hidrogênio.</p>

Um esquema geral das etapas envolvidas na construção de um modelo de QSAR é ilustrado na Figura 1.14. Inicialmente as moléculas do conjunto treinamento são fragmentadas, e em seguida os fragmentos são alocados no holograma molecular. Então a propriedade estudada é adicionada a matriz que da origem aos modelos QSAR analisados via PLS e posteriormente validados através da predição de um conjunto teste.

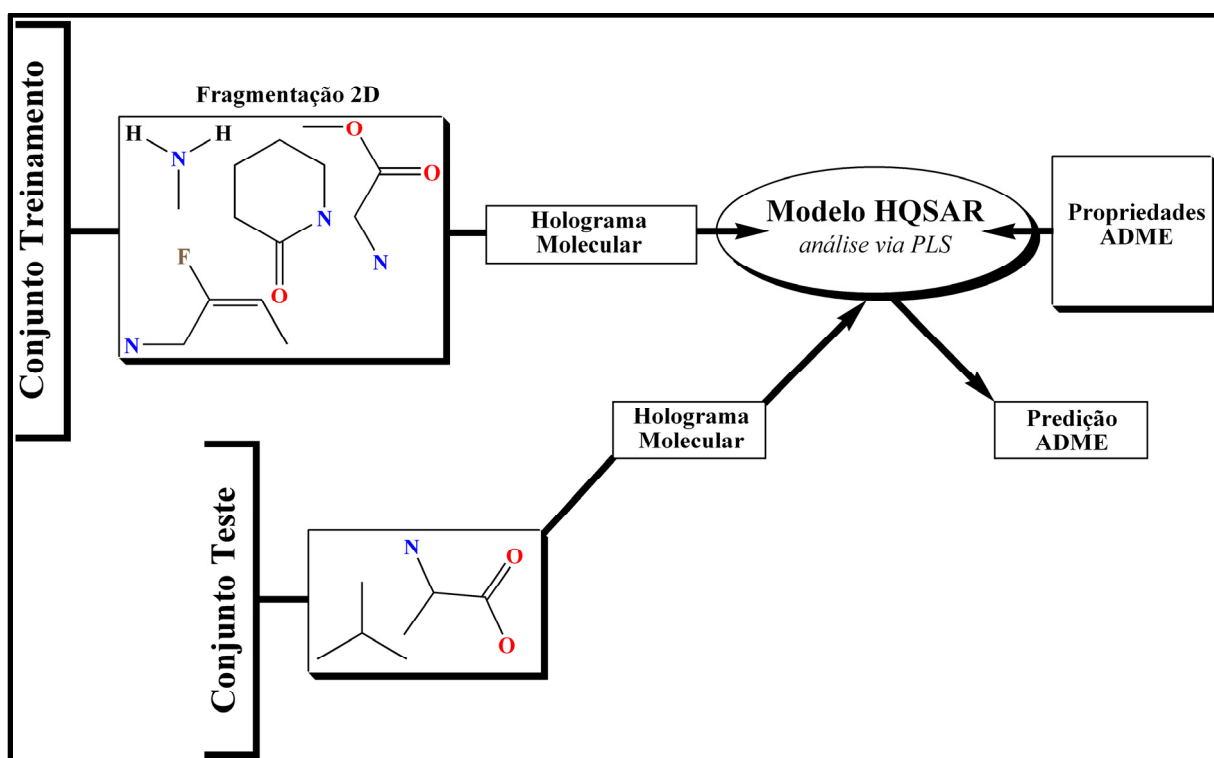


Figura 1.14 – Etapas envolvidas no método QSAR.

CAPÍTULO 2

Objetivos

2 Objetivos

Os objetivos fundamentais desta tese de doutorado são:

- ✓ Organizar amplos conjuntos padrões de dados de propriedades farmacocinéticas e físico-químicas, incluindo absorção intestinal humana, metabolismo de primeira passagem, permeabilidade da barreira hematoencefálica, inibição da glicoproteína-P e solubilidade aquosa.
- ✓ Desenvolver modelos de QSAR-ADME empregando a técnica holograma QSAR.
- ✓ Avaliar a consistência interna e externa dos modelos através do seu poder de correlação nos conjuntos treinamento e predição nos conjuntos teste.
- ✓ Determinar a robustez dos modelos e a sua aplicação como ferramenta útil em Química Medicinal para a identificação e seleção de candidatos a fármacos.
- ✓ Disponibilizar na internet com acesso livre e gratuito informações farmacocinéticas a comunidade científica brasileira e mundial através de uma base de dados.

CAPÍTULO 3

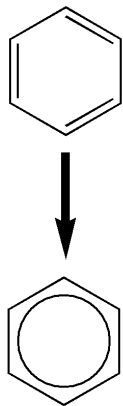
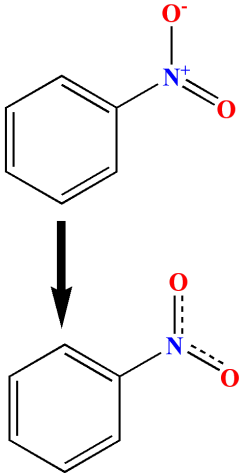
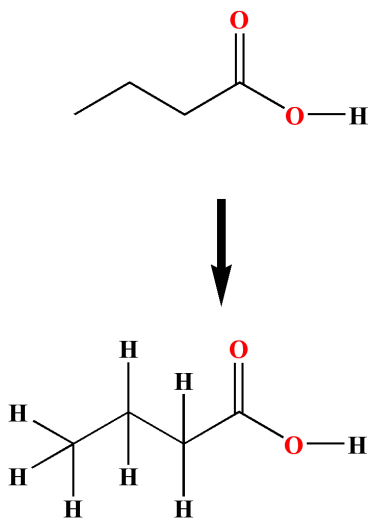
**Materiais e
Métodos**

3 Materiais e Métodos

3.1 Conjuntos Padrões de Dados

A seleção de conjuntos padrões de dados (ou bases de dados) é fundamental para o desenvolvimento de modelos *in silico* de alto poder preditivo. O princípio é encontrar padrões moleculares em conjuntos de compostos que se situam em regiões características do espectro da propriedade estudada (*i.e.* características moleculares semelhantes de compostos que apresentem alta absorção intestinal). Para que seja possível a associação correta entre os valores de propriedade e os padrões encontrados é necessário que as bases sejam corretamente curadas tanto no trabalho computacional, de criação virtual, quanto no trabalho experimental, na medida precisa dos valores atribuídos as propriedades de cada composto. Também como parte do processo de preparação dos conjuntos de dados há a normalização das estruturas que foram consideradas na forma neutra⁹⁸. A Tabela 3.1 apresenta algumas opções usadas neste trabalho de doutorado para normalizar os dados coletados.

Tabela 3.1 – Representação de algumas opções de normalização

Aromatização	Nitro-Ressonância	Adição de Hidrogênio
		

Neste trabalho, foram organizados cinco conjuntos padrões de dados, sendo quatro de propriedades farmacocinéticas e um de propriedade físico-química. Todas as bases foram curadas e normalizadas como descrito acima.

3.1.1 Conjunto de Dados de Absorção Intestinal Humana

Com uma diversidade química bastante considerável, representada por 638 compostos, sendo a maioria fármacos, um conjunto de dados foi organizado para valores de absorção intestinal humana, por forma de difusão passiva^{85,99-106}. Os valores para a modelagem molecular são expressos na forma de porcentagem da dose de princípio ativo administrada por via oral que alcança a veia porta (%HIA). Esses valores estão distribuídos na faixa entre 1 e 100% de HIA.

3.1.2 Conjunto de Dados de Metabolismo de Primeira Passagem

O metabolismo de primeira passagem não pode ser considerado o único responsável pela diferença entre a porcentagem de composto absorvido na via gastrointestinal e a porcentagem de composto biodisponível, pois no decorrer desse caminho outros processos podem reduzir a biodisponibilidade sistêmica (*i.e.* excreção pré-sistêmica). Contudo, para efeitos de modelagem QSAR-ADME com o objetivo de estimar o metabolismo, assumiu-se que a diferença entre absorção e biodisponibilidade é atribuída ao metabolismo de primeira passagem mediado pelo CYP, como descrito na Figura 3.1¹⁹. Neste estudo, foi levada em consideração a quantidade de composto metabolizado, parâmetro essencial para a biodisponibilidade dos fármacos, não tendo sido considerada a participação individual das diferentes isoformas do CYP. O conjunto de dados consiste em 390 compostos, a maioria fármacos, associados aos

respectivos valores de metabolismo¹⁰⁶. Os valores coletados da literatura estão distribuídos entre 0 e 100% (quantidade de composto absorvido que é metabolizado).

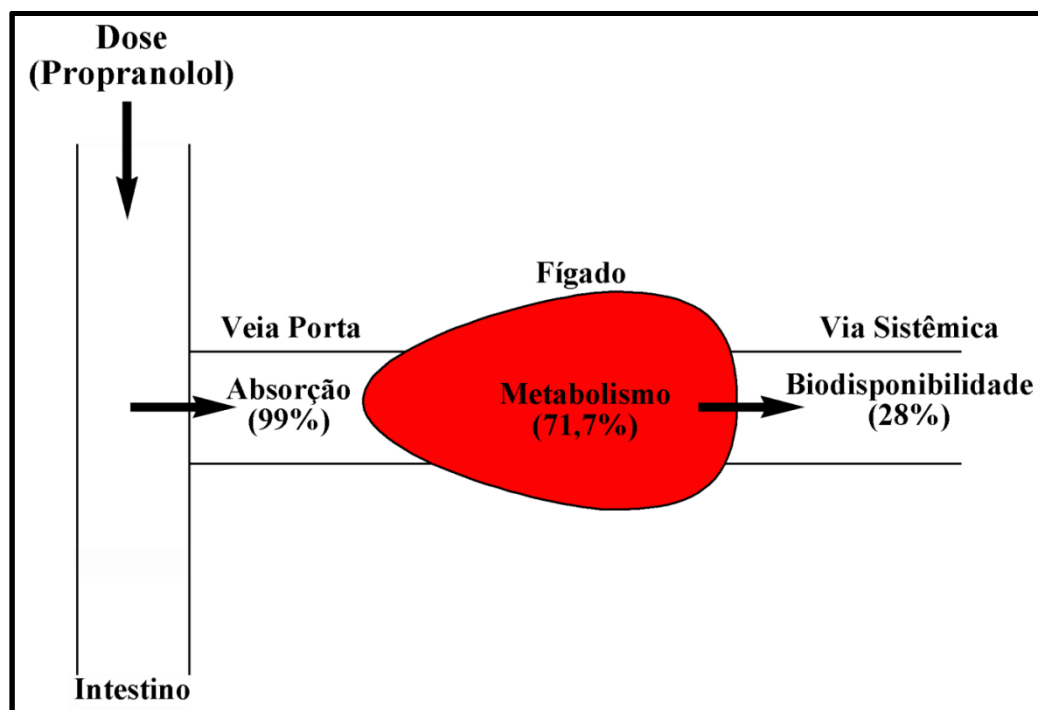


Figura 3.1 – Esquema de metabolismo de primeira passagem considerado neste estudo.

3.1.3 Conjunto de Dados de Permeabilidade da Barreira Hematoencefálica

Para a construção dos modelos *in silico*, foram coletados 255 compostos da literatura com os respectivos valores de permeabilidade da barreira hematoencefálica¹⁰⁷⁻¹¹⁷. Os valores da propriedade farmacocinética são expressos na forma de logaritmo da concentração dos compostos que se encontram no cérebro e o sangue simultaneamente ($\log[\text{cérebro}]/[\text{sangue}]$)¹¹³. Os valores de $\log BB$ para os compostos estudados nesta tese de doutorado variam entre -2 e 2, sendo que os valores negativos indicam concentrações maiores no sangue e valores positivos concentrações maiores no cérebro como ilustrado na Figura 3.2.

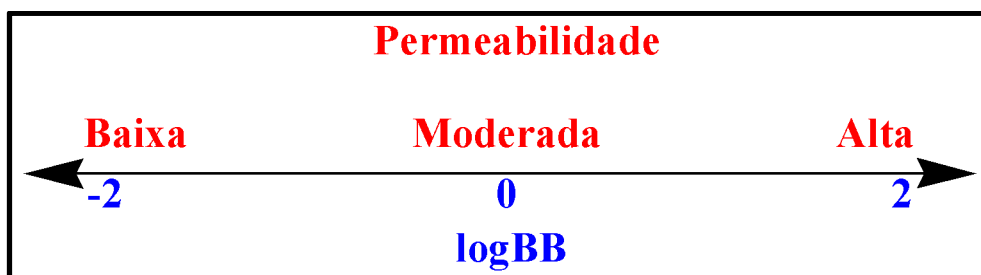


Figura 3.2 – Caracterização dos valores de permeabilidade da barreira hematoencefálica.

3.1.4 Conjunto de Dados de Inibição da Glicoproteína-P

O conjunto de dados consiste em 159 compostos, derivados de propafenona, associados aos respectivos valores de inibição da P-gp¹¹⁸. Os valores foram coletados da literatura a partir de ensaios celulares, os quais fornecem a concentração de composto necessária para induzir metade do efeito máximo de inibição (EC_{50})¹¹⁹. Os valores de EC_{50} do conjunto de dados variam entre 0,004 e 1585 μM e foram medidos sob as mesmas condições experimentais. Estes valores foram convertidos para os correspondentes valores de pEC_{50} ($-\log EC_{50}$) e usados como a variável dependente no processo de modelagem de QSAR.

3.1.5 Conjunto de Dados de Solubilidade Aquosa

O conjunto de dados é constituído de 1311 compostos associados aos seus respectivos valores de solubilidade aquosa e inclui moléculas compiladas das tradicionais bases de dados AQUASOL e PHYSPROP¹²⁰. Os dados são expressos na forma do logaritmo da solubilidade ($\log S$), onde S é a solubilidade em mol/L medida a temperaturas entre 20-25°C. Os valores de solubilidade estão distribuídos na faixa entre $2,4 \times 10^{-12}$ e 28 mol/L.

3.2 Modelagem de QSAR

Os trabalhos de modelagem, visualização, análises estatísticas e geração dos modelos de HQSAR foram realizados em estações computacionais no Laboratório de Química Medicinal e Computacional (LQMC) do Instituto de Física de São Carlos (IFSC) – Universidade de São Paulo (USP). A plataforma computacional SYBYL 8.0 (Tripos, St. Louis, EUA), operando no sistema operacional Linux, foi empregada em todos os casos.

As estruturas das moléculas dos conjuntos de dados e os valores correspondentes das propriedades farmacocinéticas e físico-química foram transferidos para uma planilha molecular (do inglês, *molecular spreadsheet* – MSS), onde toda a informação para o processo de modelagem de QSAR ficou armazenada de forma organizada. A próxima etapa consistiu na seleção de diversos parâmetros referentes à geração dos hologramas para as moléculas do conjunto treinamento. Estes parâmetros foram: comprimento do holograma, tamanho dos fragmentos e distinção de fragmentos. O holograma molecular é constituído por um conjunto de caixas (do inglês, *bins*) (Figura 3.3) contendo todos os fragmentos moleculares, os quais são designados por um algoritmo cíclico da verificação de redundância (do inglês, *Cyclic Redundancy Check* - CRC)¹²¹. O comprimento do holograma, que é determinado a partir de uma abordagem de triagem-erro com uma série padrão de hologramas de diferentes comprimentos (entre 53 e 401), define a dimensionalidade dos descritores. O processamento dos hologramas moleculares gera uma matriz de dimensão $K \times L$, onde K é o número de observações do conjunto treinamento e L é o comprimento do holograma molecular. Complementando a equação para a modelagem de QSAR, a matriz da variável dependente (propriedade farmacocinética) é adicionada à matriz $K \times L$, e torna-se o requisito necessário para que seja possível a associação estrutura e atividade (Figura 3.3)¹²².

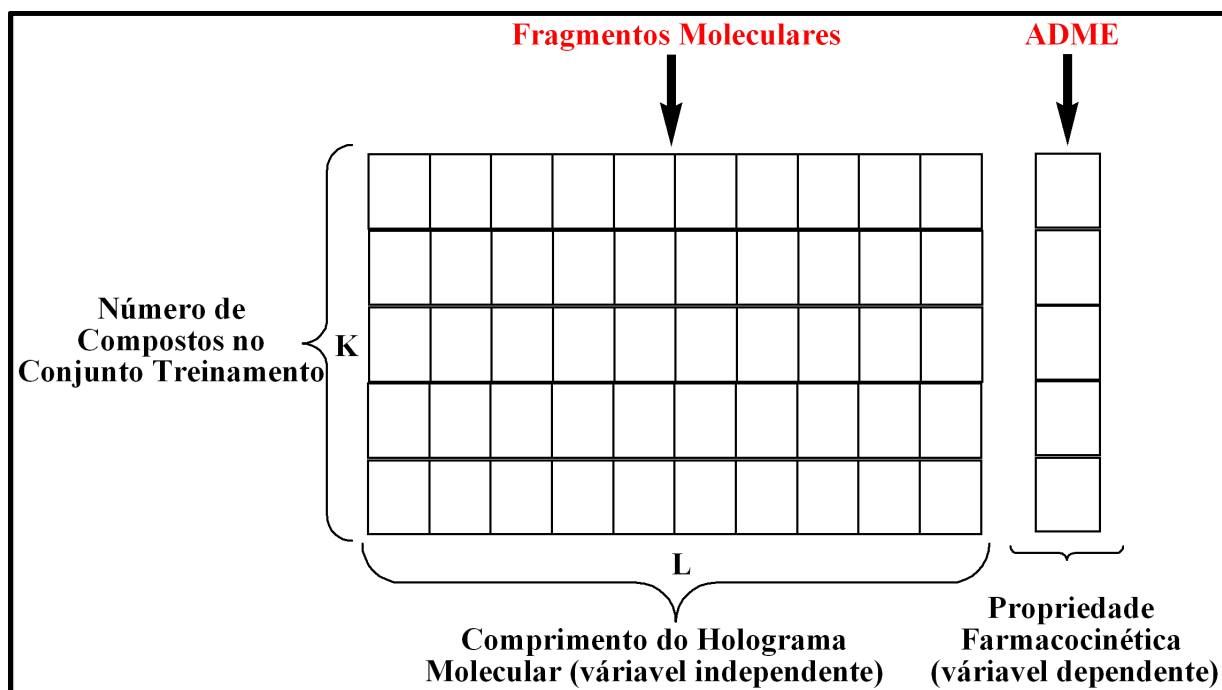


Figura 3.3 – Esquema ilustrativo da matriz de modelagem QSAR.

Várias análises PLS foram conduzidas com diversas combinações de distinção de fragmentos e tamanho dos fragmentos com o objetivo de produzir modelos de HQSAR com elevada consistência interna. Técnicas de validação cruzada determinaram o número de componentes principais (N) que produziram os modelos mais consistentes. Os modelos finais foram utilizados no processo de validação externa de conjuntos teste. Os resultados são apresentados no capítulo 4 a seguir.

CAPÍTULO 4

Resultados e Discussão

4 Resultados e Discussão

4.1 Absorção Intestinal Humana

4.1.1 Organização do Conjunto de Dados

Como descrito na sessão 3.1.1 do capítulo 3, 638 compostos foram organizados em um conjunto de dados para o processo de modelagem da absorção intestinal humana^{85,99-106}. A lista com a identificação das moléculas e os respectivos valores da propriedade farmacocinética (% HIA) é apresentada na Tabela 4.1.

Tabela 4.1 – Conjunto padrão de dados de absorção intestinal humana

Composto	HIA	Composto	HIA	Composto	HIA
AAFC (R)	32,00	Acitretina	95,00	Anagrelida	70,00
AAFC (S)	32,00	Acrivastina	88,00	Anastrozol	100,00
Acebutolol (R)	85,00	Albuterol (R)	83,00	Anfotericina B	3,00
Acebutolol (S)	85,00	Albuterol (S)	83,00	Antipirina	98,00
Acetaminofeno	85,00	Álcool Etilico	100,00	Arbekacin	1,00
Acetanilida	100,00	Alfacalcidol	100,00	Astemizol	100,00
Acetato de Ciproterona	98,00	Almitrina	90,00	Atropina (R)	94,00
Ac. ^[a] de Fludrocortisona	95,00	Almotriptano	75,00	Atropina (S)	94,00
Acetato de Megestrol	100,00	Alopurinol	85,00	Axetil Cefuroxima (S)	36,00
Acetazolamida	100,00	Alprazolam	90,00	Azatadina	90,00
Acetoexamida	80,00	Alprenolol (R)	93,00	Azelastina (R)	100,00
Ace. ^[b] de Triamcinolona	90,00	Alprenolol (S)	93,00	Azelastina (S)	100,00
Ácido Acetilsalicílico	100,00	Amantadina	90,00	Azimilida	100,00
Ácido Alendróico	1,00	Amfetamina (R)	90,00	Aztreonama	1,00
Ácido Edético	5,00	Amfetamina (S)	90,00	Baclofenol (R)	95,00
Ácido Etacrínico	90,00	Amigdalina (R)	5,00	Baclofenol (S)	95,00
Ácido Fusídico	100,00	Amigdalina (S)	5,00	Bendroflumetiazida (R)	100,00
Ácido Iotroxico	5,00	Amikacina	1,00	Benorilato	100,00
Ácido Meclofenâmico	100,00	Amilorida	50,00	Benserazida (R)	70,00
Ácido Mefenâmico	90,00	Aminoglutetimida (R)	100,00	Benserazida (S)	70,00
Ácido Nalidíxico	90,00	Aminoglutetimida (S)	100,00	Benzidamina	87,00
Ácido Nicotínico	94,00	Aminopirina	100,00	Benzobromarona	73,00
Ácido Pamidrónico	5,00	Amitriptilina	95,00	Bepiridil (S)	100,00
Ácido Risedrónico	1,00	Amosulalol (R)	100,00	Betaistina	100,00
Ácido Tiludronico	6,00	Amosulalol (S)	100,00	Betaxolol	90,00
Ácido Tranexâmico	55,00	Amoxapina	95,00	Bezafibrato	100,00
Ácido Valpróico	98,00	Amoxicilina	93,50	Bifemelano	95,00
Acipimox	95,00	Amrinona	93,00	Bisoprolol (R)	90,00

Continua

Continuação

Bisoprolol (S)	90,00	Clorexidina	5,00	Etambutol	78,00
Bromazepam	84,00	Clorfenesina (R)	100,00	Etidronato	5,00
Bromexina	70,00	Clorfenesina (S)	100,00	Etilmorfina	77,00
Budesonida (R)	100,00	Clorfeniramina	80,00	Etinilestradiol	100,00
Budesonida (S)	100,00	Cloroquina	100,00	Etionamida	90,00
Bumetanida	96,00	Clorpromazina	98,00	Etodolac (R)	70,00
Bupropiona (R)	87,00	Clortalidona (R)	65,00	Etodolac (S)	70,00
Bupropiona (S)	87,00	Clortalidona (S)	65,00	Etoposida	50,00
Buspirona	100,00	Clorzoxazona	90,00	Etoricoxibe	100,00
Cafeína	99,00	Clozapina	94,00	Famciclovir	70,00
Camazepam (R)	100,00	Codeína	93,00	Famotidina	41,00
Camazepam (S)	100,00	Corticosterona	100,00	Felbamato	90,00
Caproato de Hidroxi. ^[c]	90,00	Cotinina (R)	97,00	Felodipina (R)	94,00
Captopril	84,00	Cotinina (S)	97,00	Felodipina (S)	94,00
Carbamazepina	100,00	Coumarina	100,00	Fenazona	97,00
Carbimazol	95,00	Dantrolen	80,00	Fenazopiridina	90,00
Carfecilina (R)	99,00	Dapsona	90,00	Fenclofenaco	100,00
Carfecilina (S)	99,00	Daunorubicina	10,00	Fenfluramina (R)	95,00
Carmustina	100,00	Delmopinol (R)	95,00	Fenfluramina (S)	95,00
Carteolol (R)	95,00	Delmopinol (S)	95,00	Fenglutarimida (R)	100,00
Carteolol (S)	95,00	Desipramina	98,00	Fenglutarimida (S)	100,00
Carvedilol (R)	80,00	Desmetildiazepam	99,00	Fenilbutazona	98,00
Carvedilol (S)	80,00	Desogestrel	72,00	Fenindiona	100,00
Cefalexina	98,00	Dexametasona	90,00	Feniramina (R)	99,00
Cefatrizina	76,00	Dextromoramida	100,00	Feniramina (S)	99,00
Cefmetazol	10,00	Diacetato de Etinodiol	100,00	Fenitoína	90,00
Cefodizima	1,00	Diacetilmorfina	100,00	Fenobarbital	100,00
Cefotaxima	5,00	Diazepam	99,00	Fenoprofeno (R)	85,00
Cefpodoxima Pro. ^[d] (R)	5,00	Diazoxida	90,00	Fenoprofeno (S)	85,00
Ceftriaxona	1,00	Diciclomina	100,00	Fenprocoumona (R)	95,00
Cefuroxima	36,00	Diclofenaco	97,00	Fenprocoumona (S)	95,00
Cetanserina	100,00	Dienogeste	94,00	Feprazona	90,00
Cetoprofeno (S)	95,00	Dietilestilbestrol	90,00	Finasterida	100,00
Cetorolaco (R)	95,00	Dietilpropiona (R)	95,00	Flecainida (R)	81,00
Cetorolaco (S)	95,00	Dietilpropiona (S)	95,00	Flecainida (S)	81,00
Cetotifeno	90,00	Diflunisal	90,00	Floxacilina	80,00
Chloral_hydrate	100,00	Digoxina (R)	81,00	Flucitosina	100,00
Ciclopentiazida (R)	100,00	Digoxina (S)	81,00	Flucloxacilina	40,00
Ciclopentiazida (S)	100,00	Diidrocodeína	89,00	Fluconazol	94,00
Cicloserina	72,00	Diloxanida	90,00	Flumazenila	95,00
Cilazapril	59,00	Diltiazem	92,00	Flunarizina	80,00
Cilazaprilate	20,00	Dinitrato de Isossorbida	90,00	Flunisolido	80,00
Cilomilast	100,00	Dipiridamol	58,00	Flupentixol	100,00
Cimarin	47,00	Disopiramida (R)	95,00	Flurazepam	100,00
Ciproptadina	80,00	Disulfiram	97,00	Flurbiprofeno (R)	95,00
Ciprofibrato (R)	99,00	Dofetilida	100,00	Flurbiprofeno (S)	95,00
Ciprofibrato (S)	99,00	Dolasetrona	85,00	Flutamida	90,00
Ciprofloxacina	69,00	Domperidona	95,00	Fosmidomicina	30,00
Citalopram (R)	100,00	Dotiepina	95,00	Furosemida	61,00
Citalopram (S)	100,00	Doxazosina (R)	100,00	Galantamina	95,00
Clodronato	3,00	Doxazosina (S)	100,00	Galopamil (R)	97,00
Clofibrato	96,00	Doxepina	100,00	Galopamil (S)	97,00
Clometiazol	95,00	Doxorubicina	12,00	Gatifloxacina (R)	96,00
Clomipramina	95,00	Eflornitina (R)	55,00	Gatifloxacina (S)	96,00
Clonazepam	80,00	Eflornitina (S)	55,00	Gemfibrozil	100,00
Clopentixol	95,00	Enalapril	25,00	Gentamicina	1,00
Clopidogrel (R)	50,00	Enoximona	80,00	Gentamicina C2 (R)	1,00
Clopidogrel (S)	50,00	Eritromicina	35,00	Gentamicina C2 (S)	1,00
Clorambucil	100,00	Estavudina	100,00	Gentamicina C1	5,00
Clordiazepoxide	99,00	Estramustina	75,00	Gestodeno	100,00

Continua

Continuação

Gliburida	98,00	Levosimendan	90,00	Mifepristona	90,00
Glicina	100,00	Lidocaina	98,00	Mifobate (S)	82,00
Glimepirida	100,00	Lincomicina	28,00	Mifobato (R)	82,00
Glimidina	95,00	Linestrenol	100,00	Milrinona	83,00
Glipizida	100,00	Linezolida (R)	100,00	Minoxidila	96,00
Gliquidona	95,00	Linezolida (S)	100,00	Mirtazapina (S)	100,00
Granisetrona	100,00	Lomefloxacina (R)	99,00	Misoprostol (R)	88,00
Guanabenzol	78,00	Lomefloxacina (S)	99,00	Misoprostol (S)	88,00
Guanadrel (R)	80,00	Loracarbef	100,00	Mitoxantrona	5,00
Guanadrel (S)	80,00	Loratadina	90,00	Moclobemida	88,00
Guanfacina	100,00	Lorazepam (R)	95,00	Molécula-618	85,00
Guanoxan (R)	50,00	Lorazepam (S)	95,00	Morfina	90,00
Guanoxan (S)	50,00	Lormetazepam (R)	100,00	Moricizina	88,00
Haloperidol	100,00	Lormetazepam (S)	100,00	Moxalactam (R)	1,00
Hexobarbital (R)	95,00	Lornoxicam	100,00	Moxalactam (S)	1,00
Hexobarbital (S)	95,00	Losartan	80,00	Moxifloxacina	90,00
Hidroclorotiazida	68,00	Manitol	26,00	Moxisilita	70,00
Hidrocortisona	91,00	Maprotilina	98,00	Moxonidina	88,00
Hidroflumetiazida	67,00	Mebeverina (R)	90,00	Nabumetona	80,00
Hidroxicloroquina (R)	90,00	Mebeverina (S)	90,00	Naproxeno	99,00
Hidroxicloroquina (S)	90,00	Meperidina	100,00	Naratriptano	70,00
Ibuprofeno (R)	98,00	Meptazina (R)	100,00	Nateglinida (R)	83,00
Ibuprofeno (S)	98,00	Meptazinol (S)	100,00	Nateglinida (S)	83,00
Ibutilida (R)	82,00	Mercaptoetanol	77,00	Nefazodona	100,00
Ibutilida (S)	82,00	Mercaptopurina	50,00	Nefopam (S)	98,00
Imatinib	98,00	Mesalamina	80,00	Netivudina	28,00
Imipramina	98,00	Mesna	75,00	Nevirapina ⁴²	90,00
Indobufeno (R)	100,00	Mestranol	90,00	Nicardipina (R)	95,00
Indobufeno (S)	100,00	Metadona (R)	80,00	Nicardipina (S)	95,00
Indometacina	100,00	Metadona (S)	80,00	Nicorandil	100,00
Indoprofen (R)	100,00	Metaproterenol (R)	43,00	Nicotina (R)	100,00
Indoprofen (S)	100,00	Metaproterenol (S)	43,00	Nicotina (S)	100,00
Irbesartan	100,00	Metformina	54,00	Nifedipina	90,00
Isocarboxazida	70,00	Methsuximida (R)	100,00	Nilutamida	100,00
Isoniazida	80,00	Metildopa	50,00	Nilvadipina (R)	100,00
Isoproterenol (R)	85,00	Metilfenidato	80,00	Nilvadipina (S)	100,00
Isoproterenol (S)	85,00	Metilprednisolona	82,00	Nisoldipina (R)	90,00
Isotretinoína	90,00	Metimazol	95,00	Nisoldipina (S)	90,00
Isoxepac	98,00	Metirapona	80,00	Nitrazepam	95,00
Isoxicam	100,00	Metocarbamol (R)	100,00	Nitrendipina (R)	88,00
Isradipina (R)	92,00	Metocarbamol (S)	100,00	Nitrendipina (S)	88,00
Isradipina (S)	92,00	Metoclopramida	88,00	Nitrofurantoína	95,00
Itraconazol (R)	80,00	Metolazona (R)	64,00	Nizatidina	95,00
Itraconazol (S)	80,00	Metolazona (S)	64,00	Nordiazepam	99,00
Ivermectina	60,00	Metoprolol (R)	96,00	Noretindrona	100,00
Kanamicina	1,00	Metoprolol (S)	96,00	Norgestimato	100,00
Ketazolam	100,00	Metotrexato (R)	100,00	Norgestrel	100,00
Ketoconazol	75,00	Metotrexato (S)	100,00	Nortriptilina	100,00
Ketoprofeno (R)	95,00	Metoxianfetamina (R)	80,00	Ofloxacina (R)	100,00
Lacidipina	98,00	Metoxianfetamina (S)	80,00	Ofloxacina (S)	100,00
Lactulose	1,00	Metronidazol	95,00	Omeprazol (S)	88,50
Lamotrigina	98,00	Metsuximida (S)	100,00	Ondansetrona (R)	100,00
Lansoprazol (R)	85,00	Mexiletina (R)	98,00	Ondansetrona (S)	100,00
Lansoprazol (S)	85,00	Mexiletina (S)	98,00	Orfenadrina (R)	100,00
Letrozol	100,00	Mezlocillina	1,00	Orfenadrina (S)	100,00
Levamisol	95,00	Mianserina	70,00	Oseltamivir	75,00
Levetiracetam (R)	100,00	Mibefradil	69,00	Ouabaina	1,40
Levetiracetam (S)	100,00	Miconazol (R)	99,00	Oxamniquina (R)	80,00
Levobunolol	100,00	Miconazol (S)	99,00	Oxamniquina (S)	80,00
Levonorgestrel	100,00	Midazolam	100,00	Oxaprozina	98,00

Continua

Continuação

Oxatomida	100,00	Protriptilina	95,00	Tenoxicam	100,00
Oxazepam (R)	97,00	Proxifilina (R)	100,00	Teofilina	100,00
Oxazepam (S)	97,00	Proxifilina (S)	100,00	Terazosina (R)	91,00
Oxcarbazepina	100,00	Pseudoefedrina	100,00	Terazosina (S)	91,00
Oxibutinina (R)	100,00	Quetiapina	73,00	Terbinafina	80,00
Oxibutinina (S)	100,00	Quinalbarbitona (R)	90,00	Terbutalina (R)	63,00
Oxicodona	60,00	Quinalbarbitona (S)	90,00	Terbutalina (S)	63,00
Oxifedrina	85,00	Quinidina	81,00	Terfenadina (R)	100,00
Oxifenbutazona (R)	100,00	Rabeprazol (R)	90,00	Terfenadina (S)	100,00
Oxifenbutazona (S)	100,00	Rabeprazol (S)	90,00	Tesaglitazar	100,00
Oxprenolol (R)	95,00	Raloxifeno	60,00	Testosterona	100,00
Oxprenolol (S)	95,00	Ramipril	60,00	Tetraidrocanabinol	93,00
Pantoprazol (R)	77,00	Recainam	71,00	Tiabendazol	90,00
Pantoprazol (S)	77,00	Remoxiprida (R)	100,00	Tiacrilast	99,00
Papaverina	90,00	Remoxiprida (S)	100,00	Tiagabina	90,00
Paroxetina	100,00	Repaglinida (R)	100,00	Tianeptina (R)	99,00
Pefloxacina	95,00	Repaglinida (S)	100,00	Tianeptina (S)	99,00
Pelrinona	98,00	Reproterol (R)	60,00	Tibolona	90,00
Penbutolol (R)	100,00	Reproterol (S)	60,00	Ticlopidina	80,00
Penbutolol (S)	100,00	Riboflavin	80,00	Tilidina	100,00
Pentobarbital (R)	100,00	Riluzol	90,00	Timolol (R)	95,00
Pentobarbital (S)	100,00	Rimiterol	48,00	Timolol (S)	95,00
Pentoxifilina	95,00	Risperidona	97,00	Tinidazol	100,00
Perfenazina	100,00	Rivastigmina	98,00	Tobramicina	1,00
Pimozida	70,00	Rizatriptano	90,00	Tocainida (R)	100,00
Pindolol (R)	90,00	Ropinirol Acetil	98,00	Tocainida (S)	100,00
Pindolol (S)	90,00	Ropinirol	100,00	Tolazolina	90,00
Pirazinamida	100,00	Sacarina	91,00	Tolbutamida	90,00
Pirbuterol (R)	60,00	Scopolamina	93,00	Tolmesoxida (R)	98,00
Pirbuterol (S)	60,00	Selegiline (R)	100,00	Tolmesoxida (S)	98,00
Piretanida	85,00	Selegiline (S)	100,00	Tolmetina	99,00
Piroxicam	100,00	Sibutramina (R)	77,00	Tolrestat	66,00
Piroximona	81,00	Sibutramina (S)	77,00	Tolterodina	77,00
Pizotilin	80,00	Sotalol (R)	95,00	Topiramato	86,00
Politiazida (R)	100,00	Sotalol (S)	95,00	Toremifeno	100,00
Politiazida (S)	100,00	Sparfloxacina	90,00	Torseמידا	96,00
Practolol (R)	95,00	Streptomicina	1,00	Tramadol	95,00
Practolol (S)	95,00	Strofantosida (R)	16,00	Trandolapril	50,00
Pramipexol (R)	80,00	Strofantoside k (S)	16,00	Trapidil	96,00
Pramipexole (S)	80,00	Sulfadiazina	85,00	Trazodona	100,00
Praziquantel	96,00	Sulfametizol	85,00	Triazolam	85,00
Prazosina	100,00	Sulfametoxazol	100,00	Triexifenidil (R)	100,00
Primidona	75,00	Sulfasalazina	62,00	Triexifenidil (S)	100,00
Probenecida	100,00	Sulfinpirazona (R)	100,00	Trifluoperazina	100,00
Procainamida	85,00	Sulfinpirazona (S)	100,00	Trimeprazina (R)	80,00
Procarbazina	100,00	Sulfisomidina	90,00	Trimeprazina (S)	80,00
Prociclidina (R)	100,00	Sulfisoxazol	95,00	Trimetoprima	97,00
Prociclidina (S)	100,00	Sulindac (R)	90,00	Trimipramina (R)	80,00
Proguanil	80,00	Sulindac (S)	90,00	Trimipramina (S)	80,00
Prometazina (R)	95,00	Sultoprida (R)	89,00	Trofosfamida (R)	100,00
Prometazina (S)	95,00	Sultoprida (S)	89,00	Trofosfamida (S)	100,00
Propafenona (R)	95,00	Sumatriptano	67,00	Tropisetron	95,00
Propafenona (S)	95,00	Tamoxifeno	100,00	Trovafoxacina	88,00
Propiltiouracil	76,00	Tamsulosina	100,00	Urapidil	78,00
Propiverina	84,00	Telitromicina	90,00	Valsartan (R)	55,00
Propofol	100,00	Telmisartano	90,00	Valsartan (S)	55,00
Propoxifeno	95,00	Temazepam (R)	95,00	Vancomicina	1,00
Propranolol (R)	99,00	Temazepam (S)	95,00	Varfarina (R)	98,00
Propranolol (S)	99,00	Temozolomida	100,00	Varfarina (S)	98,00
Protionamida	90,00	Tenidap	90,00	Venlafaxina (R)	95,00

Continua

Continuação

Venlafaxina (S)	95,00	Zaleplom	100,00	Zonisamida	100,00
Viloxazina (R)	99,00	Zatebradin	79,00	Zopiclona (R)	98,00
Viloxazina (S)	99,00	Zileuton (R)	100,00	Zopiclona (S)	98,00
Vitamina A	80,00	Zileuton (S)	100,00	Zotepina	100,00
Zalcitabina	85,00	Zomepirac	100,00		

HIA (%). [a] Acetato. [b] Acetonida. [c] Hidroxiprogesterona. [d] Proxetil.

Durante o extensivo processo de coleta e padronização do conjunto de dados, alguns aspectos mereceram especial atenção. Por exemplo, não foram incluídos no conjunto de dados compostos com absorção intestinal dose dependente, ou seja, quando o valor da propriedade variava em função da dose administrada. Outro aspecto considerado foi o limite tolerável da variância do dado experimental. No conjunto de dados final foram incorporados somente compostos em que a diferença entre o menor e o maior valor descrito da propriedade não ultrapassou 20 unidades percentuais. Para ilustrar essa situação pode ser citado o fármaco anti-hiperuricêmico alopurinol (Zyloric[®]), usado no tratamento da gota. Esse fármaco possuía valores entre 60 e 80% (no limite de 20 unidades percentuais) e, portanto, foi incorporado ao conjunto de dados. No caso de compostos que apresentavam na literatura mais de um valor percentual de absorção intestinal, foi usada a média dos valores. Desta maneira, foi atribuído um valor de 70% para o alopurinol.

Os fármacos quirais com um centro estereogênico e sem configuração definida foram considerados separadamente e incluídos no conjunto de dados como enantiômeros (R) e (S) com o mesmo valor de HIA¹²³⁻¹²⁴. Todas as moléculas do conjunto de dados foram consideradas em sua forma neutra, como descrito na sessão 3.1 do capítulo 3.

A alta diversidade química das moléculas do conjunto de dados, associada à ampla distribuição de valores de absorção intestinal humana, garante notável cobertura do espaço químico-biológico da propriedade alvo. Entre os fármacos do conjunto de dados se encontram inibidores enzimáticos, agonistas e antagonistas de receptores biológicos, abrangendo diversas classes terapêuticas. Podem ser citados agentes sedativos, antibacterianos, analgésicos, antidepressivos, hipoglicemiantes, antieméticos, anticonvulsivantes, antifúngicos, anti-hipertensivos, antiinflamatórios, antivirais, anticâncer, entre outros. Alguns fármacos são mostrados na Figura 4.1 para ilustrar a representatividade do conjunto de dados.

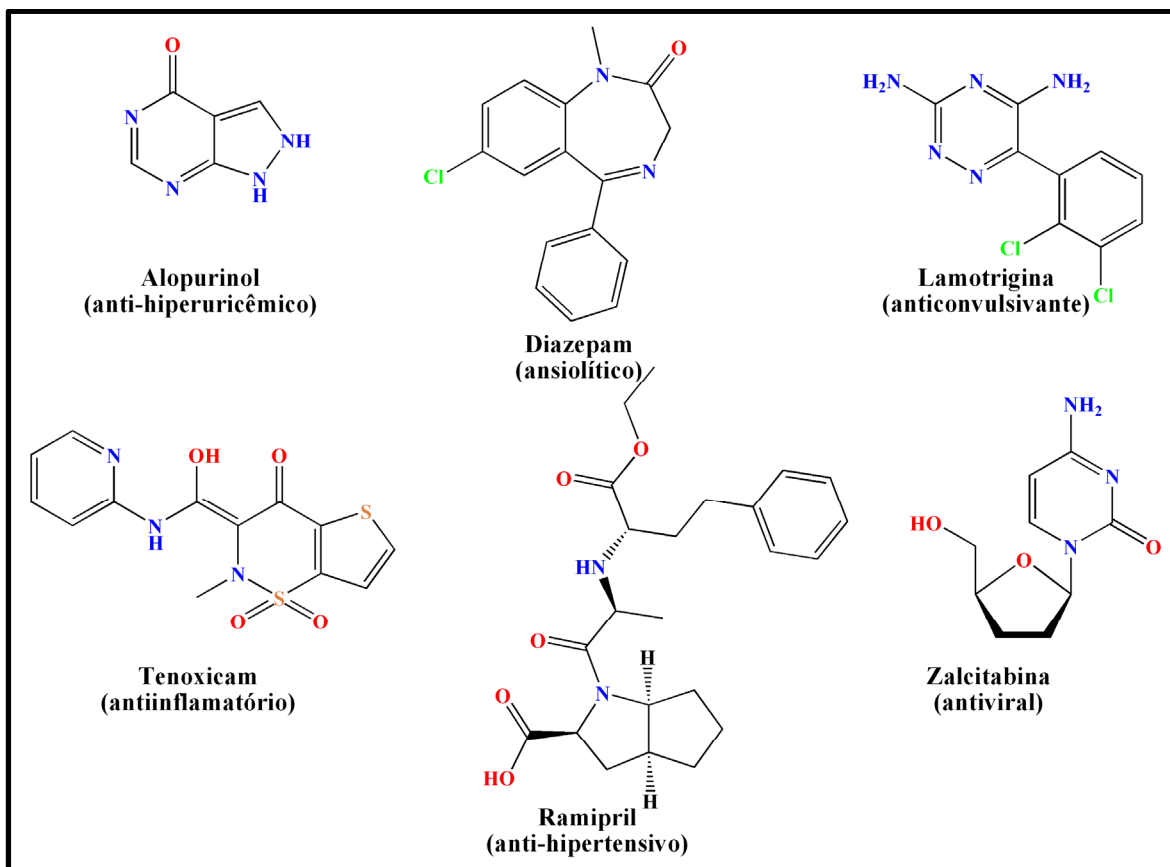


Figura 4.1 – Exemplos representativos do conjunto de dados da absorção intestinal humana.

Para racionalizar a análise do conjunto de dados, os compostos foram divididos em três faixas de distribuição de valores de HIA, como mostrado na Figura 4.2. Como pode ser observado, 7,84% dos 638 compostos do conjunto de dados encontram-se na faixa de distribuição de baixa HIA (com valores de até 40%). Uma fração de 37,77% dos compostos se encontra na faixa de distribuição de moderada HIA (valores entre 40 e 90%). Os 54,39% restantes encontram-se na faixa de alta HIA (valores maiores que 90%). Os dados corroboram uma tendência observada de que os fármacos possuem alta HIA, propriedade diretamente relacionada com a biodisponibilidade e efeito terapêutico.

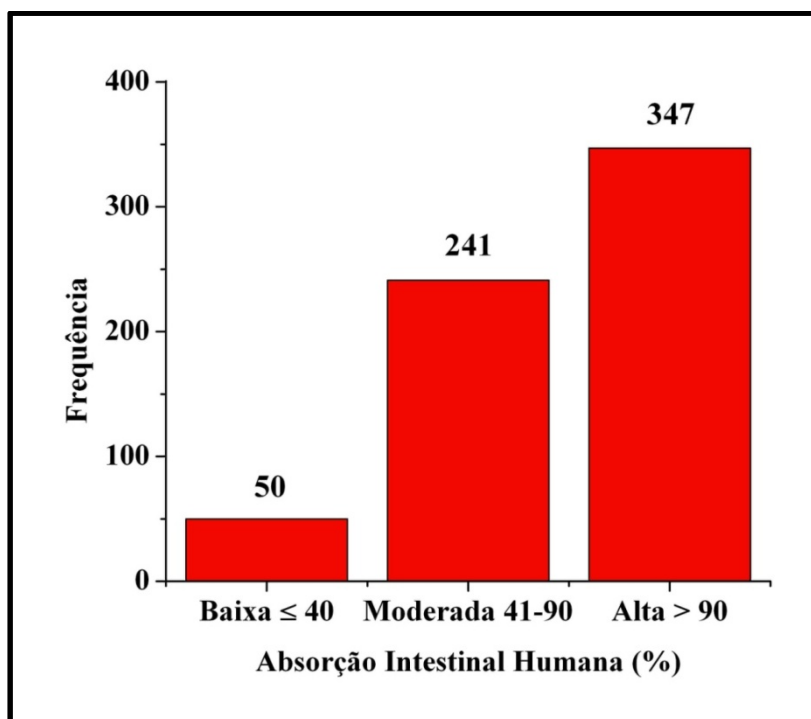


Figura 4.2 – Histograma da distribuição da absorção intestinal humana do conjunto de dados.

4.1.2 Desenvolvimento de Modelos de HQSAR-HIA

Devido a sua importância no desenvolvimento de candidatos a fármacos, existe grande interesse no desenvolvimento de modelos *in silico* para estimar e prever a absorção intestinal de NCEs. A proposta de nosso estudo é a geração de modelos preditivos a partir de um conjunto padrão de dados, empregando um método de QSAR baseado em fragmentos moleculares. As moléculas do conjunto de dados (Tabela 4.1) possuem cobertura significativa da faixa de distribuição de valores de HIA, entre 1 e 100%.

Para o trabalho de modelagem de HQSAR, os 638 compostos do conjunto de dados foram divididos em conjuntos treinamento e teste. O conjunto treinamento, empregado para o desenvolvimento dos modelos, é formado por 510 compostos, enquanto que o conjunto teste, usado para validação externa dos modelos, é formado por 128 compostos. A razão do número de compostos dos conjuntos treinamento e teste é 3,9:1, ou seja, cerca de 80:20 (%). Fundamental para o bom desempenho dos

modelos QSAR-ADME, a distribuição existente nos conjuntos treinamento e teste é representativa em relação ao conjunto padrão de dados original. O histograma da Figura 4.3 mostra a distribuição dos valores de HIA para os três conjuntos, salientando a boa distribuição existente.

O conjunto teste, em particular, apresenta uma distribuição representativa em relação ao conjunto treinamento. Essa ampla distribuição da propriedade farmacocinética representa um aspecto fundamental na investigação da consistência interna e externa de modelos de QSAR.

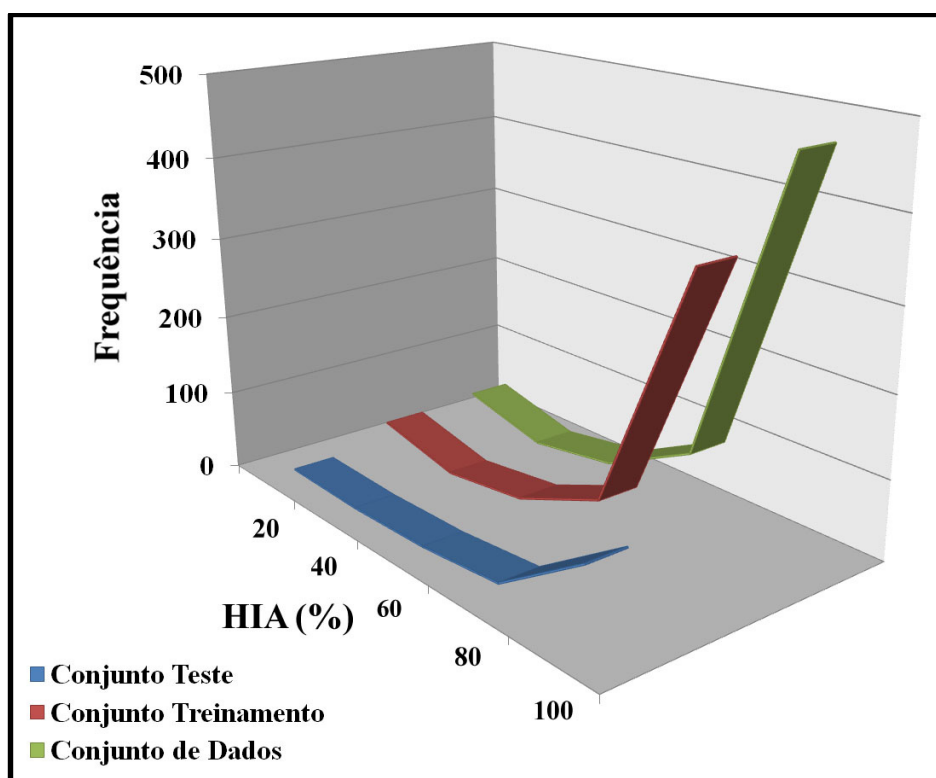


Figura 4.3 – Representação da distribuição dos valores de HIA.

Uma clara vantagem no processo de modelagem de propriedades farmacocinéticas é a determinação do dado experimental em valores percentuais. Os dados de propriedades biológicas ou bioquímicas, comumente usados na modelagem de QSAR, são representados em escala logarítmica (e.g., $pIC_{50} = -\log IC_{50}$) que lineariza os dados que podem variar em fatores superiores a 200.000 vezes¹¹⁸. No caso das propriedades de ADME, a maioria dos conjuntos de dados apresentam valores distribuídos em uma faixa definida entre 0 e 100%. Essa característica delinea uma faixa de distribuição do valor da propriedade para todos os compostos considerados no

processo de modelagem, sejam do conjunto treinamento (correlação) ou do conjunto teste (predição). Assim, o processo de extrapolação, que constitui uma das principais dificuldades e desafios da modelagem de QSAR de propriedades farmacodinâmicas, é minimizado e até mesmo eliminado.

Após a padronização do conjunto de dados, o passo seguinte foi o trabalho de modelagem de QSAR. A análise de HQSAR requer a seleção de alguns parâmetros essenciais para a composição dos hologramas moleculares: comprimento do holograma, tamanho dos fragmentos e distinção de fragmentos. Neste estudo, os hologramas foram gerados empregando o módulo de HQSAR da plataforma SYBYL 8.0.

O processo de obtenção dos fragmentos moleculares foi realizado utilizando as seguintes distinções de fragmentos: átomos (A), ligações (B), conectividade (C), átomos de hidrogênio (H), quiralidade (Ch) e doadores e aceptores (DA) (Tabela 1.3, sessão 1.7.1, capítulo 1). Foram empregadas oito combinações destes parâmetros para o tamanho padrão do fragmento (4-7): A/B/C/Ch, A/Ch/DA, A/B/C/H/Ch, A/C/Ch/DA, A/H/Ch/DA, A/B/C/Ch/DA, A/B/H/Ch/DA, A/B/C/H/Ch/DA. A análise de HQSAR foi realizada considerando os 12 comprimentos de hologramas disponíveis entre 53 e 401 (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401). A distinção quiralidade (Ch) foi sempre considerada na modelagem da absorção intestinal, pois vários pares de enantiômeros (R) e (S) estavam presentes no conjunto de dados. Se a distinção Ch não fosse considerada, os enantiômeros seriam analisados como moléculas idênticas afetando a qualidade estatística dos modelos. Os resultados obtidos a partir das análises PLS para os 510 compostos do conjunto treinamento, empregando várias combinações de distinção de fragmentos, são mostrados na Tabela 4.2.

Tabela 4.2 – Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão do fragmento (4-7)

Modelo	Distinção de Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B/C/Ch	0,51	0,79	11,79	8	353
2	A/Ch/DA	0,63	0,79	11,65	8	307
3	A/B/C/H/Ch	0,48	0,75	12,68	8	401
4	A/B/Ch/DA	0,62	0,85	9,99	8	401
5	A/C/Ch/DA	0,67	0,84	10,17	7	401
6	A/B/C/Ch/DA	0,57	0,75	12,79	5	353
7	A/B/H/Ch/DA	0,55	0,81	11,10	8	401
8	A/B/C/H/Ch/DA	0,57	0,82	10,74	8	307

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma. Distinção do fragmento: A, átomos; B, ligação; C, conexão; H, átomos de hidrogênio; Ch, quiralidade; DA, doador e receptor.

Os melhores resultados estatísticos foram obtidos usando as combinações A/Ch/DA do modelo 2 ($q^2 = 0,63$ e $r^2 = 0,79$), A/B/Ch/DA do modelo 4 ($q^2 = 0,62$ e $r^2 = 0,85$) e A/C/Ch/DA do modelo 5 ($q^2 = 0,67$ e $r^2 = 0,84$).

Os modelos 2, 4, 5, destacados em cinza na Tabela 4.2, foram selecionados para a investigação da influência da variação do tamanho dos fragmentos sobre os parâmetros estatísticos. O parâmetro referente ao tamanho dos fragmentos controla o comprimento mínimo e máximo dos fragmentos a serem incluídos no holograma. Estes parâmetros podem ser modificados de forma a incluir fragmentos moleculares maiores ou menores. Desta maneira, foram realizadas análises individuais empregando os seguintes tamanhos dos fragmentos: 2-5, 3-6, 4-7, 5-8 e 6-9, representados nestes intervalos pelo número mínimo e máximo de átomos. A Tabela 4.3 apresenta os resultados desta análise.

Tabela 4.3 – Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando as melhores distinções de fragmentos: A/Ch/DA, A/B/Ch/DA, A/C/Ch/DA

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
Modelo 2					
2-5	0,60	0,76	12,36	8	401
3-6	0,58	0,77	12,13	8	307
4-7	0,63	0,79	11,65	8	307
5-8	0,56	0,72	13,40	6	307
6-9	0,55	0,69	14,16	5	151
Modelo 4					
2-5	0,62	0,79	11,63	8	401
3-6	0,66	0,84	10,21	8	353
4-7	0,62	0,85	9,99	8	401
5-8	0,57	0,78	12,05	6	401
6-9	0,56	0,80	11,46	6	401
Modelo 5					
2-5	0,70	0,86	9,67	8	401
3-6	0,71	0,87	9,29	8	401
4-7	0,67	0,84	10,17	7	401
5-8	0,64	0,84	10,24	7	401
6-9	0,55	0,77	12,11	5	401

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma.

A variação do tamanho dos fragmentos levou a parâmetros estatísticos mais significativos para os modelos 4 e 5, quando comparados com o tamanho padrão do fragmento (4-7). Como pode ser observado na Tabela 4.3, os melhores resultados estatísticos (destacados em cinza) dentre todos os modelos testados foram obtidos com o tamanho do fragmento 3-6, para os modelos 4 e 5. Para o modelo 2 (distinção de fragmento A/Ch/DA), a variação do tamanho dos fragmentos não levou a melhores modelos. Estes resultados confirmam a importância da avaliação da influência do tamanho dos fragmentos na geração de hologramas moleculares. Além disso, os resultados ilustrados nas Tabelas 4.2 e 4.3 mostram que não há uma relação entre o

poder preditivo dos modelos de HQSAR e o comprimento do holograma, o que não surpreende, pois o padrão de ocupação no holograma molecular varia de uma maneira não trivial em relação às mudanças na composição do holograma.

4.1.3 Validação dos Modelos de HQSAR-HIA

Os modelos de HQSAR gerados com o conjunto treinamento possuem a validação cruzada, com o método “deixe um de fora” (do inglês, *Leave-One-Out* – LOO), como indicativo maior de sua robustez. Para melhor avaliar a consistência interna desses modelos, foi empregado o método de validação cruzada com “deixe muitos de fora” (do inglês, *Leave-Many-Out* – LMO). Neste processo, grupos de 15 moléculas foram aleatoriamente retirados e os resultados não indicaram uma influência significativa nos parâmetros estatísticos. Isso indica a boa capacidade de correlação dos dados no conjunto treinamento e a elevada consistência interna dos modelos. Contudo, apenas esta validação não pode ser considerada suficiente, pois bons coeficientes de correlação associados a uma elevada consistência interna muitas vezes não resultam em modelos com capacidade de prever a propriedade alvo para novos compostos.

Para que modelos de QSAR apresentem bom desempenho no planejamento de NCEs, é necessário que os mesmos passem por um processo de validação externa para uma avaliação da sua real capacidade preditiva e poder de predição. Como a estrutura molecular codificada nos hologramas moleculares, no caso de modelos HQSAR, está diretamente relacionada com a propriedade alvo das moléculas do conjunto treinamento, os modelos de HQSAR devem ser capazes de prever a absorção intestinal de novos candidatos baseados nessa informação. A validação externa dos melhores modelos de HQSAR, obtidos para o conjunto treinamento de 510 moléculas utilizando as distinções de fragmento A/Ch/DA, A/B/Ch/DA e A/C/Ch/DA, e o tamanho do fragmento 4-7, 3-6 e 3-6, respectivamente, foi feita com o emprego de um conjunto teste de 128 moléculas que foram completamente excluídas durante o

treinamento do modelo. Os resultados do processo de validação externa são apresentados na Tabela 4.4.

Tabela 4.4 – Valores experimentais e preditos de absorção intestinal humana, associados aos resíduos para os compostos do conjunto teste

No.	Composto	Exp ^[a]	Pred 1 ^[b]	Pred 2 ^[c]	Pred 3 ^[d]	Cons ^[e]	Res ^[f]
Conjunto Teste							
001	Ácido Acetilsalicílico	100,00	86,63	86,05	84,55	85,74	14,26
002	Ácido Alendrônico	1,00	-7,41	-9,20	-18,13	-11,58	12,58
003	Ácido Meclofenâmico	100,00	88,46	85,66	89,07	87,73	12,27
004	Ácido Nalidíxico	90,00	78,17	94,00	79,59	83,92	6,08
005	Ácido Pamidrônico	5,00	-6,32	-11,29	-16,96	-11,52	16,52
006	Ácido Valpróico	98,00	85,99	86,06	79,90	83,99	14,01
007	Acipimox	95,00	83,38	84,26	88,75	85,46	9,54
008	Álcool Etilico	100,00	88,99	89,94	87,27	88,73	11,27
009	Almotriptano	75,00	93,37	103,26	77,73	91,46	-16,46
010	Amigdalina (R)	5,00	19,50	15,13	18,65	17,76	-12,76
011	Antipirina	98,00	90,69	95,50	91,68	92,62	5,38
012	Arbekacin	1,00	5,82	-3,02	1,15	1,32	-0,32
013	Atropina (R)	94,00	88,77	103,99	97,22	96,66	-2,66
014	Benserazida (S)	70,00	61,68	64,88	63,79	63,45	6,55
015	Bezafibrato	100,00	95,94	94,14	83,70	91,26	8,74
016	Bifemelano	95,00	91,31	94,06	96,71	94,02	0,98
017	Bisoprolol (R)	90,00	91,44	86,19	91,56	89,73	0,27
018	Bromexina	70,00	96,53	90,93	95,38	94,28	-24,28
019	Bupropiona (R)	87,00	87,76	86,40	89,23	87,80	-0,80
020	Bupropiona (S)	87,00	85,64	88,08	88,42	87,38	-0,38
021	Camazepam (S)	100,00	99,74	87,80	101,66	96,40	3,60
022	Cefuroxima	36,00	55,44	48,09	55,26	52,93	-16,93
023	Clomipramina	95,00	96,86	93,73	105,53	98,71	-3,71
024	Clonazepam	80,00	101,63	93,89	88,99	94,84	-14,84
025	Clordiazepoxide	99,00	101,63	90,72	79,99	90,78	8,22
026	Clorfenesina (S)	100,00	89,31	91,89	84,57	88,59	11,41
027	Codeína	93,00	87,82	91,13	97,40	92,12	0,88
028	Desmetildiazepam	99,00	102,40	94,95	94,31	97,22	1,78
029	Diazepam	99,00	105,16	101,21	97,36	101,24	-2,24
030	Diciclomina	100,00	87,56	92,83	82,47	87,62	12,38
031	Dienogeste	94,00	96,91	93,33	84,28	91,51	2,49
032	Dietilpropiona (R)	95,00	90,30	89,19	87,02	88,84	6,16
033	Dietilpropiona (S)	95,00	86,43	92,16	87,40	88,66	6,34
034	Diflunisal	90,00	82,13	80,83	75,17	79,38	10,62
035	Diidrocodeína	89,00	85,70	92,71	93,47	90,63	-1,63
036	Diltiazem	92,00	82,69	84,11	95,46	87,42	4,58
037	Doxepina	100,00	92,34	89,31	94,35	92,00	8,00
038	Enoximona	80,00	90,47	99,07	80,02	89,86	-9,86
039	Etilmorfina	77,00	98,95	87,62	71,71	86,09	-9,09
040	Etionamida	90,00	84,19	89,12	78,23	83,85	6,15
041	Fenazona	97,00	90,69	95,50	91,68	92,62	4,38
042	Fenclofenaco	100,00	89,87	99,20	99,72	96,26	3,74
043	Fenindiona	100,00	75,28	100,67	93,78	89,91	10,09
044	Feprazona	90,00	92,88	91,62	87,14	90,54	-0,54
045	Finasterida	100,00	111,00	116,83	105,51	111,11	-11,11
046	Flucitosina	100,00	88,71	86,72	95,01	90,15	9,85
047	Flumazenila	95,00	86,62	94,48	95,15	92,08	2,92
048	Gatifloxacina (S)	96,00	91,80	94,59	88,55	91,64	4,36
049	Gestodeno	100,00	105,25	96,44	84,08	95,26	4,74
050	Glimepirida	100,00	79,41	98,78	85,54	87,91	12,09

Continua

Continuação

051	Glimidina	95,00	84,71	86,99	88,65	86,78	8,22
052	Glipizida	100,00	83,29	93,73	96,84	91,29	8,71
053	Gliquidona	95,00	79,86	101,51	94,49	91,95	3,05
054	Haloperidol	100,00	91,40	100,69	103,32	98,47	1,53
055	Indoprofen (R)	100,00	90,58	97,45	103,89	97,31	2,69
056	Irbesartan	100,00	92,70	77,13	90,08	86,63	13,37
057	Isocarboxazida	70,00	89,46	97,75	80,87	89,36	-19,36
058	Isoproterenol (R)	85,00	76,48	69,02	64,64	70,05	14,95
059	Isoproterenol (S)	85,00	79,40	73,05	68,20	73,55	11,45
060	Kanamicina	1,00	-12,01	-2,13	2,45	-3,89	4,89
061	Ketoprofeno (R)	95,00	82,34	85,96	88,42	85,57	9,43
062	Levonorgestrel	100,00	103,08	102,22	107,44	104,25	-4,25
063	Lidocaína	98,00	96,96	87,16	86,29	90,14	7,86
064	Loracarbef	100,00	66,31	57,91	75,88	66,70	33,30
065	Maprotilina	98,00	93,91	84,46	101,63	93,33	4,67
066	Mebeverina (S)	90,00	85,79	93,28	99,70	92,92	-2,92
067	Mesalamina	80,00	72,58	68,66	70,78	70,67	9,33
068	Metadona (R)	80,00	87,25	94,36	93,56	91,72	-11,72
069	Metaproterenol (R)	43,00	73,90	67,29	63,68	68,29	-25,29
070	Metildopa	50,00	78,47	71,84	82,60	77,64	-27,64
071	Metolazona (R)	64,00	95,20	73,12	72,71	80,34	-16,34
072	Mexiletina (R)	98,00	93,39	91,40	94,06	92,95	5,05
073	Miconazol (S)	99,00	93,18	85,43	88,04	88,88	10,12
074	Mifobate (S)	82,00	66,52	110,73	72,28	83,18	-1,18
075	Milrinona	83,00	98,19	85,60	79,59	87,79	-4,79
076	Molécula-618	85,00	70,76	82,86	92,13	81,91	3,09
077	Nabumetona	80,00	89,18	87,85	93,80	90,27	-10,27
078	Nateglinida (R)	83,00	83,83	88,64	94,47	88,98	-5,98
079	Netivudina	28,00	51,23	55,03	56,42	54,23	-26,23
080	Nevirapina	90,00	85,06	119,87	68,26	91,06	-1,06
081	Nicotina (R)	100,00	92,87	94,72	82,58	90,06	9,95
082	Nicotina (S)	100,00	90,50	94,96	89,31	91,59	8,41
083	Nitrofurantoína	95,00	87,43	86,74	85,32	86,50	8,50
084	Nordiazepam	99,00	102,40	94,95	78,09	91,81	7,19
085	Ofloxacina (S)	100,00	98,66	111,34	100,24	103,41	-3,41
086	Omeprazol (S)	88,50	76,71	90,65	79,65	82,33	6,17
087	Oxcarbazepina	100,00	98,29	82,43	100,52	93,75	6,25
088	Oxibutinina (R)	100,00	88,30	88,62	92,66	89,86	10,14
089	Oxifenbutazona (R)	100,00	87,16	82,87	79,28	83,10	16,90
090	Pantoprazol (S)	77,00	81,74	75,19	89,90	82,28	-5,28
091	Papaverina	90,00	86,42	84,62	95,11	88,71	1,29
092	Pentobarbital (R)	100,00	98,08	94,25	90,68	94,33	5,67
093	Pentobarbital (S)	100,00	95,93	94,25	90,28	93,49	6,51
094	Pentoxifilina	95,00	94,57	99,90	102,57	99,01	-4,01
095	Pramipexole (S)	80,00	72,04	79,18	70,75	73,99	6,01
096	Propiltiouracil	76,00	92,90	92,67	88,50	91,36	-15,36
097	Propoxifeno	95,00	99,33	94,45	91,01	94,93	0,07
098	Propranolol (R)	99,00	93,70	88,75	93,80	92,09	6,91
099	Quinidina	81,00	68,25	85,68	88,10	80,68	0,32
100	Rabeprazol (S)	90,00	86,33	93,86	80,79	86,99	3,01
101	Ramipril	60,00	71,95	54,92	52,27	59,71	0,29
102	Repaglinida (R)	100,00	92,61	94,46	110,93	99,33	0,67
103	Reproterol (S)	60,00	76,27	71,93	82,08	76,76	-16,76
104	Rimiterol	48,00	74,00	72,46	71,63	72,69	-24,69
105	Ropinirol Acetil	98,00	91,41	87,82	92,08	90,44	7,56
106	Ropinirol	100,00	94,16	99,42	106,16	99,92	0,08
107	Selegiline (S)	100,00	89,82	95,12	90,77	91,90	8,10
108	Sotalol (S)	95,00	83,94	90,26	81,58	85,26	9,74
109	Strofantoside k (S)	16,00	9,15	5,59	4,94	6,56	9,44
110	Sulfametizol	85,00	92,30	75,39	87,54	85,08	-0,08

Continua

Continuação

111	Sulindac (R)	90,00	93,06	101,71	104,77	99,85	-9,85
112	Sumatriptano	67,00	91,13	97,82	79,02	89,32	-22,32
113	Tamoxifeno	100,00	95,86	102,13	97,43	98,47	1,53
114	Tenidap	90,00	83,39	83,01	80,45	82,29	7,71
115	Ticlopidina	80,00	83,86	91,14	76,41	83,81	-3,81
116	Tilidina	100,00	93,62	91,70	93,70	93,00	7,00
117	Tinidazol	100,00	89,25	83,89	86,51	86,55	13,45
118	Tocainida (R)	100,00	93,15	87,82	92,45	91,14	8,86
119	Tolazolina	90,00	90,53	89,27	84,24	88,01	1,99
120	Tolbutamida	90,00	82,14	90,06	85,20	85,80	4,20
121	Trapidil	96,00	96,14	96,57	96,56	96,42	-0,42
122	Trofosfamida (R)	100,00	94,38	110,82	101,85	102,35	-2,35
123	Urapidil	78,00	101,13	82,53	78,86	87,51	-9,51
124	Valsartan (S)	55,00	80,53	64,90	57,22	67,55	-12,55
125	Venlafaxina (R)	95,00	91,17	93,54	96,38	93,70	1,30
126	Vitamina A	80,00	97,84	83,79	94,87	92,17	-12,17
127	Zaleplom	100,00	103,91	105,37	89,84	99,71	0,29
128	Zonisamida	100,00	95,23	98,94	81,52	91,90	8,10
$r^2_{\text{pred}}[g]$			0,72	0,75	0,76	0,80	

[a] Valor experimental de HIA (%). [b] Valor (%) predito pelo modelo 2. [c] Valor (%) predito pelo modelo 4. [d] Valor (%) predito pelo modelo 5. [e] Predição consenso. [f] Resíduo: diferença entre o valor experimental e a predição consenso. [g] coeficiente de correlação entre o valor experimental e predito do conjunto teste.

O resultado da validação externa mostra que o conjunto teste, o qual representa diferentes características estruturais quando comparado com o conjunto treinamento, é razoavelmente bem predito em todos os modelos selecionados. Entretanto, ao utilizar o método de predição consenso $[(\text{Pred } 1 + \text{Pred } 2 + \text{Pred } 3)/3]$, o resultado final mostrou-se mais robusto e de maior confiabilidade. Quando comparado com os modelos individuais 2, 4 e 5 o método de predição consenso apresenta melhor coeficiente de correlação para o conjunto teste (r^2_{pred}) (Tabela 4.4), podendo assim ser aplicado para a predição de novos compostos. Os valores preditos se aproximam dos experimentais, desviando-se em não mais do que 20 unidades percentuais. As únicas exceções são os compostos 18, 64, 69, 70, 79, 104, 112, que representam 5,47% do conjunto teste, para os quais os valores apresentam uma diferença mais significativa (entre 22,32 e 33,30 unidades percentuais). A Figura 4.4 apresenta o gráfico dos valores experimentais contra os valores preditos de absorção intestinal humana para os conjuntos treinamento e teste.

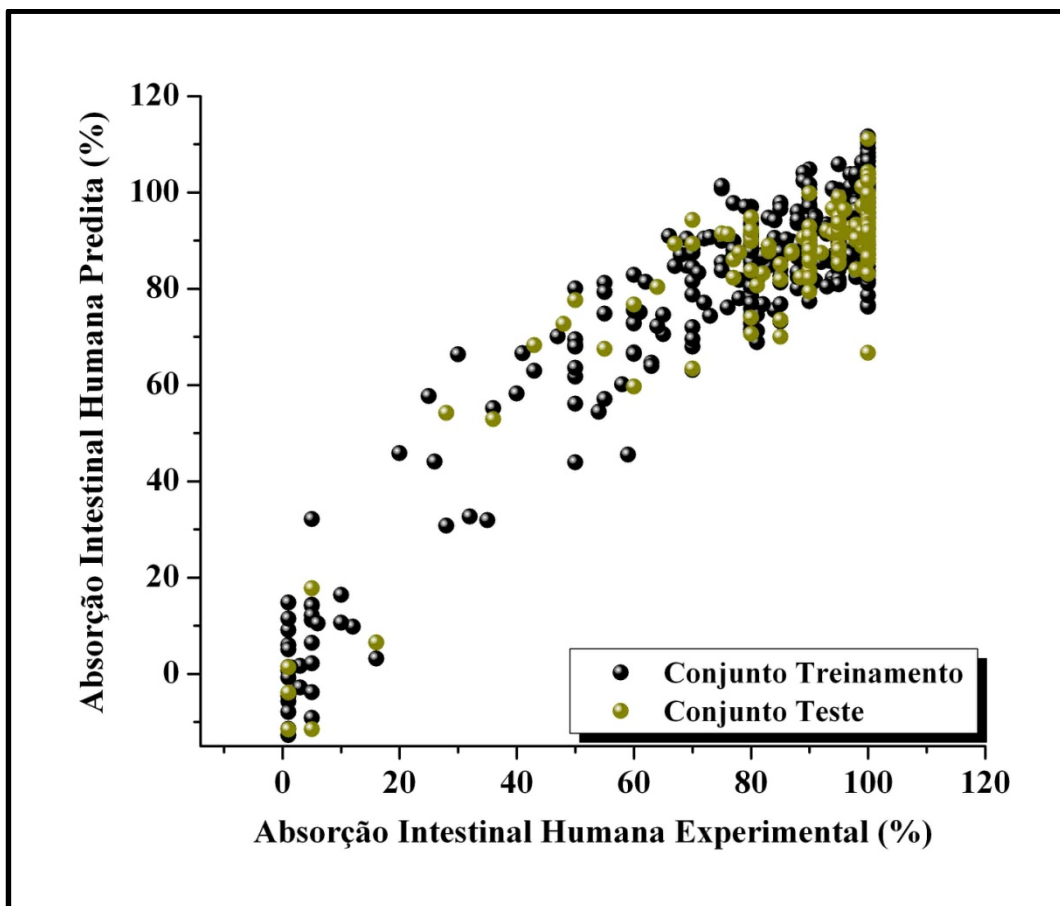


Figura 4.4 – Gráfico de HQSAR dos valores experimentais e preditos de absorção intestinal para os 638 compostos do conjunto de dados.

A análise gráfica confirma a boa concordância dos valores preditos e experimentais. O modelo final gerado através do método de predição consenso apresenta, portanto, considerável consistência interna e externa. O modelo possui capacidade de classificação nas faixas de baixa, moderada e alta HIA, podendo ser aplicado na predição de novos compostos.

4.2 Metabolismo de Primeira Passagem

4.2.1 Organização do Conjunto de Dados

Foram coletados 390 compostos (a maioria fármacos) da literatura juntamente com os respectivos valores de absorção intestinal e biodisponibilidade¹⁰⁶. Os valores foram convertidos no valor de metabolismo de primeira passagem para cada um dos compostos, como descrito na seção 3.1.2 do capítulo 3. A Tabela 4.5 apresenta o conjunto de dados usado para a modelagem do metabolismo de primeira passagem.

Tabela 4.5 – Conjunto padrão de dados do metabolismo de primeira passagem

Composto	MET ^[a]	Composto	MET ^[a]	Composto	MET ^[a]
Acamprosato	0,00	Ampicilina	0,00	Carnitina (S)	0,00
Acarbosa	50,00	Amrinona	0,00	Carteolol (R)	7,37
Acetaminofeno	0,00	Astemizol	95,00	Carteolol (S)	7,37
Acetazolamida	1,00	Atropina (R)	46,81	Carvedilol (R)	68,75
Ace. ^[b] de Triamcinolona	74,44	Atropina (S)	46,81	Carvedilol (S)	68,75
Aciclovir	13,00	Azitromicina	2,70	Cefalexina	6,25
Ácido Acetilsalicílico	14,63	Azosemida	0,00	Cefixima	21,66
Ácido Alendrónico	0,00	Baclofeno	26,31	Cefpodoxima Prox. ^[c] (R)	100,00
Ácido Clavulânico	0,00	Baclofeno (S)	26,31	Cefpodoxima Prox. ^[c] (S)	100,00
Ácido Fólico	0,00	Benazepril	0,00	Cefradina	1,05
Ácido Risedrónico	0,00	Bendroflumetiazida (R)	10,00	Ceftizoxima	100,00
Ácido Salicílico	1,00	Bendroflumetiazida (S)	10,00	Ceftriaxona	100,00
Ácido Tiludrónico	0,00	Bepridil (R)	39,00	Celiprolol (R)	20,00
Ácido Valpróico	1,02	Bepridil (S)	39,00	Celiprolol (S)	20,00
Acipimox	5,26	Betaxolol (R)	5,55	Cicaprost	0,00
Acitretina	37,90	Betaxolol (S)	5,55	Ciclopentiazida (R)	0,00
Albuterol (R)	38,55	Bezafibrato	1,00	Ciclopentiazida (S)	0,00
Albuterol (S)	38,55	Biotina	1,00	Ciclosporina A	9,09
Almitrina	30,00	Biperidena	70,00	Cimetidina	31,00
Alprazolam	0,00	Bisoprolol (R)	0,00	Ciprofibrato (R)	0,00
Amantadina	0,00	Bretilium	0,00	Ciprofibrato (S)	0,00
Amfotericina B	0,00	Bromocriptina	83,87	Cisaprida	59,18
Amilorida	0,00	Bumetanida	6,25	Citalopram (R)	20,00
Amitriptilina	49,47	Bupropiona	19,54	Citalopram (S)	20,00
Amobarbital	0,00	Burpropiona (R)	19,54	Clindamicina	13,80
Amosulalol (R)	0,00	Cafeina	1,00	Clodronato	66,66
Amosulalol (S)	0,00	Captopril	6,76	Clomipramina	46,31
Amoxapina	62,10	Carbamazepina	30,00	Clorambucil	18,00
Amoxicilina	2,17	Carnitina (R)	0,00	Cloramfenicol	7,86

Continua

Continuação

Clordiazepoxida	0,00	Fenoprofeno (S)	5,88	Linezolida (S)	0,00
Cloridrato de Trospium	0,00	Fenoterol	96,66	Lisinopril	3,70
Cloroquina (R)	13,00	Finasterida	30,00	Lomefloxacina (R)	2,02
Cloroquina (S)	13,00	Flubiprofeno (S)	3,16	Lomefloxacina (S)	2,02
Clorotiazida	0,00	Flucitosina	16,00	Lorazepam (R)	2,10
Clorpromazina	74,50	Flunisolida	75,00	Lorazepam (S)	2,10
Clortalidona (R)	0,00	Fluoxetina (R)	25,00	Lormetazepam (R)	25,00
Clortalidona (R)	1,54	Fluoxetina (S)	25,00	Lormetazepam (S)	25,00
Clortalidona (S)	1,54	Flupentixol	60,00	Losartam	58,75
Clozapina	41,49	Flurbiprofeno (R)	3,16	Lovastatina	50,00
Cotinina (R)	0,00	Fluvoxamina	41,11	Meperidina	48,00
Cotinina (S)	0,00	Foscarnet	27,77	Mesalamina	6,25
Cromolim	0,00	Fosinopril	2,94	Mesna	33,33
Desipramina	61,22	Gabapentina	3,50	Metadona (R)	0,00
Desmetildiazepam	0,00	Galopamil (R)	80,41	Metadona (S)	0,00
Dexametasona	13,33	Galopamil (S)	80,41	Metaproterenol (R)	39,50
Dextromoramida (R)	1,00	Gatifloxacina (R)	0,00	Metaproterenol (S)	39,50
Dextromoramida (S)	1,00	Gatifloxacina (S)	0,00	Metformina	1,85
Dextrose (R)	0,00	Gemfibrozil	2,00	Metildopa (S)	39,02
Dextrose (S)	0,00	Genaconazol	0,00	Metilergonovina	40,00
Diacetilmorfina	99,00	Gliburida	21,43	Metimazole	2,10
Diazepam	4,00	Glicopirrolato	5,55	Metoclopramida	43,18
Diazoxida	0,00	Glimepirida	0,00	Metronidazol	0,00
Diclofenaco	44,33	Glipizida	5,00	Mexiletina (R)	11,22
Dienogest	0,00	Granisetron	40,00	Mexiletina (S)	11,22
Diflunisal	0,00	Guanfacina	19,00	Miconazol (R)	74,78
Digitoxina	0,00	Haloperidol	40,00	Miconazol (S)	74,78
Digitoxina (R)	10,25	Hexobarbital (R)	0,00	Mifepristona	23,33
Digoxina (S)	10,25	Hexobarbital (S)	0,00	Minoxidil	1,04
Dihidroergotamina	97,14	Hidroflumetiazida	25,37	Mirtazapina	50,00
Diltiazem	51,65	Hidroxicloroquina (R)	17,77	Misoprostol (R)	9,09
Dipiridamol	13,79	Hidroxicloroquina (S)	17,77	Misoprostol (S)	9,09
Disopiramida (R)	12,63	Ibuprofeno (R)	13,26	Moclobemida	36,36
Disopiramida (S)	12,63	Ibuprofeno (S)	13,26	Molsidomina	52,00
Distigmina	0,00	Imipramina	54,08	Mononitrato de Isso. ^[d] 2	1,00
Dofetilida	2,00	Indobufeno (R)	15,00	Morfina	73,33
Domperidona	84,21	Indobufeno (S)	15,00	Moricizina	56,81
Doxazosina (R)	37,00	Indometacina	1,00	Moxonidina	0,00
Doxazosina (R)	37,00	Irbesartano	30,00	Nabumetona	56,25
Eflornitina (R)	0,00	Isoniazida	0,00	Nadolol (R)	6,25
Eflornitina (S)	0,00	Isotretinoína	63,33	Nadolol (S)	6,25
Enalapril	19,05	Isradipina	79,34	Nalbufina	84,00
Enoximone	32,50	Itraconazol (R)	31,25	Naloxona	97,80
Ergotamina	98,00	Itraconazol (S)	31,25	Naltrexona	76,84
Eritromicina	0,00	Kanamicina	0,00	Nateglinida (R)	12,05
Etambutol	1,28	Ketoprofeno (R)	10,53	Nateglinida (S)	12,05
Etidronato	0,00	Ketoprofeno (S)	10,53	Nedocromil	0,00
Etinilestradiol	49,00	Ketorolac	0,00	Neomicina	0,00
Etoricoxibe	0,00	Labetalol	73,68	Neostigmina	0,00
Felbamato	0,00	Lacidipina	81,63	Nicardipina	75,80
Felodipine (R)	84,04	Lactulose	0,00	Nicorandil	10,00
Felodipine (S)	84,04	Lamivudina	1,16	Nifedipina	46,66
Fenfluramina (R)	12,63	Letrozol	0,00	Nisoldipina (R)	94,44
Fenfluramina (S)	12,63	Levamisol (R)	31,56	Nisoldipina (S)	94,44
Fenilbutazona	8,16	Levamisol (S)	31,56	Nitrazepam	17,90
Fenilpropanolamina	0,00	Levetiracetam (R)	0,00	Nitrendipina (R)	84,09
Fenindione	3,00	Levetiracetam (S)	0,00	Nitrendipina (S)	84,09
Feniramina (R)	0,00	Levodopa (R)	60,46	Nitrofurantoína	4,21
Feniramina (S)	0,00	Levodopa (S)	60,46	Nivadipina (R)	86,00
Fenoprofeno (R)	5,88	Lincomicina	10,71	Nivadipina (S)	86,00

Continua

Continuação

Noretindrone	36,00	Rabeprazol	42,22	Tetrabenazina (R)	95,00
Nortriptilina	45,00	Ramipril	21,66	Tetrabenazina (S)	95,00
Ofloxacina (R)	0,00	Ranitidina	8,77	Tianeptina (R)	0,00
Ofloxacina (S)	0,00	Repaglinida (R)	40,00	Tianeptina (S)	0,00
Omeprazol	41,25	Repaglinida (S)	40,00	Timolol (R)	34,73
Ondansetrona (S)	40,00	Riluzol	31,11	Timolol (S)	34,73
Orfenadrina (R)	10,00	Risperidona	33,00	Tinidazol	2,00
Orfenadrina (S)	10,00	Ritodrina	63,16	Tocainide (R)	5,00
Oxaprozina	0,00	Rivastigmina (R)	59,18	Tocainide (S)	5,00
Oxazepam (R)	0,00	Rivastigmina (S)	59,18	Tolbutamida	1,11
Oxazepam (S)	0,00	Rizatriptam	47,77	Tolmesoxide	13,26
Oxcarbazepina	5,00	Sacarina	7,69	Tolmetina	4,04
Papaverina	68,88	Saquinavir	86,66	Torse mida	16,66
Paroxetina	50,00	Scopolamina	70,96	Tramadol	26,31
Pefloxacina	5,26	Sotalol (R)	2,10	Trapidil	0,00
Penbutolol (R)	7,00	Sotalol (S)	2,10	Trazodona	25,00
Penbutolol (S)	7,00	Spironolactona	65,75	Triazolam	41,17
Pentazocina	81,05	Stavudina	18,00	Triexifenidil (R)	0,00
Pentobarbital (R)	1,00	Streptomicina	0,00	Triexifenidil (S)	0,00
Pentobarbital (S)	1,00	Sulfadiazina	3,22	Trimipramina (R)	50,00
Perfenazina	65,00	Sulfametazina	0,00	Trimipramina (S)	50,00
Perindopril	30,52	Sulfametoxazol	0,00	Trofesfamida	0,00
Pimozida	42,86	Sulfipirazona	0,00	Valsartam (R)	58,18
Piroxicam	1,00	Sulindaco	2,22	Valsartam (S)	58,18
Piroximone	0,00	Sulpirida	32,50	Varfarina (R)	4,12
Prazosina	32,63	Sumatriptano	79,10	Varfarina (S)	4,12
Prednisolona	18,18	Tacrina	71,58	Verapamil (R)	77,89
Probenecida	0,00	Tamsulosina (R)	0,00	Verapamil (S)	77,89
Prociclidina (R)	25,00	Tamsulosina (S)	0,00	Viloxazina (R)	29,29
Prociclidina (S)	25,00	Telitromicina	63,33	Viloxazine (S)	29,29
Progesterona	73,96	Telmisartam	52,22	Xamoterol (R)	0,00
Prometazina (R)	73,68	Temazepam (R)	4,21	Xamoterol (S)	0,00
Prometazina (S)	73,68	Temazepam (S)	4,21	Zanamivir	0,00
Propranolol (R)	71,70	Tenoxicam	0,00	Ziprasidona	0,00
Propranolol (S)	71,70	Teofilina	2,04	Zolpidem	26,31
Protriptilina	10,52	Terazosina (R)	5,49	Zonisamida	0,00
Proxifilina (R)	3,00	Terazosina (S)	5,49	Zopicolne (R)	18,36
Proxifilina (S)	3,00	Terbutalina (R)	77,77	Zopicolne (S)	18,36
Pseudoefedrina	0,00	Terbutallina (S)	77,77		
Quinapril	26,66	Tesaglitazar	0,00		

[a] Metabolismo (%). [b] Acetonida. [c] Proxetil. [d] Isossorbida.

Como demonstrado anteriormente para o conjunto de dados de HIA, os fármacos quirais, com um centro estereogênico e sem configuração definida, foram considerados separadamente e incluídos no conjunto de dados como enantiômeros (R) e (S), assumindo-se o mesmo valor da propriedade para as duas formas. Os fármacos com mais de um centro estereogênico sem configuração definida foram excluídos do conjunto. Todas as moléculas foram consideradas em sua forma neutra. A diversidade química das moléculas e as várias classes terapêuticas contempladas garantem ampla cobertura do espaço químico-biológico em relação ao metabolismo de primeira passagem.

A Figura 4.5 apresenta a distribuição dos valores, onde 74,36% dos compostos estão na faixa de baixo metabolismo, 22,05% encontram-se com metabolismo moderado e 3,59% com alto metabolismo de primeira passagem. De acordo com esses dados, observa-se a predominância de um baixo metabolismo entre os fármacos.

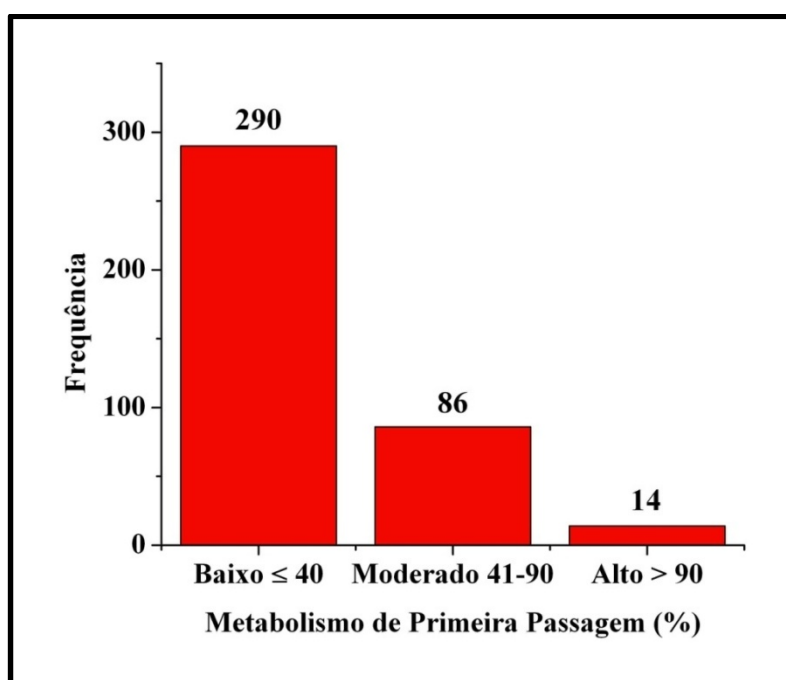


Figura 4.5 – Histograma da distribuição do metabolismo de primeira passagem do conjunto de dados.

4.2.2 Desenvolvimento de Modelos de HQSAR-MET

A importância do metabolismo de primeira passagem justifica a geração de modelos *in silico* baseados em uma aproximação dessa propriedade. Os modelos foram gerados na plataforma computacional Sybyl 8.0. Inicialmente os 390 compostos do conjunto de dados foram divididos em conjuntos treinamento e teste, respeitando-se a proporção desejável de 80/20. O conjunto treinamento é formado por 313 compostos, e o conjunto teste por 77. A relação entre os conjuntos e a representatividade do conjunto teste pode ser observada na Figura 4.6.

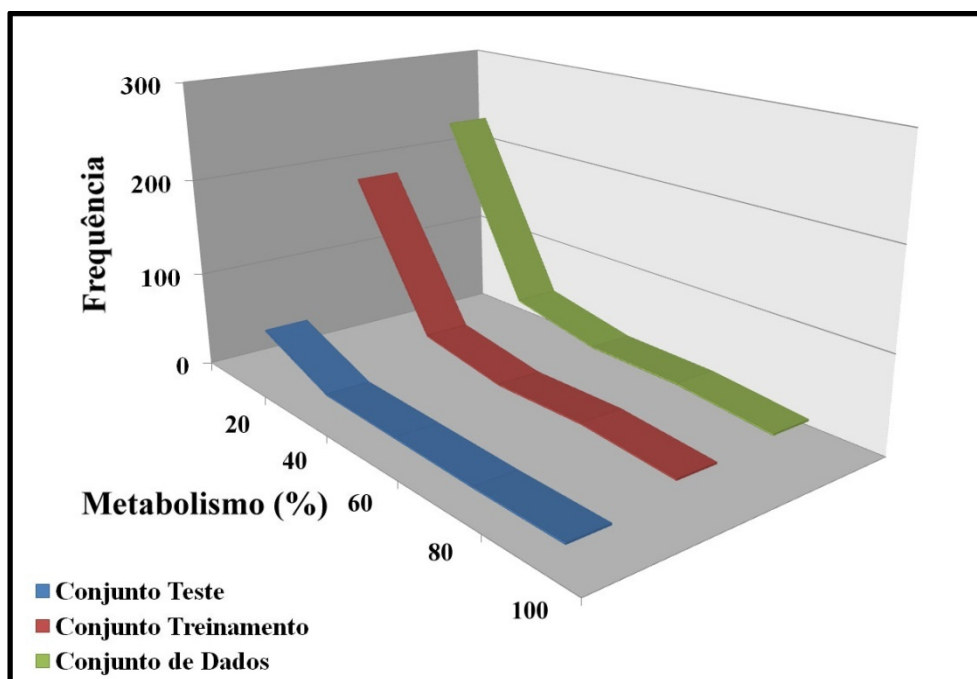


Figura 4.6 – Representação da distribuição dos valores de metabolismo de primeira passagem.

Estabelecida a divisão do conjunto de dados foi dado início a modelagem QSAR usando o método de hologramas moleculares. A aplicação de várias distinções de fragmentos, mantendo-se constante o parâmetro de número de átomos (4-7), forneceu modelos HQSAR de boa qualidade. Os resultados estatísticos significantes obtidos a partir da análise PLS são apresentados na Tabela 4.6.

Tabela 4.6 – Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão do fragmento (4-7)

Modelo	Distinção do Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B/C/Ch	0,57	0,83	12,25	7	401
2	A/Ch/DA	0,49	0,73	15,66	8	307
3	A/B/C/H/Ch	0,49	0,71	16,18	7	257
4	A/B/Ch/DA	0,58	0,83	12,25	8	401
5	A/C/Ch/DA	0,60	0,82	12,76	7	353
6	A/B/C/Ch/DA	0,58	0,84	12,03	7	401
7	A/B/H/Ch/DA	0,59	0,83	12,32	8	401
8	A/B/C/H/Ch/DA	0,54	0,82	12,80	8	307

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma. Distinção do fragmento: A, átomos; B, ligação; C, conexão; H, átomos de hidrogênio; Ch, quiralidade; DA, doador e receptor.

Os melhores resultados estatísticos foram obtidos usando as combinações A/B/C/Ch do modelo **1** ($q^2 = 0,57$ e $r^2 = 0,83$), A/B/Ch/DA do modelo **4** ($q^2 = 0,58$ e $r^2 = 0,83$), A/C/Ch/DA do modelo **5** ($q^2 = 0,60$, e $r^2 = 0,82$) e A/B/C/Ch/DA do modelo **6** ($q^2 = 0,58$, e $r^2 = 0,84$) (destacadas em cinza na Tabela 4.6). Esses modelos foram selecionados para a investigação da influência da variação do tamanho dos fragmentos (2-5, 3-6, 5-8 e 6-9) sobre os parâmetros estatísticos. A Tabela 4.7 apresenta os resultados obtidos nesta análise.

Tabela 4.7 – Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando as melhores distinções observadas: A/B/C/Ch, A/B/Ch/DA, A/C/Ch/DA e A/B/C/Ch/DA

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
Modelo 1					
2-5	0,61	0,79	13,85	8	207
3-6	0,61	0,82	12,80	8	307
4-7	0,57	0,83	12,25	7	401
5-8	0,54	0,85	11,41	8	307
6-9	0,49	0,70	16,41	5	151
Modelo 4					
2-5	0,55	0,75	14,91	8	401
3-6	0,60	0,80	13,28	8	401
4-7	0,58	0,83	12,25	8	401
5-8	0,53	0,79	13,56	7	401
6-9	0,53	0,76	14,60	6	401
Modelo 5					
2-5	0,60	0,81	13,17	8	353
3-6	0,60	0,78	13,98	7	257
4-7	0,60	0,82	12,76	7	353
5-8	0,57	0,79	13,49	6	401
6-9	0,47	0,73	15,54	5	401
Modelo 6					
2-5	0,68	0,84	11,76	8	307
3-6	0,63	0,82	12,72	7	307
4-7	0,58	0,84	12,03	7	401
5-8	0,55	0,84	12,09	7	401
6-9	0,51	0,78	14,02	6	401

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma.

Pode ser observada a melhora dos parâmetros estatísticos para os modelos **1**, **4** e **6**, com tamanhos dos fragmentos 3-6, 3-6 e 2-5, respectivamente (destacados em cinza na Tabela 4.7). Apenas o modelo **5** não apresentou melhora em função da variação do tamanho do fragmento.

4.2.3 Validação dos Modelos de HQSAR-MET

De maneira similar aos modelos de HIA, os modelos finais HQSAR-MET possuem validação cruzada do tipo LOO e foram validados pelo método LMO com até 12 grupos. Apenas esses métodos não são suficientes para avaliar o real poder preditivo dos modelos, que pode ser investigado através da predição do metabolismo de primeira passagem para moléculas do conjunto teste (validação externa), que não fizeram parte da construção dos modelos.

Um conjunto teste de 77 moléculas foi submetido ao processo de validação externa com os modelos **1**, **4**, **5** e **6**, destacados em cinza na Tabela 4.7. Os resultados são apresentados na Tabela 4.8.

Tabela 4.8 – Valores experimentais e preditos de metabolismo de primeira passagem associados aos resíduos para os compostos do conjunto teste

No.	Composto	Exp ^[a]	Pred 1 ^[b]	Pred 2 ^[c]	Pred 3 ^[d]	Pred 4 ^[e]	Cons ^[f]	Res ^[g]
Conjunto Teste								
001	Acetaminofeno	0,00	-0,08	18,64	8,74	6,60	8,48	-6,60
002	Acetazolamida	1,00	14,72	14,74	13,34	10,47	13,32	-9,47
003	Aciclovir	13,00	19,57	-4,12	27,61	12,94	14,00	0,06
004	Ácido Clavulânico	0,00	21,51	7,62	5,59	3,48	9,55	-3,48
005	Ácido Fólico	0,00	19,10	8,51	-13,43	0,61	3,70	-0,61
006	Amilorida	0,00	5,25	10,36	23,26	-1,70	9,29	1,70
007	Ampicilina	0,00	26,24	-5,59	-6,51	-5,19	2,24	5,19
008	Baclofeno (S)	26,31	31,49	16,60	15,61	16,02	19,93	10,29
009	Bepiridil (R)	39,00	23,53	21,92	43,22	23,04	27,93	15,96
010	Bumetanida	6,25	13,51	16,52	9,39	5,17	11,14	1,08
011	Burpropiona (R)	19,54	18,80	27,56	16,89	14,63	19,47	4,91
012	Captopril	6,76	13,26	10,44	9,53	8,07	10,32	-1,31
013	Carbamazepina	30,00	10,08	46,56	33,49	30,03	30,04	-0,02
014	Carnitina (R)	0,00	9,61	3,40	7,54	2,53	5,77	-2,53
015	Cefixima	21,66	70,40	40,10	24,32	29,65	41,12	-7,99
016	Ceftriaxona	100,00	70,34	84,93	94,41	70,84	80,13	29,16
017	Ciclopentiazida (R)	0,00	0,58	-4,38	10,88	-1,12	1,49	1,12
018	Ciclosporina A	9,09	-48,99	-81,53	-56,14	29,82	-39,21	-20,73
019	Cisaprida	59,18	57,83	59,19	15,51	46,12	44,66	13,06
020	Citalopram (S)	20,00	28,19	36,63	24,00	24,23	28,26	-4,23
021	Clindamicina	13,80	6,92	10,39	2,14	7,55	6,75	6,25
022	Clodronato	66,66	13,40	-7,67	1,70	38,88	11,58	27,79
023	Clomipramina	46,31	67,05	60,93	57,15	54,29	59,86	-7,98
024	Clorpromazina	74,50	75,89	76,96	20,63	57,05	57,63	17,45
025	Clortalidona (R)	1,54	26,33	21,09	20,89	11,88	20,05	-10,34
026	Dextrose (R)	0,00	-1,23	8,87	-2,37	6,09	2,84	-6,09
027	Dextrose (S)	0,00	-0,71	6,69	-0,55	4,44	2,47	-4,44
028	Digoxina (S)	10,25	-2,37	-3,29	1,01	-6,61	-2,81	16,86
029	Dipiridamol	13,79	8,49	31,63	77,23	17,96	33,83	-4,17
030	Doxazosina (R)	37,00	9,85	15,96	18,01	17,19	15,25	19,81

Continua

Continuação

031	Etidronato	0,00	-4,04	2,09	-13,96	7,63	-2,07	-7,63
032	Etinilestradiol	49,00	14,95	35,19	43,38	31,86	31,34	17,14
033	Feniramina (R)	0,00	5,50	5,87	14,43	4,36	7,54	-4,36
034	Fenoterol	96,66	42,65	59,18	57,02	74,89	58,43	21,77
035	Finasterida	30,00	24,98	45,43	43,30	30,36	36,02	-0,36
036	Flucitosina	16,00	9,59	6,98	3,82	24,02	11,10	-8,02
037	Foscarnet	27,77	-2,99	14,69	-3,28	17,73	6,54	10,04
038	Genaconazol	0,00	18,92	8,80	21,86	11,10	15,17	-11,10
039	Glipizida	5,00	-6,11	5,32	13,78	-13,22	-0,06	18,22
040	Haloperidol	40,00	41,64	52,58	28,12	33,74	39,02	6,26
041	Isradipina	79,34	72,07	74,51	74,06	90,74	77,85	-11,40
042	Kanamicina	0,00	2,10	-1,14	-32,98	9,46	-5,64	-9,46
043	Lacidipina	81,63	80,13	89,62	74,01	97,14	85,22	-15,51
044	Lamivudina	1,16	3,92	11,05	15,65	-9,93	5,17	11,09
045	Levetiracetam (S)	0,00	1,06	0,38	7,01	4,14	3,15	-4,14
046	Levodopa (R)	60,46	33,05	37,18	28,79	42,36	35,34	18,10
047	Lisinopril	3,70	9,83	13,91	16,19	15,62	13,89	-11,92
048	Lorazepam (S)	2,10	19,10	17,97	24,55	20,44	20,51	-18,34
049	Metildopa (S)	39,02	33,80	34,41	29,31	40,52	34,51	-1,50
050	Metoclopramida	43,18	27,50	39,37	33,43	35,65	33,99	7,53
051	Molsidomina	52,00	23,67	27,38	13,70	40,19	26,23	11,81
052	Moricizina	56,81	56,70	56,24	47,56	63,38	55,97	-6,57
053	Nalbufina	84,00	46,82	93,38	83,83	60,16	71,05	23,84
054	Nateglinida (R)	12,05	11,20	13,95	22,97	13,33	15,36	-1,28
055	Neostigmina	0,00	13,32	32,01	14,08	7,23	16,66	-7,23
056	Nicorandil	10,00	7,21	15,35	-1,29	5,38	6,66	4,62
057	Nizatidina	0,00	22,89	11,81	12,05	18,55	16,33	-18,55
058	Ofloxacina (S)	0,00	-0,37	-5,65	8,08	-2,68	-0,15	2,68
059	Orfenadrina (S)	10,00	19,81	24,19	12,14	14,85	17,75	-4,85
060	Pentobarbital (R)	1,00	9,00	-9,67	4,21	1,01	1,14	-0,01
061	Prazosina	32,63	30,62	9,57	28,16	29,97	24,58	2,66
062	Prometazina (S)	73,68	54,91	56,25	62,27	66,19	59,91	7,49
063	Protriptilina	10,52	0,33	18,60	13,28	1,72	8,48	8,80
064	Sacarina	7,69	-8,66	1,33	-14,37	12,62	-2,27	-4,93
065	Saquinavir	86,66	40,43	41,42	33,66	49,92	41,36	36,74
066	Stavudina	18,00	25,62	-1,61	6,95	15,12	11,52	2,88
067	Sulfametoxazol	0,00	-8,41	4,76	4,81	-8,06	-1,72	8,06
068	Telitromicina	63,33	58,21	25,97	53,09	80,71	54,49	-17,38
069	Tianeptina (S)	0,00	-0,91	9,88	-6,80	-0,18	0,50	0,18
070	Timolol (R)	34,73	25,02	4,18	41,24	24,43	23,72	10,30
071	Tolmetina	4,04	13,49	6,20	13,77	12,64	11,52	-8,60
072	Torseמידa	16,66	14,38	40,51	18,66	9,71	20,81	6,95
073	Trapidil	0,00	5,45	4,21	22,51	-18,60	3,39	18,60
074	Trazodona	25,00	26,51	24,24	25,39	17,87	23,50	7,13
075	Triazolam	41,17	30,95	35,86	43,94	40,76	37,88	0,41
076	Trimipramina (S)	50,00	65,57	65,56	51,27	67,57	62,49	-17,57
077	Verapamil (R)	77,89	71,57	74,72	75,08	76,48	74,46	1,41
	$r^2_{pred} [h]$		0,57	0,56	0,51	0,82	0,72	

[a] Valor experimental de metabolismo (%). [b] Valor predito pelo modelo 1. [c] Valor predito pelo modelo 4. [d] Valor predito pelo modelo 5. [e] Valor predito pelo modelo 6. [f] Predição consenso. [g] Resíduo: diferença entre o valor experimental e a predição gerada pelo modelo 6. [h] coeficiente de correlação entre o valor experimental e predito do conjunto teste.

O resultado da validação externa revela que o conjunto teste é bem predito pelo modelo 6 (Pred 4, Tabela 4.8), cujo valor dos resíduos consideravelmente baixo, desviando em não mais que 20 unidades percentuais. As exceções incluem os

compostos **18**, **22**, **34**, **53** e **65** (6,49% do conjunto), para os quais os valores apresentam uma diferença mais significativa (entre 20,73 e 36,74 unidades percentuais).

As predições realizadas pelos modelos **1**, **4**, **5** e também pelo método de predição consenso ($[\text{Pred } 1 + \text{Pred } 2 + \text{Pred } 3 + \text{Pred } 4]/4$) ficaram abaixo dos índices aconselháveis para aplicação desses modelos na predição de novos compostos. A predição realizada pelos modelos individuais apresenta valores de resíduo significativos para alguns compostos, e valor de coeficiente de correlação (r^2_{pred}) não superior a 0,57. O mesmo ocorre para o método de predição consenso, com $r^2_{\text{pred}} = 0,72$, um valor 0,1 abaixo do melhor coeficiente de correlação gerado pelo modelo **6** (0,82). A Figura 4.7 apresenta o gráfico, gerado a partir do modelo **6**, dos valores experimentais contra os valores preditos de metabolismo de primeira passagem para os conjuntos treinamento e teste.

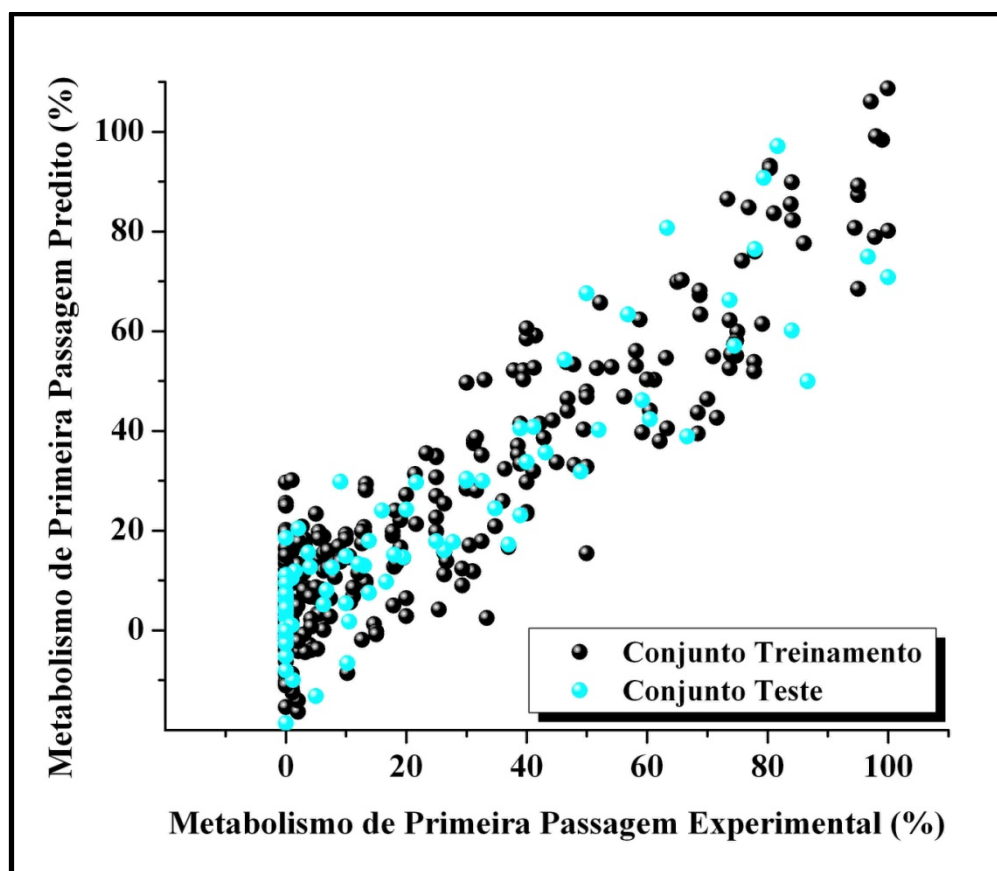


Figura 4.7 – Gráfico de HQSAR dos valores experimentais e preditos de metabolismo para os 390 compostos do conjunto de dados.

A boa concordância entre os valores preditos e experimentais (Figura 4.7) confirma a utilidade do modelo **6** na predição de novos compostos, bem como a sua consistência interna e externa. O modelo gerado pode ser utilizado para prever faixas de metabolismo auxiliando na escolha de NCEs qualificadas.

4.3 Permeabilidade da Barreira Hematoencefálica

4.3.1 Organização do Conjunto de Dados

Como descrito na sessão 3.1.3 do capítulo 3, foram coletados 255 compostos para modelagem da permeabilidade da barreira hematoencefálica¹⁰⁷⁻¹¹⁷. A Tabela 4.9 apresenta a lista com a identificação das moléculas e os correspondentes valores de logBB.

Tabela 4.9 – Conjunto padrão de dados de permeabilidade da barreira hematoencefálica

Composto	logBB	Composto	logBB	Composto	logBB
1,1 1-tricloroetano	0,40	1,1 1-trifluoro-2-cloroetano	0,08	1-hidroimidazolam	-0,07
1-propanol	-0,16	2 2-dimetilbutano	1,04	2-metilpentano	0,97
2-metilpropanol	-0,17	2-propanol	-0,15	3-metilhexano (R)	0,90
3-metilhexano (S)	0,90	3-metilpentano	1,01	4-hidroimidazolam (R)	-0,30
4-hidroimidazolam (S)	-0,30	9-hidroxisperidona (R)	-0,67	9-hidroxisperidona (S)	-0,67
Ácido Acetilsalicílico	-0,50	Ácido Salicílico	-0,44	Ácido Valpróico	-0,22
Alprazolam	0,04	Aminopirina	0,00	Amitripatalina	0,89
Amobarbital	0,04	Antipirina	-0,10	Atenolol (R)	-1,42
Atenolol (S)	-1,42	BBcpd10	-1,17	BCNU	-0,52
Benzeno	0,37	Bis Hidroxilato L663581	-1,82	Bretazenil	-0,09
Butanona	-0,08	Cafeína	-0,06	Carbamazepina	-0,14
Ciclohexano	0,92	Ciclopropano	0,00	Cimetidina	-1,42
Cimetidina - Análogo 12	-0,67	Cimetidina - Análogo 13	-0,66	Clobazam	0,35
Clonidina	0,11	Clorpromazina	1,06	Codeína	0,55
Comp-3	-2,00	Comp-4	-1,30	Comp-5	-1,06
Comp-6	0,11	Comp-10	-0,82	Comp-11 1	-1,17
Comp-14	-0,66	Comp-15	-0,12	Comp-19	-1,54
Comp-20	-1,12	Comp-20	-0,72	Comp-21	-0,73
Comp-22	-0,67	Comp-22	-0,27	Comp-23	-0,28
Comp-27	0,69	Comp-28	-0,42	Comp-28	0,44
Comp-29	0,14	Comp-31	-1,88	Comp-32	-0,50
Comp-54	0,00	Comp-54	1,04	Comp-56	0,03
Comp-56	0,40	Comp-60	0,13	Comp-69	-0,16
Comp-78	0,39	Comp-80	0,42	Comp-94	1,00
Comp-95	1,00	Comp-100	1,20	Comp-102	1,44
Composto 02	-0,04	Composto 05	-1,06	Composto 15	-0,12
Composto 16	-0,18	Composto 18	-1,57	Composto 19	-1,54
Composto 21	-0,73	Composto 24	-0,46	Composto 25	-0,24
Composto 27	0,69	Composto 28	0,44	Composto 29	0,14
Composto 30	0,22	Composto 31	0,00	Composto 32	-0,34
Composto 33	-0,30	Composto 34	-1,34	Composto 35 (R)	-1,82
Composto 35 (S)	-1,82	Desipramina	1,20	Desmetildesipramina	1,06
Desmetildiazepam	0,50	Desmonometilpromazina	0,59	Diazepam	0,52
Diclorometano	-0,11	Didanosina	-1,30	Dietil Éter	0,00
Dissulfeto de Carbono	0,60	Divinil Éter	0,11	Domperidona	-0,78

Continua

Continuação

Enflurano (R)	0,24	Enflurano (S)	0,24	Epóxido de Carbam ^[c]	-0,34
Etanol	-0,16	Etilbenzeno	0,20	FCHCHOCHCHF	0,13
Fenilbutazona	-0,52	Fenitoin	-0,04	Fenserina	1,00
Fisostigmina	0,08	Flumazenil	-0,29	Flunitrazepam	0,06
Fluroxeno	0,13	Guanidinoiazol - A ^[a] 15	-0,18	Guanidinoiazol - A ^[a] 16	-1,57
Guanidinoiazol - A ^[a] 17	-1,54	Guanidinoiazol - A ^[a] 57	-1,15	Guanidinoiazol - A ^[a] 58	-1,54
Halotano (R)	0,35	Halotano (S)	0,35	Heptano	0,81
Hexano	0,80	Hexobarbital (R)	0,10	Hexobarbital (S)	0,10
Hidroxizina (R)	0,39	Hidroxizina (S)	0,39	Ibuprofeno (R)	-0,18
Ibuprofeno (S)	-0,18	ICI 17148	-0,04	Icotidina	-2,00
Imipramina	1,07	Indinavir (R)	-0,74	Indinavir (S)	-0,74
Indometacina	-1,26	Isoflurano (R)	0,42	Isoflurano (S)	0,42
L663581	-0,30	L663581 - Mono Hidro ^[b]	-1,34	Lupitidina	-1,06
Mepiramina	0,49	Mesoridazina (R)	-0,36	Mesoridazina (S)	-0,36
Meta-Xileno	0,29	Metilciclopentano	0,93	Metoxiflurano	0,25
Mianserina	0,99	Midazolam	0,36	Mirtazapina	0,53
Molécula 62	-1,30	Molécula 63	-1,40	Molécula 64	0,43
Molécula 65	0,25	Molécula 66	-0,30	Molécula 67	-0,06
Molécula 68	-0,42	Molécula 69	-0,16	Morfina	-0,16
Nevirapina	0,00	Nitrogênio	0,03	Nor 1 Clorpromazina	1,37
Nor 2 Clorpromazina	0,97	Nortioridazina (R)	0,75	Nortioridazina (S)	0,75
Org32104	0,52	Org34167 (R)	0,00	Org34167 (S)	0,00
Org4428	0,82	Org5222	1,03	Orto-Xileno	0,37
Oxazepam (R)	0,61	Oxazepam (S)	0,61	Óxido Nitroso	0,03
Para-Acetamidofenol	-0,31	Paraxantina	0,06	Para-Xileno	0,31
Pentano	0,76	Promazina	1,23	Propanona	-0,15
Propranolol (R)	0,64	Propranolol (S)	0,64	Quinidina	-0,46
R019 4603	-0,25	Ranitidina	-1,23	Ranitidina - Análogo 17	-1,12
Ranitidina - Análogo 18	-0,27	Ranitidina - Análogo 19	-0,28	Ranitidina - Análogo 20	-0,46
Ranitidina - Análogo 21	-0,24	Ranitidina - Análogo 22	-0,02	Ranitidina - Análogo 23	0,69
Ranitidina - Análogo 24	0,44	Ranitidina - Análogo 26	0,22	Ranitidina - Análogo 60	-0,73
Risperidona	-0,02	SB222200 (R)	0,30	SB222200 (S)	0,30
Sulforidazina (R)	0,18	Sulforidazina (S)	0,18	Teflurano (R)	0,27
Teflurano (S)	0,27	Temelastina	-1,88	Teobromina	-0,28
Teofilina	-0,29	Tertbutilclorambucil	1,00	Teste 1	-1,30
Teste 2	-1,40	Teste 3	-0,43	Teste 4	0,25
Teste 5	-0,30	Teste 6	-0,06	Teste 8	-0,16
Teste 9	0,00	Teste 10	-0,34	Teste 11	-0,30
Teste 12	-1,34	Teste 13	-1,82	Teste2 3	-0,29
Teste2 4	-0,06	Teste2 5	-0,10	Teste2 8	-0,31
Teste2 9	-0,18	Teste2 10	0,11	Teste2 11	0,55
Teste2 13	0,12	Teste2 14	-1,42	Teste2 17	0,50
Teste2 18	-1,26	Teste2 20	0,39	Teste2 22	1,30
Teste2 24	0,36	Teste2 25	-0,70	Teste2 27	1,06
Teste2 29	0,24	Tibolona	0,40	Tiopental (R)	-0,14
Tiopental (S)	-0,14	Tioperamida	-0,16	Tioridazina (R)	0,24
Tioridazina (S)	0,24	Tiotidina	-0,82	Tolueno	0,37
Triazolam	0,74	Tricloroetileno	0,34	Triclorometano	0,29
Trifluoroperazina	1,44	Verapamil (R)	-0,70	Verapamil (S)	-0,70
Y G14	-0,42	Zidovudina	-0,72	Zolantidina	0,14

logBB (log[cérebro]/[sangue]). [a] Análogo. [b] Hidróxido [c] Carbamazepina.

Os principais aspectos discutidos na organização do conjunto de dados de HIA são aplicáveis a este conjunto de dados de BBB e, portanto, não serão discutidos

novamente. O conjunto padrão de dados de BBB possui também alta diversidade química, com notável cobertura do espaço químico-biológico.

A medida da concentração de compostos no cérebro humano é uma tarefa difícil que exige o uso de métodos invasivos de monitoramento contínuo. Por essa razão, os dados disponíveis na literatura são majoritariamente provenientes de testes em animais. Assim, o conjunto de dados organizado neste trabalho é composto por valores da permeabilidade da barreira hematoencefálica de ratos, ao contrário dos conjuntos de HIA e metabolismo que são baseados em estudos realizados em humanos.

Como apresentado na Figura 3.2 da sessão 3.1.3 do capítulo 3, a permeabilidade da barreira hematoencefálica pode ser dividida em três faixas: baixa, moderada e alta. Como podem ser observados na Figura 4.8, 23,92% dos 255 compostos do conjunto de dados encontram-se na faixa de distribuição de baixa permeabilidade (com valores de $\log BB \leq -0,5$). Uma fração de 57,26% dos compostos cai na faixa de distribuição de permeabilidade moderada (valores entre $\log BB -0,49$ e $\log BB 0,5$). Os 18,82% restantes encontram-se na faixa de alta permeabilidade (valores de $\log BB > 0,5$).

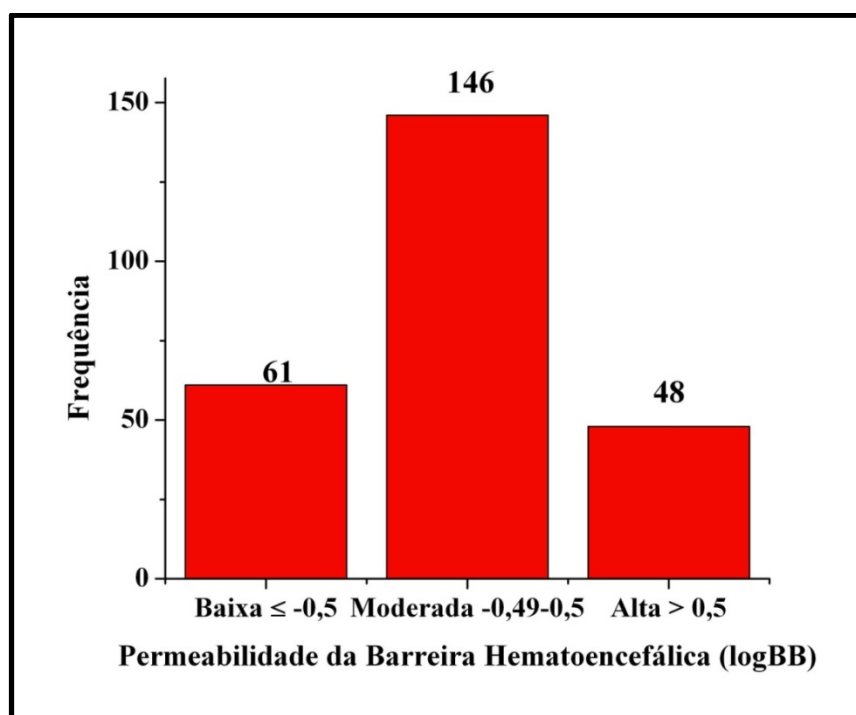


Figura 4.8 – Histograma da distribuição dos valores de permeabilidade da barreira hematoencefálica do conjunto de dados.

4.3.2 Desenvolvimento de Modelos de HQSAR-BBB

O conjunto inicial de 255 compostos foi dividido em conjunto treinamento (utilizado para a geração dos modelos) com 207 compostos (81%) e conjunto teste (utilizado para avaliar a capacidade de predição externa dos modelos) com 48 compostos (19%), seguindo a proporção padrão de 80/20.

O histograma da Figura 4.9 mostra a distribuição do conjunto de dados e dos conjunto treinamento e teste em função do valor da propriedade.

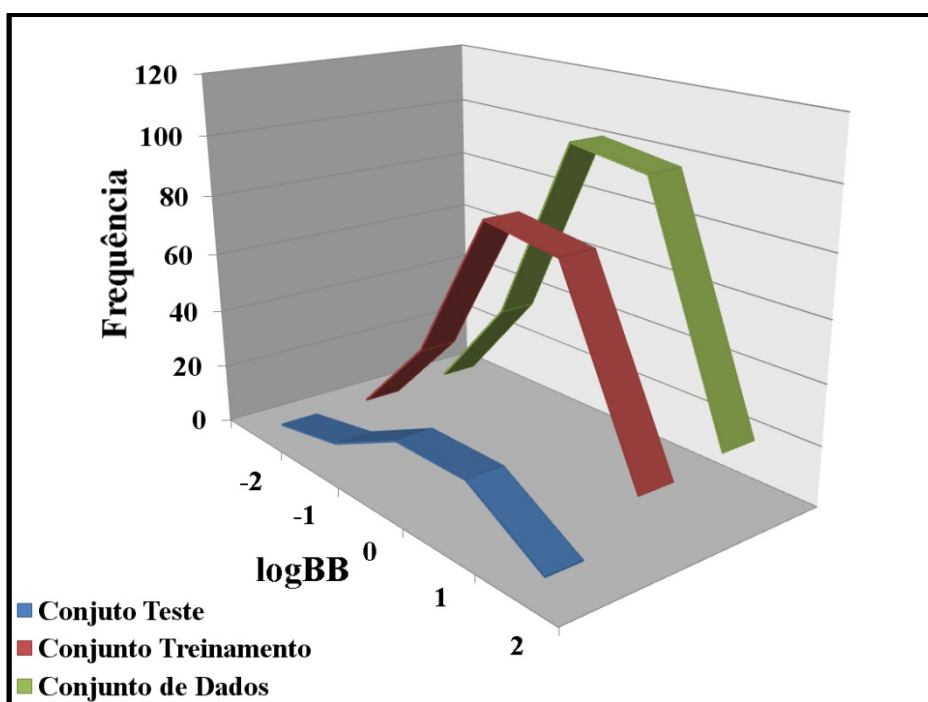


Figura 4.9 – Representação gráfica da distribuição dos valores de logBB.

O conjunto treinamento foi utilizado para a geração de vários modelos através da variação da distinção de fragmentos, mantendo-se constante o tamanho dos fragmentos no intervalo padrão de 4 a 7 átomos. Os resultados estatísticos mais relevantes obtidos a partir da análise PLS são mostrados na Tabela 4.10.

Tabela 4.10 – Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão de fragmento (4-7)

Modelo	Distinção de Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B/C/Ch	0,61	0,81	0,33	8	199
2	A/Ch/DA	0,58	0,79	0,35	8	401
3	A/B/C/H/Ch	0,66	0,87	0,27	8	353
4	A/C/Ch/DA	0,55	0,86	0,29	8	353
5	A/H/Ch/DA	0,66	0,86	0,29	8	353
6	A/B/C/Ch/DA	0,53	0,86	0,29	7	353
7	A/B/H/Ch/DA	0,68	0,88	0,26	8	307
8	A/B/C/H/Ch/DA	0,69	0,91	0,22	8	401

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma. Distinção do fragmento: A, átomos; B, ligação; C, conexão; H, átomos de hidrogênio; Ch, quiralidade; DA, doador e receptor.

O descritor quiralidade (Ch) foi sempre considerado na fragmentação 2D em função do número significativo de moléculas quirais. Os melhores modelos (destacados em cinza), baseados nos valores de q^2 e r^2 e SEE, foram obtidos utilizando-se as combinações A/B/C/H/Ch ($q^2 = 0,66$ e $r^2 = 0,87$), A/H/Ch/DA ($q^2 = 0,66$ e $r^2 = 0,86$), A/B/H/Ch/DA ($q^2 = 0,68$ e $r^2 = 0,88$) e A/B/C/H/Ch/DA ($q^2 = 0,69$ e $r^2 = 0,91$). Foi conduzida também a avaliação da influência da variação do tamanho dos fragmentos (2-5, 3-6, 4-7, 5-8 e 6-9) sobre os parâmetros estatísticos dos melhores modelos.

Tabela 4.11 – Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando as melhores distinções observadas: A/B/C/H/Ch, A/H/Ch/DA, A/B/H/Ch/DA e A/B/C/H/Ch/DA

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
Modelo 3					
2-5	0,68	0,85	0,30	8	307
3-6	0,71	0,87	0,28	8	401
4-7	0,66	0,87	0,27	8	353
5-8	0,60	0,81	0,33	7	307
6-9	0,53	0,80	0,34	7	307
Modelo 5					
2-5	0,64	0,78	0,36	8	71
3-6	0,65	0,84	0,30	8	353
4-7	0,66	0,86	0,29	8	353
5-8	0,64	0,86	0,29	8	353
6-9	0,62	0,86	0,29	8	257
Modelo 7					
2-5	0,71	0,85	0,29	8	199
3-6	0,68	0,86	0,28	8	353
4-7	0,68	0,88	0,26	8	307
5-8	0,68	0,87	0,27	8	307
6-9	0,61	0,84	0,31	7	307
Modelo 8					
2-5	0,68	0,89	0,25	8	307
3-6	0,70	0,91	0,23	8	401
4-7	0,69	0,91	0,22	8	401
5-8	0,64	0,85	0,29	6	401
6-9	0,60	0,85	0,29	6	353

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N, número ótimo de componentes; HL, comprimento do holograma.

A investigação da variação do tamanho dos fragmentos levou a melhores parâmetros estatísticos para os modelos **3**, **7** e **8**, sendo os novos fragmentos de tamanho 3-6, 2-5, 3-6 átomos, respectivamente (destacados em cinza na Tabela 4.11).

O modelo 5 (A/H/Ch/DA) não mostrou melhora nos parâmetros com a variação do número de átomos que compõe os fragmentos.

4.3.3 Validação dos Modelos de HQSAR-BBB

Os melhores modelos descritos na Tabela 4.11, já validados pelos métodos LOO e LMO (com 12 grupos), foram empregados no processo de validação externa através da predição de um conjunto teste contendo 48 moléculas. Os resultados são apresentados na Tabela 4.12.

Tabela 4.12 – Valores experimentais e preditos de permeabilidade da barreira hematoencefálica associados aos resíduos para os compostos do conjunto teste

No.	Composto	Exp ^[a]	Pred 1 ^[b]	Pred 2 ^[c]	Pred 3 ^[d]	Pred 4 ^[e]	Cons ^[f]	Res ^[g]
Conjunto Teste								
01	1,1 1-tricloroetano	0,40	0,01	0,09	0,18	0,06	0,09	0,22
02	1-hidroximidazolam	-0,07	0,20	0,01	0,42	0,07	0,18	-0,49
03	Aminopirina	0,00	-0,38	-0,03	-0,15	-0,77	-0,33	0,15
04	Cimetidina - Análogo 13	-0,66	-0,63	-0,51	-0,77	-0,62	-0,63	0,11
05	Clonidina	0,11	-0,30	-1,17	-0,10	-0,91	-0,62	0,21
06	Clorpromazina	1,06	0,86	1,12	0,95	0,43	0,84	0,11
07	Comp-10	-0,82	-1,61	-1,22	-1,47	-1,51	-1,45	0,65
08	Comp-20	-1,12	-0,69	-0,66	-0,80	-0,90	-0,76	-0,32
09	Comp-20	-0,72	-1,36	-0,95	-1,38	-0,99	-1,17	0,66
10	Comp-29	0,14	0,32	0,16	0,44	0,64	0,39	-0,30
11	Comp-32	-0,50	0,07	0,25	-0,26	0,09	0,04	-0,24
12	Comp-5	-1,06	-1,72	-1,61	-1,84	-2,08	-1,81	0,78
13	Comp-56	0,03	0,23	0,08	0,26	0,14	0,18	-0,23
14	Comp-94	1,00	0,77	0,43	0,97	0,37	0,63	0,03
15	Composto 25	-0,24	0,29	0,28	0,29	0,06	0,23	-0,53
16	Composto 35 (S)	-1,82	-1,62	-1,76	-1,70	-1,34	-1,61	-0,12
17	Flunitrazepam	0,06	0,27	0,54	0,40	0,48	0,42	-0,34
18	Fluroxeno	0,13	0,24	-0,04	0,43	0,04	0,17	-0,30
19	Guanidinothiazol – A ^[h] 58	-1,54	-1,64	-1,17	-1,17	-0,99	-1,24	-0,37
20	Icotidine	-2,00	-1,55	-1,62	-1,82	-1,58	-1,65	-0,18
21	Indinavir (S)	-0,75	-1,22	-1,03	-1,05	-1,04	-1,08	0,30
22	Indometacina	-1,26	-0,42	-0,49	-0,59	0,14	-0,34	-0,67
23	Mepiramina	0,49	0,15	0,23	0,47	0,45	0,32	0,02
24	Mianserina	0,99	1,02	1,12	0,91	1,27	1,08	0,08
25	Molécula 64	0,43	-0,35	-0,11	-0,23	-0,05	-0,18	0,66
26	Molécula 68	-0,42	-0,59	-0,61	-0,67	-1,05	-0,73	0,25
27	Molécula 69	-0,16	-0,12	-0,02	-0,07	-0,15	-0,09	-0,09
28	Nor 2 Clorpromazina	0,97	0,60	0,95	0,59	0,55	0,67	0,38
29	Oxazepam (S)	0,61	0,50	0,36	0,25	0,33	0,36	0,36

Continua

Continuação

30	Paraxantina	0,06	-0,19	-0,21	-0,39	0,13	-0,17	0,45
31	Para-Xileno	0,31	0,38	-0,01	0,10	0,31	0,19	0,21
32	Propanona	-0,15	-0,05	0,00	0,09	-0,11	-0,02	-0,24
33	Propranolol (R)	0,64	-0,15	-0,40	-0,21	0,29	-0,12	0,85
34	R019 4603	-0,25	0,10	0,03	-0,12	0,17	0,04	-0,13
35	Risperidona	-0,02	-0,56	-0,15	-0,16	-0,40	-0,32	0,14
36	SB222200 (R)	0,30	-0,32	0,32	0,02	0,08	0,03	0,28
37	Teflurano (R)	0,27	0,34	0,08	0,29	0,17	0,22	-0,02
38	Teofilina	-0,29	-0,31	-0,71	-0,43	-1,00	-0,61	0,14
39	Teste 10	-0,34	-0,03	0,34	-0,25	0,32	0,09	-0,09
40	Teste 2	-1,40	-0,66	-0,52	-0,88	-0,87	-0,73	-0,52
41	Teste2 4	-0,06	-0,26	-0,06	-0,10	-0,53	-0,24	0,04
42	Teste2 5	-0,10	-0,36	-0,24	-0,08	-0,25	-0,23	-0,02
43	Teste2 8	-0,31	-0,32	-0,22	-0,61	-0,16	-0,33	0,30
44	Tioperamida	-0,16	0,05	-0,10	-0,11	0,04	-0,03	-0,05
45	Tioridazina (R)	0,24	0,24	0,06	0,47	0,31	0,27	-0,23
46	Tolueno	0,37	0,41	0,10	0,24	0,34	0,27	0,13
47	Triazolam	0,74	0,27	0,38	0,18	0,27	0,28	0,56
48	Trifluoroperazina	1,44	0,98	1,39	1,47	1,46	1,32	-0,03
$r^2_{\text{pred}}^{[i]}$		0,72	0,69	0,79	0,62	0,75		

[a] Valor experimental de logBB. [b] Valor predito pelo modelo 3. [c] Valor predito pelo modelo 5. [d] Valor predito pelo modelo 7. [e] Valor predito pelo modelo 8. [f] Predição consenso. [g] Resíduo: diferença entre o valor experimental e a predição realizada pelo modelo 7. [h] Análogo. [i] coeficiente de correlação entre o valor experimental e predito do conjunto teste.

O resultado da validação externa mostra que o conjunto teste, o qual representa diferentes características químico-estruturais quando comparado com o conjunto treinamento, é razoavelmente bem predito em todos os modelos selecionados. Entretanto, nesse caso, a análise dos r^2_{pred} para os modelos selecionados não indicou a predição consenso como melhor escolha. O coeficiente de correlação do conjunto teste para o modelo 7 ($r^2_{\text{pred}} = 0,79$) mostra maior capacidade de predição quando comparado com os modelos 3 ($r^2_{\text{pred}} = 0,72$), 5 ($r^2_{\text{pred}} = 0,69$), 8 ($r^2_{\text{pred}} = 0,62$) e o método de predição consenso ($r^2_{\text{pred}} = 0,75$). Assim, o modelo 7, com tamanho dos fragmentos 3-6, foi escolhido para a predição da propriedade de novas moléculas bioativas.

Em geral, os valores preditos pelo modelo 7 desviam dos experimentais por não mais que 0,5 unidades logarítmicas. As exceções são os compostos 7, 9, 12, 15, 25, 33, 40 e 47, para os quais os valores preditos apresentam erro mais considerável (entre 0,52 e 0,85 unidades logarítmicas). A boa concordância entre os valores de logBB experimental e predito para os compostos do conjunto teste indica a robustez do modelo preditivo obtido com o método de HQSAR. O gráfico das relações entre os valores experimentais e preditos de logBB dos compostos dos conjuntos treinamento e teste é apresentado na Figura 4.10.

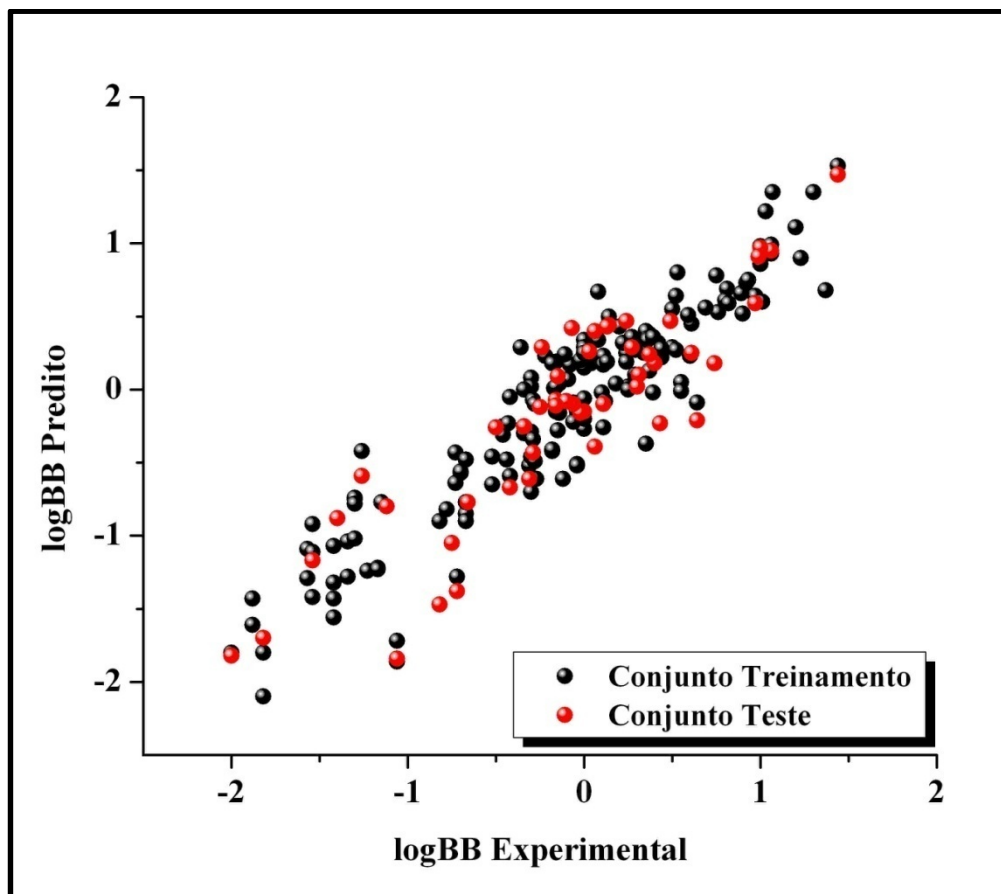


Figura 4.10 – Gráfico de HQSAR dos valores experimentais e preditos de permeabilidade da barreira hematoencefálica para os 255 compostos do conjunto de dados.

O gráfico confirma a boa correlação do conjunto de dados, indicando que o modelo final de HQSAR é útil para a predição de logBB de moléculas candidatas a novos fármacos.

4.4 Inibição da Glicoproteína-P

4.4.1 Organização do Conjunto de Dados

Como descrito na sessão 3.1.4 do capítulo 3, foram coletados 159 compostos para modelagem da inibição da P-gp¹¹⁸. A lista com as estruturas dos inibidores da P-gp e os valores de pEC₅₀ é apresentada na Tabela 4.13.

Tabela 4.13 – Estruturas químicas e valores associados de inibição (pEC₅₀) do conjunto de dados de moduladores da glicoproteína-P

Composto	Estrutura	R1-R4	L	A	pEC ₅₀
Mol-1	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A9	2,22
Mol-2	A	R1 = COCH ₂ CH ₂ -(4-N(CH ₃) ₂ -Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,89
Mol-3	A	R1 = COCH ₂ CH ₃	-CH ₂ CH(OH)CH ₂ -	A2	-2,32
Mol-4	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A2	-1,83

Continua

Continuação

Mol-5	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A2	-2,11
Mol-6	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-1,69
Mol-7	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₆ -	A1	0,69
Mol-8	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₇ -	A1	0,75
Mol-9	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₈ -	A1	2,40
Mol-10	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph) ₂)	-0,77
Mol-11	A	R1 = CH(OH)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,98
Mol-12	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	0,55
Mol-13	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A9	-0,13
Mol-14	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	-0,19
Mol-15	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A9	0,42
Mol-16	C	R1 = CH ₂ CH ₂ Ph	-CO-	A1	-0,82
Mol-17	C	R1 = CH ₃	-CO-	A1	-2,02
Mol-18	C	R1 = CH ₂ CH ₂ Ph	-CH ₂ -	A1	-0,21
Mol-19	C	R1 = CH ₂ CH ₂ Ph	-CO-	A1	-0,82
Mol-20	C	R1 = CH ₃	-CO-	A1	-2,02
Mol-21	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A9	0,86
Mol-22	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(nPr)	0,48
Mol-23	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(C ₂ H ₅) ₂	0,04
Mol-24	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,22
Mol-25	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	0,42
Mol-26	A	R1 = COCH ₂ CH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-1,16
Mol-27	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-1,50
Mol-28	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A3	0,21
Mol-29	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A4	0,64
Mol-30	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A5	0,17
Mol-31	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A6	0,59
Mol-32	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A7	1,57
Mol-33	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A10	0,18
Mol-34	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A8	1,15
Mol-35	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A8	-0,32
Mol-36	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(CH ₃) ₂	-0,39
Mol-37	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(CH ₃)nPr	0,38
Mol-38	A	R1 = CPh	-CH ₂ CH(OH)CH ₂ -	-N(C ₂ H ₅) ₂	-0,37
Mol-39	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,56
Mol-40	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A12	1,24
Mol-41	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,01
Mol-42	A	R1 = CH(OH)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,05
Mol-43	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,76
Mol-44	B	R1 = CH ₂ CH ₂ Ph	-(CH ₂) ₂ -	A1	0,01
Mol-45	B	R1 = CH ₂ CH ₃	-CH(OH)CH ₂ -	-NH(nPr)	-1,06
Mol-46	B	R1 = CH ₂ CH ₂ Ph	-CH(OH)CH ₂ -	-NH(nPr)	-0,02
Mol-47	B	R1 = CH ₂ CH ₂ Ph	-CH(OH)CH ₂ -	-NH(nPr)	-0,04
Mol-48	B	R1 = CH ₂ CH ₂ Ph	-CH(OH)CH ₂ -	A4	0,75
Mol-49	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A11	0,59
Mol-50	A	R1 = COCH ₂ CH ₂ Ph, R3 = OH	-CH ₂ CH(OH)CH ₂ -	-NH(nPr)	-0,48
Mol-51	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A8	-0,40
Mol-52	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,37
Mol-53	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,84
Mol-54	A	R1 = CH(OH)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A8	0,17
Mol-55	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A8	0,65

Continua

Continuação

Mol-56	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A8	-0,05
Mol-57	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	0,04
Mol-58	A	R1 = CH(OH)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	-0,24
Mol-59	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	0,18
Mol-60	A	R1 = COCH ₂ CH ₂ -(1-naftil)	-CH ₂ CH(OH)CH ₂ -	A1	0,76
Mol-61	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH(Ph)) ₂	0,10
Mol-62	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH(Ph)) ₂	-1,02
Mol-63	A	R1 = COCH ₂ CH ₂ -(1-naftil)	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph)) ₂	0,14
Mol-64	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₃ -	A1	0,19
Mol-65	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₂ -	A1	-0,16
Mol-66	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₄ -	A1	0,27
Mol-67	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₅ -	A1	0,62
Mol-68	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₆ -	A1	0,69
Mol-69	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₇ -	A1	0,75
Mol-70	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₈ -	A1	0,86
Mol-71	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph)) ₂	0,12
Mol-72	A	R1 = CH(OH)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,98
Mol-73	A	R1 = CH(OCH ₃)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,26
Mol-74	A	R1 = COCH ₂ CH ₂ Ph, R3 = OBz	-CH ₂ CH(OH)CH ₂ -	-NH(nPr)	0,96
Mol-75	A	R1 = COCH ₂ CH ₂ Ph, R3 = OBz	-CH ₂ CH(OH)CH ₂ -	A1	0,76
Mol-76	A	R1 = COCH ₂ CH ₂ Ph, R3 = OH	-CH ₂ CH(OH)CH ₂ -	A1	-0,24
Mol-77	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph)) ₂	0,14
Mol-78	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(Ph)	-0,81
Mol-79	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(ciclohexil)	0,47
Mol-80	A	R1 = COCH ₂ CH ₂ Ph, R3 = OH	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	-0,02
Mol-81	A	R1 = COCH ₂ CH ₂ Ph, R3 = OBz	-CH ₂ CH(OH)CH ₂ -	A10	0,94
Mol-82	A	R1 = CH(OiPr)CH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A1	0,34
Mol-83	A	R1 = COPh	-CH ₂ CH(OH)CH ₂ -	A12	0,51
Mol-84	A	R1 = COPh	-CH ₂ CH(OH)CH ₂ -	A8	0,83
Mol-85	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A11	0,45
Mol-86	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A12	-0,50
Mol-87	A	R1 = COCH ₂ CH ₂ (4-Cl-Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,72
Mol-88	A	R1 = COCH ₂ CH ₂ (4-CH ₃ -Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,74
Mol-89	A	R1 = COCH ₂ CH ₂ (4-OCH ₃ -Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,85
Mol-90	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OAc)CH ₂ -	A1	0,47
Mol-91	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(Ph-4-COOCH ₃)	-0,19
Mol-92	A	R1 = COCH ₂ CH ₂ Ph, R3 = Cl	-CH ₂ CH(OH)CH ₂ -	A7	1,23
Mol-93	A	R1 = COCH ₂ CH ₂ Ph, R3 = OCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,75
Mol-94	A	R1 = COCH ₂ CH ₂ Ph, R3 = CH ₃	-CH ₂ CH(OH)CH ₂ -	A7	1,59
Mol-95	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(Ph-4-CF ₃)	-0,93
Mol-96	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(Ph-4-NO ₂)	-0,41
Mol-97	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(COCH ₂ CH ₃)Bz	-0,22
Mol-98	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₄ -	A2	0,00
Mol-99	A	R1 = COCH ₂ CH ₂ Ph	-(CH ₂) ₄ -	A8	0,72
Mol-100	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-N(nPr)COPh	-0,23
Mol-101	A	R1 = COCH ₂ CH ₂ -(1-naftil)	-CH ₂ CH(OH)CH ₂ -	A2	0,14
Mol-102	A	R1 = COCH ₂ CH ₂ -(1-naftil)	-CH ₂ CH(OH)CH ₂ -	-N(iPr) ₂	0,33
Mol-103	A	R1 = COCH ₂ CH ₃	-CH ₂ CH(OH)CH ₂ -	A12	0,52
Mol-104	A	R1 = COCH ₂ CH ₂ -(1-naftil)	-CH ₂ CH(OH)CH ₂ -	A12	1,12
Mol-105	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-0,96
Mol-106	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A8	-1,00

Continua

Continuação

Mol-107	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A13	0,89
Mol-108	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A8	-1,08
Mol-109	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A2	-2,48
Mol-110	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NHCH(iPr)Ph	0,68
Mol-111	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NHCH(ciclohexil)Ph	0,84
Mol-112	A	R1 = COCH ₂ CH ₂ Ph, R3 = Cl, R4 = Cl	-CH ₂ CH(OH)CH ₂ -	A7	1,77
Mol-113	A	R1 = COCH ₃ , R3 = CH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-0,74
Mol-114	E	R1 = H, R2 = O, R3 = H,H	-CH ₂ CH(OH)CH ₂ -	A1	-0,99
Mol-115	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NHCH(tBu)Ph	0,53
Mol-116	E	R1 = H, R2 = H,H, R3 = O	-CH ₂ CH(OH)CH ₂ -	A1	-1,05
Mol-117	E	R1 = CH ₂ Ph, R2 = O, R3 = H,H	-CH ₂ CH(OH)CH ₂ -	A1	0,67
Mol-118	E	R1 = CH ₂ Ph, R2 = H,H, R3 = O	-CH ₂ CH(OH)CH ₂ -	A1	0,85
Mol-119	E	R1 = CH ₂ Ph, R2 = O, R3 = H,H	-CH ₂ CH(OH)CH ₂ -	A9	0,78
Mol-120	D	-----	-CH ₂ -	A1	-2,25
Mol-121	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ COOEt	-----	-1,37
Mol-122	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ COOH	-----	-3,20
Mol-123	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A9	-0,13
Mol-124	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	-0,19
Mol-125	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A9	0,42
Mol-126	C	R1 = CH ₂ CH ₂ Ph	-CO-	A1	-0,82
Mol-127	C	R1 = CH ₃	-CO-	A1	-2,02
Mol-128	C	R1 = CH ₂ CH ₂ Ph	-CH ₂ -	A1	-0,21
Mol-129	C	R1 = CH ₃	-CH ₂ -	A1	-1,69
Mol-130	A	R2 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,27
Mol-131	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A9	0,70
Mol-132	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A9	0,64
Mol-133	A	R1 = COCH ₂ CH ₂ Ph, R4 = OCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,53
Mol-134	A	R1 = COCH ₃ , R3 = OCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,07
Mol-135	A	R1 = COCH ₃ , R3 = CH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,40
Mol-136	A	R1 = COCH ₃ , R3 = COCH ₃ , R4 = CH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,68
Mol-137	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,60
Mol-138	A	R1 = COCH ₃ , R4 = CH ₃	-CH ₂ CH(OH)CH ₂ -	A7	0,79
Mol-139	A	R1 = COCH ₃ , R3 = Cl, R4 = Cl	-CH ₂ CH(OH)CH ₂ -	A7	0,90
Mol-140	A	R1 = COCH ₂ CH ₂ -(4-N(CH ₃) ₂ -Ph), R4 = OCH ₃	-CH ₂ CH(OH)CH ₂ -	A7	1,40
Mol-141	A	R3 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph ₂))	0,75
Mol-142	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A7	1,13
Mol-143	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph ₂))	0,72
Mol-144	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A9	0,86
Mol-145	A	R2 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A2	-0,54
Mol-146	A	R3 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	-NH(CH ₂ CH ₂ CH(Ph ₂))	0,30
Mol-147	A	R1 = COCH ₂ CH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-1,16
Mol-148	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A1	-1,14
Mol-149	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A3	0,21
Mol-150	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A4	0,14
Mol-152	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A5	-0,08
Mol-153	A	R1 = COCH ₂ CH ₂ Ph	-CH ₂ CH(OH)CH ₂ -	A6	0,59
Mol-154	A	R1 = CPh	-CH ₂ CH(OH)CH ₂ -	A12	0,51
Mol-155	A	R1 = CPh	-CH ₂ CH(OH)CH ₂ -	A8	0,83

Continua

Continuação

Mol-156	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A11	0,45
Mol-157	A	R1 = COCH ₃	-CH ₂ CH(OH)CH ₂ -	A12	-0,50
Mol-158	A	R1 = COCH ₂ CH ₂ (4-Cl-Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,72
Mol-159	A	R1 = COCH ₂ CH ₂ (4-CH ₃ -Ph)	-CH ₂ CH(OH)CH ₂ -	A7	1,74

Os valores de inibição (EC₅₀) encontram-se entre 0,006 e 1588 µM e correspondem a um fator superior a 250.000 vezes. Os valores da propriedade foram convertidos em pEC₅₀ (-logEC₅₀), como apresentados na Tabela 4.13, e variam em sete ordens de magnitude, estando distribuídos de forma adequada.

4.4.2 Desenvolvimento de Modelos de HQSAR-P-gp

Para a geração dos modelos *in silico* utilizou-se a plataforma computacional Sybyl 8.0. Inicialmente os 159 compostos do conjunto de dados foram divididos em conjuntos treinamento e teste. O conjunto treinamento é formado por 127 (79,9%) compostos, enquanto que o conjunto teste por 32 (20,1%) (Figura 4.11), em uma proporção de 80:20.

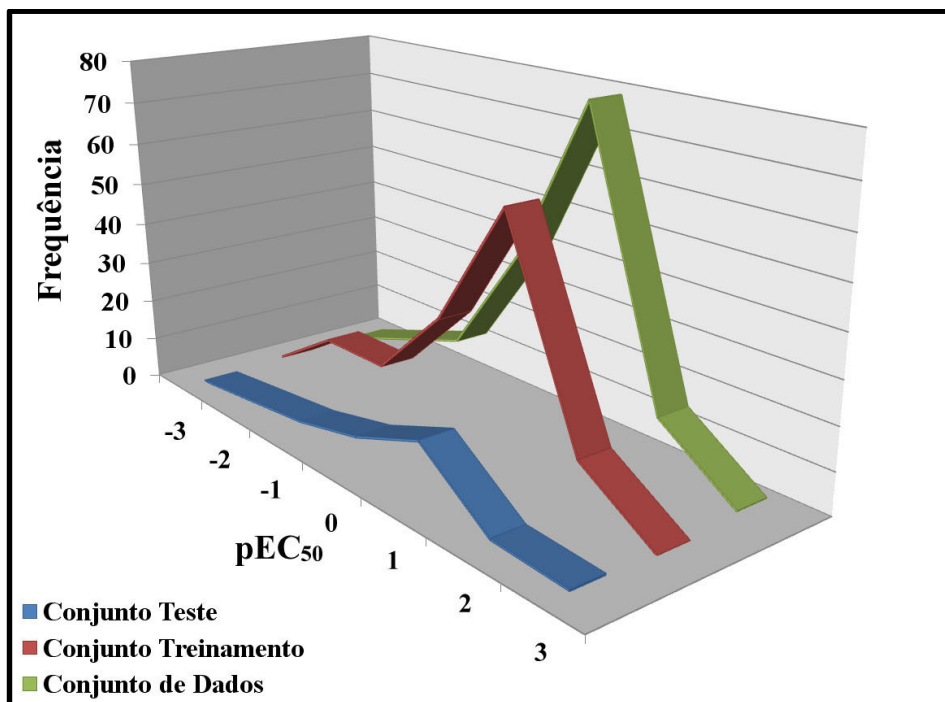


Figura 4.11 – Representação gráfica da distribuição dos valores de pEC₅₀.

No processo de modelagem de QSAR, o conjunto treinamento (127 compostos) foi usado na geração de vários modelos de HQSAR através da análise da variação de distinção de fragmentos, mantendo-se constante o parâmetro de tamanho dos fragmentos, no intervalo padrão de 4-7. Os resultados estatísticos obtidos a partir da análise PLS para os 127 compostos do conjunto treinamento, empregando as várias combinações de distinção de fragmentos, são mostrados na Tabela 4.14. Os melhores resultados estatísticos foram obtidos usando as combinações: A/B/H, modelo **3** ($q^2 = 0,70$ e $r^2 = 0,88$); A/B/C/H, modelo **5** ($q^2 = 0,71$ e $r^2 = 0,78$), A/H/DA, modelo **8** ($q^2 = 0,70$ e $r^2 = 0,84$) e A/C/H/DA, modelo **9** ($q^2 = 0,70$ e $r^2 = 0,88$).

Tabela 4.14 – Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão do fragmento (4-7)

Modelo	Distinção de Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B	0,67	0,79	0,50	6	53
2	A/B/C	0,66	0,89	0,39	8	307
3	A/B/H	0,70	0,88	0,38	8	257
4	A/DA	0,57	0,69	0,61	5	199
5	A/B/C/H	0,71	0,78	0,51	5	59
6	A/B/DA	0,65	0,87	0,41	8	401
7	A/C/DA	0,65	0,87	0,39	8	151
8	A/H/DA	0,70	0,84	0,45	7	199
9	A/C/H/DA	0,70	0,88	0,39	8	151
10	A/B/C/DA	0,64	0,82	0,47	8	59
11	A/B/H/DA	0,69	0,81	0,49	6	59
12	A/B/C/H/DA	0,67	0,82	0,47	7	59

q^2 , coeficiente de correlação com validação cruzada. r^2 , coeficiente de correlação sem validação cruzada. SEE, erro padrão sem validação cruzada. N , número ótimo de componentes. HL, comprimento do holograma. Distinção do fragmento: A, átomos. B, ligação. C, conexão. H, átomos de hidrogênio. Ch, quiralidade. DA, doador e receptor.

A distinção quiralidade (Ch) não foi considerada, pois na base de dados não há moléculas com centros quirais. Os modelos **3**, **5**, **8** e **9**, destacados em cinza na Tabela 4.14, foram selecionados para a investigação da influência da variação do tamanho dos fragmentos. Esses parâmetros podem ser modificados com o objetivo de incluir fragmentos maiores ou menores, alterando os padrões de reconhecimento molecular do holograma. Assim, foram realizadas análises individuais empregando diferentes tamanhos dos fragmentos: 2-5, 3-6, 4-7, 5-8 e 6-9, como mostra a Tabela 4.15.

Tabela 4.15 – Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando as melhores distinções observadas A/B/H, A/B/C/H, A/H/DA e A/C/H/DA

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
Modelo 3					
2-5	0,73	0,78	0,52	3	59
3-6	0,70	0,78	0,52	5	151
4-7	0,7	0,88	0,38	8	257
5-8	0,69	0,87	0,40	8	257
6-9	0,70	0,84	0,45	6	257
Modelo 5					
2-5	0,72	0,77	0,53	3	53
3-6	0,70	0,78	0,51	4	83
4-7	0,71	0,78	0,51	5	59
5-8	0,70	0,85	0,43	6	307
6-9	0,71	0,81	0,48	6	97
Modelo 8					
2-5	0,72	0,82	0,47	6	307
3-6	0,71	0,79	0,51	4	151
4-7	0,70	0,84	0,45	7	199
5-8	0,71	0,86	0,42	8	199
6-9	0,70	0,85	0,44	7	199
Modelo 9					
2-5	0,72	0,78	0,51	4	61
3-6	0,71	0,78	0,51	4	71
4-7	0,70	0,88	0,39	8	151
5-8	0,71	0,88	0,39	8	199
6-9	0,71	0,86	0,42	7	307

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma.

A variação do tamanho dos fragmentos influenciou positivamente os modelos **3**, **5**, **8** e **9**, usando tamanho dos fragmentos de 2-5 átomos.

4.4.3 Validação dos Modelos de HQSAR-P-gp

Os modelos finais possuem validação cruzada LOO e LMO (com 13 grupos). Conforme discutido anteriormente, os modelos de HQSAR devem ser capazes de prever valores da propriedade alvo. A validação externa dos melhores modelos de HQSAR, obtidos com o conjunto treinamento de 127 compostos (distinções de fragmento: A/B/H, A/B/C/H, A/H/DA e A/C/H/DA; tamanho do fragmento: 2-5), foi feita a partir de um conjunto teste de 32 compostos. O resultado do processo de predição é apresentado na Tabela 4.16.

Tabela 4.16 – Valores experimentais e preditos de inibição da P-gp associados aos resíduos para os compostos do conjunto teste

No.	Composto	Exp ^[a]	Pred 1 ^[b]	Pred 2 ^[c]	Pred 3 ^[d]	Pred 4 ^[e]	Cons ^[f]	Res ^[g]
Conjunto Teste								
001	Mol-1	2,22	1,47	1,44	1,38	1,43	1,43	0,79
002	Mol-6	-1,69	-1,18	-1,01	-0,98	-0,83	-1,00	-0,69
003	Mol-9	-2,40	-2,34	-2,14	-1,70	-2,06	-2,06	-0,34
004	Mol-10	-0,77	-1,47	-1,38	-1,78	-1,14	-1,44	0,67
005	Mol-12	0,55	0,56	0,44	0,27	0,26	0,38	0,17
006	Mol-22	0,48	-0,71	-0,78	-0,53	-0,78	-0,70	1,18
007	Mol-25	0,42	-0,76	-0,77	-0,65	-0,96	-0,78	1,21
008	Mol-27	-1,50	-1,14	-0,98	-0,89	-0,84	-0,96	-0,54
009	Mol-30	0,17	0,49	0,69	0,70	0,70	0,65	-0,48
010	Mol-40	1,24	0,54	0,43	0,89	0,31	0,54	0,69
011	Mol-45	-1,06	-0,85	-1,10	-0,89	-1,22	-1,02	-0,05
012	Mol-46	-0,02	-0,18	-0,50	-0,50	-0,58	-0,44	0,41
013	Mol-52	0,37	-0,22	-0,14	-0,16	-0,07	-0,15	0,52
014	Mol-57	0,04	-0,07	-0,10	-0,06	-0,32	-0,14	0,18
015	Mol-64	0,19	-0,11	-0,11	-0,11	-0,05	-0,10	0,28
016	Mol-72	-0,98	-0,39	-0,58	-0,68	-0,62	-0,57	-0,41
017	Mol-74	0,96	0,05	0,00	0,06	-0,08	0,01	0,95
018	Mol-76	-0,24	-0,21	-0,13	0,09	0,00	-0,06	-0,18
019	Mol-84	0,83	-0,29	-0,03	-0,36	0,05	-0,16	0,99
020	Mol-87	1,72	0,93	1,00	0,96	0,99	0,97	0,75
021	Mol-96	-0,41	-0,77	-0,80	-0,72	-0,90	-0,80	0,39
022	Mol-106	-1,00	-0,62	-0,40	-0,80	-0,60	-0,60	-0,40
023	Mol-110	0,68	-0,08	-0,05	0,38	-0,10	0,04	0,64
024	Mol-124	-0,19	-0,03	0,07	0,06	0,27	0,09	-0,28
025	Mol-126	-0,82	-0,18	-0,29	-0,81	-0,34	-0,40	-0,42
026	Mol-127	-2,02	-1,14	-1,22	-1,54	-1,25	-1,29	-0,73
027	Mol-134	0,07	0,08	0,23	0,40	0,34	0,26	-0,20
028	Mol-137	0,60	0,01	0,09	0,15	0,26	0,13	0,48
029	Mol-141	0,75	-0,49	-0,62	-0,60	-0,52	-0,56	1,31
030	Mol-144	0,86	0,33	0,48	0,07	0,16	0,26	0,60
031	Mol-150	0,14	-1,29	-1,04	-1,25	-1,17	-1,19	1,33
032	Mol-151	-0,08	-0,39	-0,49	-0,51	-0,44	-0,46	0,38
r^2_{pred} ^[h]			0,65	0,65	0,64	0,61	0,66	

Continua

Continuação

[a] Valor experimental de pEC₅₀. [b] Valor predito pelo modelo 3. [c] Valor predito pelo modelo 5. [d] Valor predito pelo modelo 8. [e] Valor predito pelo modelo 9. [f] Predição consenso. [g] Resíduo: diferença entre o valor experimental e a predição consenso. [h] coeficiente de correlação entre o valor experimental e predito do conjunto teste.

Em todos os modelos selecionados, o conjunto teste foi predito de forma bastante satisfatória. De maneira similar aos modelos de QSAR-HIA, o método de predição consenso ($[\text{Pred } 1 + \text{Pred } 2 + \text{Pred } 3 + \text{Pred } 4]/4$) foi o mais robusto (maior r^2_{pred}). A diferença entre os valores experimentais e preditos (resíduos da Tabela 4.16) indica a boa capacidade preditiva do método consenso. Os valores preditos se aproximam dos experimentais, desviando-se em não mais que 1 unidade logarítmica. As únicas exceções são os compostos 6, 7, 29 e 31 para os quais os valores apresentam uma diferença mais significativa (entre 1,18 e 1,33 unidades logarítmicas). A Figura 4.12 apresenta o gráfico dos valores experimentais contra os valores preditos de pEC₅₀ para os conjuntos treinamento e teste.

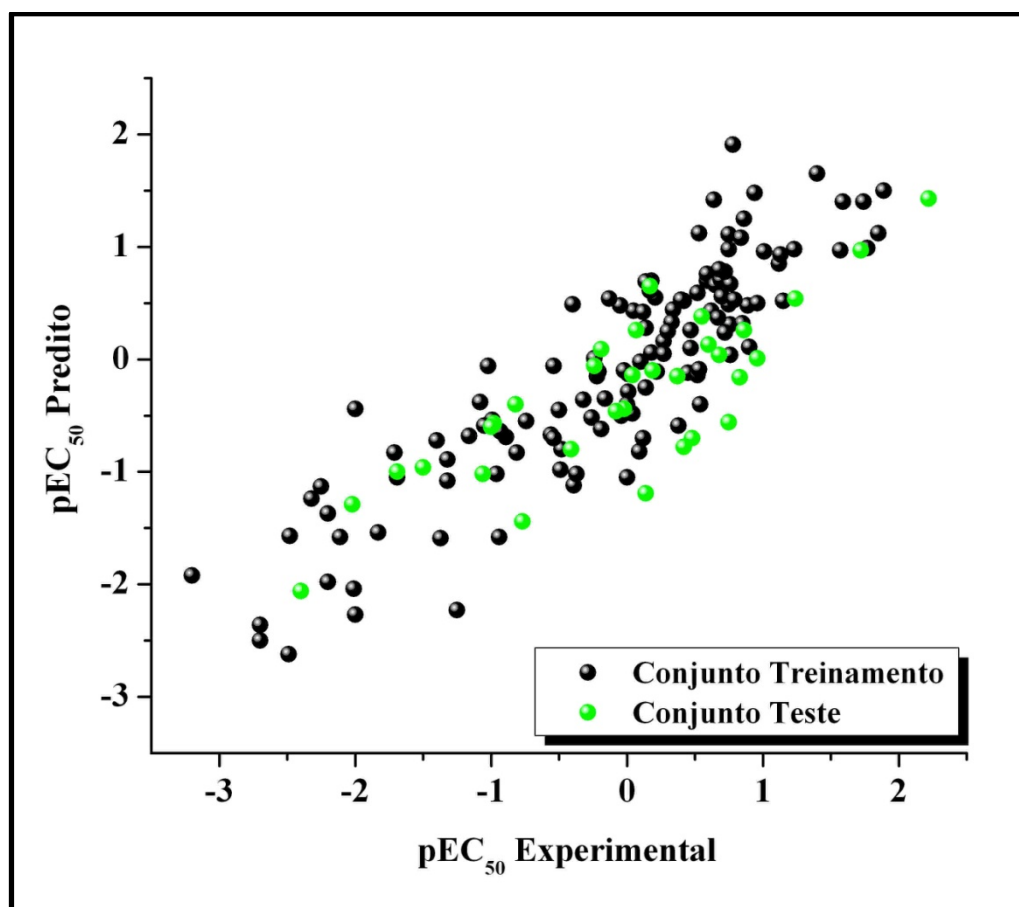


Figura 4.12 – Gráfico de HQSAR dos valores experimentais e preditos de pEC₅₀ para os 159 compostos do conjunto de dados.

Pode-se observar a boa concordância entre os valores preditos e experimentais, indicando a consistência interna e externa dos modelos de HQSAR para a inibição da glicoproteína-P. Os modelos são significativos e podem ser empregados no planejamento de NCEs.

4.4.4 Contribuição dos Fragmentos

Além de prever o valor da propriedade de novas moléculas não incluídas no conjunto de treinamento, os modelos de HQSAR fornecem importantes indicadores sobre os diversos fragmentos moleculares que estão diretamente relacionados com a propriedade biológica em estudo. Estas informações estruturais, diretamente associadas com as contribuições em termos de inibição da P-gp nos modelos gerados, são úteis no planejamento de novas moléculas^{121,125}. As informações são codificadas nos mapas de contribuições de HQSAR, que determinam as contribuições individuais de diferentes fragmentos em uma determinada molécula com relação à propriedade alvo. O módulo de HQSAR implementado na plataforma SYBYL 8.0 utiliza um código de cores para atribuir as diferentes magnitudes de contribuição de cada átomo para a propriedade alvo, refletindo a contribuição dos diferentes fragmentos da molécula. Cores na região do vermelho do espectro (vermelho, laranja avermelhado e laranja) refletem contribuições desfavoráveis ou negativas. Cores na região do verde (amarelo, azul esverdeado e verde) refletem contribuições favoráveis ou positivas. Átomos com contribuições intermediárias ou neutras são representados em branco. A Figura 4.13 mostra a contribuição atômica individual para o composto mais potente (**001**, pEC₅₀ = 2,22 (Tabela 4.16)) do conjunto de dados. Particularmente, foi revelada a importância da fenila e da carbonila atachada, bem como do oxigênio fenólico em *orto*. Em contrapartida, para quaisquer moléculas do conjunto de dados, regiões com contribuições negativas ou intermediárias podem ser identificadas como alvos potenciais de modificação para a síntese e estudos da relação entre a estrutura e atividade (SAR).

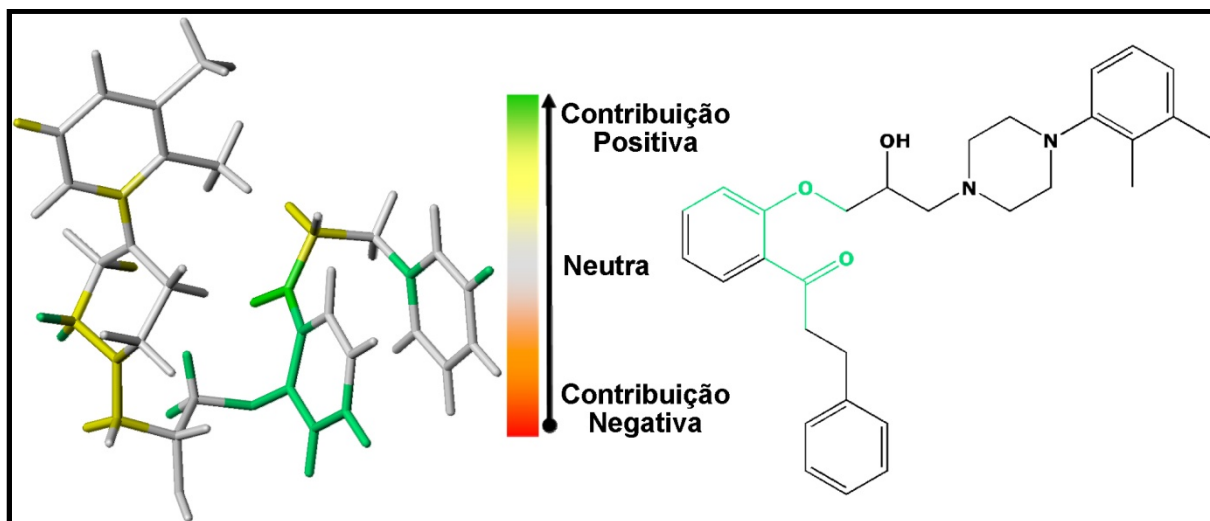


Figura 4.13 – Mapa de contribuição de HQSAR para a atividade inibitória sobre a P-gp do derivado de propafenona mais potente da serie estudada.

4.5 Solubilidade Aquosa

4.5.1 Organização do Conjunto de Dados

Como descrito anteriormente, foram coletados 1311 compostos para o trabalho de modelagem da solubilidade aquosa¹²⁰. Por questão de conveniência, a lista com as nomenclaturas CAS (do inglês, *Chemical Abstracts Service*) das moléculas e os valores de solubilidade expressos em logS é apresentada na Tabela A-1 da sessão anexo desta tese de doutorado.

O grande número moléculas que constitui o conjunto de dados confere alta diversidade química e notável cobertura do espaço químico favorável e desfavorável para a solubilidade. A análise da distribuição dos dados de solubilidade (Figura 4.14) permite observar que 40,96% dos compostos estão na faixa de baixa solubilidade ($\log S \leq -3$). A faixa de solubilidade moderada conta com a percentual de 39,97 do conjunto de dados, enquanto que a faixa de alta solubilidade conta com 19,07% dos compostos.

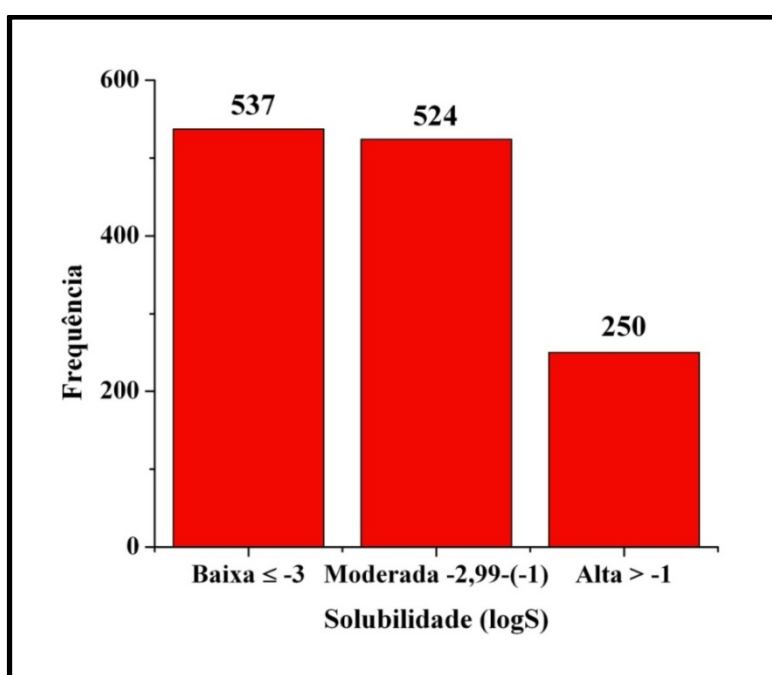


Figura 4.14 – Histograma da distribuição dos valores de solubilidade do conjunto de dados.

4.5.2 Desenvolvimento de Modelos de HQSAR-SOL

A solubilidade em meio aquoso apresenta importância vital no planejamento de moléculas bioativas, influenciando em várias etapas que vão desde a avaliação biológica em modelos experimentais em *in vitro* e *in vivo* até os processos de desintegração e solubilização de princípios ativos. Também tem papel fundamental no processo de absorção intestinal. Por essas razões, a solubilidade aquosa é uma propriedade físico-química atrativa para a geração de modelos de QSAR.

Para o processo de modelagem, o conjunto padrão de dados de 1311 compostos foi dividido em conjuntos treinamento (1053 compostos, 81%) e teste (258 compostos, 19%), de acordo com o histograma apresentado na Figura 4.15.

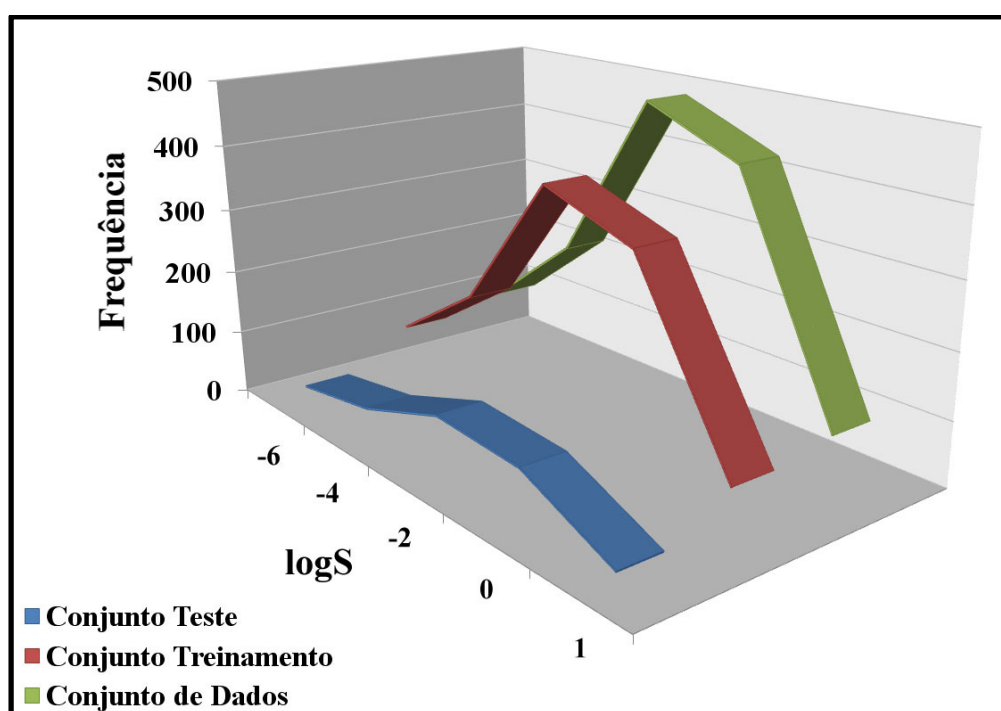


Figura 4.15 – Representação gráfica da distribuição dos valores de logS.

O conjunto treinamento foi submetido à geração de vários modelos através da variação da distinção de fragmentos, mantendo-se constante o tamanho dos fragmentos no padrão de fragmento 4-7. Os resultados estatísticos obtidos a partir da análise PLS são mostrados na Tabela 4.17. O melhor resultado estatístico foi obtido usando as combinações A/B/H, modelo 3 ($q^2 = 0,76$ e $r^2 = 0,79$); A/B/C/H, modelo 5 ($q^2 = 0,74$ e

$r^2 = 0,78$); A/B/H/DA, modelo **8** ($q^2 = 0,74$ e $r^2 = 0,79$); e A/B/C/H/Ch, modelo **10** ($q^2 = 0,73$ e $r^2 = 0,79$).

Tabela 4.17 – Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão do fragmento (4-7)

Modelo	Distinção do Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B	0,71	0,75	1,00	8	401
2	A/B/C	0,70	0,77	0,97	8	401
3	A/B/H	0,76	0,79	0,93	8	401
4	A/B/DA	0,68	0,74	1,02	8	353
5	A/B/C/H	0,74	0,78	0,94	8	401
6	A/B/C/Ch	0,70	0,77	0,97	8	401
7	A/B/C/DA	0,66	0,78	0,95	8	401
8	A/B/H/DA	0,74	0,79	0,93	8	401
9	A/B/Ch/DA	0,71	0,75	1,01	8	401
10	A/B/C/H/Ch	0,73	0,79	0,94	8	401
11	A/B/C/Ch/DA	0,64	0,73	1,04	7	353
12	A/B/C/H/Ch/DA	0,68	0,78	0,96	8	401

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma. Distinção do fragmento: A, átomos; B, ligação; C, conexão; H, átomos de hidrogênio; Ch, quiralidade; DA, doador e receptor.

Os modelos **3**, **5**, **8** e **10** foram selecionados para a investigação da influência da variação do tamanho dos fragmentos sobre os parâmetros estatísticos. Para isso, foram realizadas análises individuais empregando diferentes combinações de tamanho dos fragmentos: 2-5, 3-6, 4-7, 5-8, 6-9, de acordo com a Tabela 4.18.

Tabela 4.18 – Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando as melhores distinções observadas A/B/H, A/B/C/H, A/B/H/DA e A/B/C/H/Ch

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
Modelo 3					
2-5	0,82	0,85	0,79	8	353
3-6	0,79	0,82	0,87	8	307
4-7	0,76	0,79	0,93	8	401
5-8	0,69	0,74	1,03	8	401
6-9	0,63	0,71	1,10	8	307
Modelo 5					
2-5	0,80	0,83	0,83	8	401
3-6	0,78	0,81	0,88	8	401
4-7	0,74	0,78	0,94	8	401
5-8	0,67	0,73	1,04	8	353
6-9	0,59	0,68	1,14	8	401
Modelo 8					
2-5	0,81	0,85	0,79	8	401
3-6	0,79	0,83	0,84	8	401
4-7	0,74	0,79	0,93	8	401
5-8	0,66	0,74	1,03	8	401
6-9	0,58	0,68	1,114	8	353
Modelo 10					
2-5	0,80	0,83	0,84	8	401
3-6	0,78	0,81	0,88	8	401
4-7	0,73	0,79	0,94	8	401
5-8	0,67	0,73	1,05	8	353
6-9	0,57	0,70	1,10	8	353

q^2 , coeficiente de correlação com validação cruzada; r^2 , coeficiente de correlação sem validação cruzada; SEE, erro padrão sem validação cruzada; N , número ótimo de componentes; HL, comprimento do holograma.

Os melhores modelos foram gerados com tamanho dos fragmentos 2-5 para todas as distinções selecionadas. Esses modelos, destacados em cinza na Tabela 4.18, possuem parâmetros estatísticos superiores aos gerados com o tamanho padrão do

fragmento (4-7) (Tabela 4.17). O procedimento LMO, com até 15 grupos, confirmou a consistência interna dos modelos.

4.5.3 Validação dos Modelos de HQSAR-SOL

Os modelos finais de HQSAR foram avaliados quanto à sua capacidade preditiva através de um processo de validação externa empregando um conjunto teste com 258 moléculas que não foram incluídas no treinamento para geração dos modelos. Os resultados do processo de validação externa são apresentados na Tabela 4.19.

Tabela 4.19 – Valores experimentais e preditos de solubilidade associados aos resíduos para os compostos do conjunto teste

No.	Composto (CAS)	Exp ^[a]	Pred 1 ^[b]	Pred 2 ^[c]	Pred 3 ^[d]	Pred 4 ^[e]	Cons ^[f]	Res ^[g]
Conjunto Teste								
001	100-41-4	-2,77	-2,19	-2,01	-2,61	-2,08	-2,22	-0,55
002	100-42-5	-2,82	-2,08	-2,26	-2,22	-2,18	-2,18	-0,64
003	100-46-9	-1,53	-1,54	-1,91	-1,31	-1,92	-1,67	0,14
004	100-52-7	-1,19	-1,68	-1,93	-1,95	-1,88	-1,86	0,67
005	100-68-5	-2,39	-1,72	-1,77	-2,09	-1,71	-1,82	-0,57
006	1024-57-3	-6,29	-7,56	-6,77	-6,82	-6,65	-6,95	0,66
007	10311-84-9	-6,34	-4,54	-4,86	-4,51	-5,04	-4,74	-1,60
008	104-76-7	-2,11	-2,68	-2,73	-2,44	-2,66	-2,63	0,52
009	105-55-5	-1,46	-1,97	-1,81	-0,66	-2,02	-1,61	0,15
010	106-51-4	-0,99	-0,95	-1,03	-0,81	-0,91	-0,93	-0,06
011	106-95-6	-1,50	-1,20	-1,15	-1,00	-1,15	-1,13	-0,37
012	107-04-0	-1,32	-1,24	-1,00	-1,07	-0,97	-1,07	-0,25
013	1077-16-3	-5,21	-4,08	-3,84	-4,48	-3,92	-4,08	-1,13
014	108-70-3	-4,48	-3,98	-3,75	-4,03	-3,63	-3,85	-0,63
015	108-95-2	0,00	-1,29	-1,29	-1,43	-1,28	-1,32	1,32
016	109-00-2	-0,46	-0,56	-0,38	-0,65	-0,38	-0,49	0,03
017	109-67-1	-2,68	-1,91	-1,98	-2,09	-2,05	-2,01	-0,67
018	109-87-5	0,48	0,02	0,34	0,56	0,46	0,35	0,13
019	109-97-7	-0,17	-0,60	-0,49	-0,49	-0,44	-0,51	0,34
020	110-74-7	-0,49	-0,61	-1,05	-0,80	-1,08	-0,89	0,40
021	110-83-8	-2,59	-2,51	-2,52	-2,76	-2,60	-2,60	0,01
022	110-91-8	1,06	0,10	0,85	0,23	0,85	0,51	0,55
023	110-94-1	1,00	0,15	0,25	-0,08	0,20	0,13	0,87
024	111-26-2	-1,10	-1,82	-2,15	-1,81	-2,18	-1,99	0,89
025	1114-71-2	-3,35	-3,40	-3,26	-3,99	-3,46	-3,53	0,18
026	112-30-1	-3,63	-3,77	-3,72	-3,60	-3,72	-3,70	0,07
027	112-37-8	-3,55	-3,80	-3,76	-4,10	-3,77	-3,86	0,31
028	112-53-8	-4,67	-4,69	-4,63	-4,53	-4,61	-4,61	-0,06
029	117-81-7	-6,96	-7,11	-8,06	-7,51	-7,90	-7,65	0,69
030	118-10-5	-3,09	-3,15	-3,37	-2,78	-3,29	-3,14	0,05
031	1185-33-7	-1,04	-1,07	-2,01	-1,04	-1,85	-1,49	0,45
032	121-34-6	-2,05	-1,54	-1,80	-1,39	-1,86	-1,65	-0,40

Continua

Continuação

033	121-75-5	-3,36	-1,86	-2,23	-1,58	-2,69	-2,09	-1,27
034	122-34-9	-4,55	-2,77	-2,35	-3,04	-2,46	-2,66	-1,89
035	122-59-8	-1,10	-1,68	-1,11	-1,74	-1,14	-1,42	0,32
036	122-88-3	-2,29	-2,54	-1,99	-2,55	-2,00	-2,27	-0,02
037	122-99-6	-0,71	-1,55	-1,19	-1,37	-1,29	-1,35	0,64
038	123-19-3	-1,30	-1,64	-1,50	-2,06	-1,55	-1,68	0,38
039	123-31-9	-0,17	-1,07	-1,30	-1,13	-1,32	-1,20	1,03
040	123-56-8	0,30	0,13	0,02	0,48	-0,03	0,15	0,15
041	123-75-1	1,15	-0,19	0,02	0,51	0,02	0,09	1,06
042	123-86-4	-1,24	-1,14	-1,38	-1,16	-1,47	-1,29	0,05
043	124-07-2	-2,30	-2,43	-2,40	-2,70	-2,43	-2,49	0,19
044	125-33-7	-2,64	-2,32	-1,96	-2,18	-2,07	-2,13	-0,51
045	127-18-4	-2,54	-0,82	-0,52	-0,79	-0,44	-0,64	-1,90
046	127-31-1	-3,43	-3,89	-3,16	-3,68	-2,78	-3,37	-0,06
047	129-20-4	-3,73	-4,15	-4,49	-4,70	-4,57	-4,48	0,75
048	13110-37-7	-3,26	-2,91	-3,69	-3,10	-3,65	-3,34	0,08
049	133-90-4	-2,47	-2,84	-3,34	-2,75	-3,26	-3,05	0,58
050	13476-55-6	-3,95	-3,37	-4,14	-3,57	-4,09	-3,79	-0,16
051	136-60-7	-3,48	-2,74	-3,09	-2,93	-3,08	-2,96	-0,52
052	138-52-3	-0,85	-0,94	-0,91	-1,64	-0,77	-1,06	0,21
053	140-67-0	-2,92	-2,84	-2,77	-2,69	-2,83	-2,78	-0,14
054	141-78-6	-0,04	-0,27	-0,36	-0,14	-0,50	-0,31	0,27
055	141-84-4	-1,77	-0,89	-0,85	-0,51	-0,76	-0,75	-1,02
056	14214-32-5	-4,16	-4,14	-3,49	-3,36	-3,58	-3,64	-0,52
057	142-68-7	-0,03	-1,18	-0,95	-1,23	-0,87	-1,06	1,03
058	142-96-1	-1,85	-2,34	-2,20	-2,53	-2,23	-2,33	0,48
059	143-07-7	-4,62	-4,26	-4,22	-4,57	-4,22	-4,31	-0,31
060	14309-40-1	-4,60	-3,83	-4,60	-4,04	-4,54	-4,25	-0,35
061	145-13-1	-4,65	-4,93	-5,27	-4,53	-5,27	-5,00	0,35
062	145-73-3	-0,27	-0,33	-0,05	-0,53	0,12	-0,20	-0,07
063	151-67-7	-1,71	-0,78	-0,84	-0,47	-0,85	-0,73	-0,98
064	15271-41-7	-2,08	-3,48	-3,08	-2,49	-2,66	-2,93	0,85
065	15414-82-1	-0,06	-1,10	-0,98	-1,00	-0,95	-1,01	0,95
066	15687-27-1	-3,99	-3,06	-3,05	-3,64	-3,12	-3,22	-0,77
067	15862-07-4	-6,27	-6,10	-6,09	-6,25	-6,01	-6,11	-0,16
068	1678-91-7	-4,25	-3,60	-3,43	-3,70	-3,36	-3,52	-0,73
069	1698-60-8	-2,87	-2,09	-2,99	-2,79	-2,92	-2,70	-0,17
070	1746-81-2	-2,57	-3,37	-3,11	-2,84	-3,07	-3,10	0,53
071	17560-51-9	-3,78	-5,86	-5,64	-5,18	-5,75	-5,61	1,83
072	1861-40-1	-5,53	-5,74	-5,09	-5,38	-5,28	-5,37	-0,16
073	1912-24-9	-3,55	-3,17	-3,15	-3,61	-3,29	-3,30	-0,25
074	192-97-2	-7,80	-9,05	-8,05	-8,09	-7,83	-8,25	0,45
075	1929-82-4	-3,76	-1,92	-1,32	-1,90	-1,14	-1,57	-2,19
076	1982-49-6	-4,11	-3,97	-3,94	-4,19	-3,81	-3,98	-0,13
077	1982-55-4	-4,98	-3,65	-2,36	-3,57	-2,26	-2,96	-2,02
078	204443-27-6	-0,91	-1,47	-1,27	-0,99	-1,33	-1,27	0,36
079	204443-30-1	-1,55	-1,43	-1,43	-1,25	-1,49	-1,40	-0,15
080	2050-67-1	-5,80	-5,51	-5,43	-5,66	-5,32	-5,48	-0,32
081	2051-24-3	-11,62	-10,69	-10,56	-10,33	-10,52	-10,53	-1,09
082	206-44-0	-6,00	-6,45	-6,22	-6,96	-6,17	-6,45	0,45
083	2164-17-2	-3,43	-2,88	-2,55	-2,79	-2,58	-2,70	-0,73
084	217-59-4	-6,74	-7,44	-6,99	-6,85	-6,96	-7,06	0,32
085	218-01-9	-8,06	-7,42	-7,35	-6,95	-7,40	-7,28	-0,78
086	22862-76-6	-1,61	-1,37	-1,02	-0,75	-0,99	-1,03	-0,58
087	229-87-8	-2,78	-4,20	-4,47	-3,89	-4,44	-4,25	1,47
088	23103-98-2	-1,95	-2,11	-1,12	-2,11	-1,26	-1,65	-0,30
089	2363-58-8	-5,41	-5,24	-5,02	-5,38	-5,22	-5,22	-0,19
090	2432-99-7	-2,70	-2,90	-3,15	-2,91	-3,14	-3,03	0,33
091	2437-79-8	-6,51	-7,01	-7,00	-7,11	-6,89	-7,00	0,49
092	24691-80-3	-3,30	-2,68	-2,50	-2,76	-2,48	-2,60	-0,70

Continua

Continuação

093	2536-31-4	-4,18	-4,90	-5,19	-5,15	-4,90	-5,04	0,86
094	271-44-3	-2,16	-2,08	-2,00	-3,47	-1,97	-2,38	0,22
095	27314-13-2	-4,04	-2,81	-3,82	-2,75	-3,88	-3,31	-0,73
096	2921-88-2	-5,67	-4,59	-4,05	-4,62	-4,28	-4,38	-1,29
097	2939-80-2	-5,40	-2,42	-2,75	-2,46	-2,80	-2,61	-2,79
098	298-00-0	-3,68	-4,08	-3,96	-3,55	-3,99	-3,89	0,21
099	3124-46-7	-0,80	-0,76	-2,44	-1,15	-2,19	-1,64	0,84
100	31251-03-3	-8,40	-6,80	-5,91	-6,33	-6,05	-6,27	-2,13
101	31431-39-7	-3,88	-4,56	-4,48	-4,29	-4,28	-4,40	0,52
102	327-54-8	-2,38	-2,23	-2,33	-1,78	-2,27	-2,15	-0,23
103	330-54-1	-3,76	-3,53	-3,40	-3,92	-3,36	-3,55	-0,21
104	33284-50-3	-5,25	-5,37	-5,39	-5,65	-5,31	-5,43	0,18
105	333-41-5	-3,76	-3,89	-3,92	-3,95	-4,38	-4,03	0,27
106	3380-34-5	-4,46	-5,45	-4,94	-5,10	-4,96	-5,11	0,65
107	352-97-6	-1,51	-0,30	-0,54	-0,91	-0,54	-0,57	-0,94
108	35693-92-6	-6,14	-6,10	-6,07	-6,34	-5,98	-6,12	-0,02
109	35694-08-7	-9,16	-9,54	-9,44	-9,23	-9,38	-9,40	0,24
110	357-57-3	-2,09	-4,01	-4,17	-3,50	-4,16	-3,96	1,87
111	3625-25-0	-2,64	-4,52	-4,45	-3,98	-4,36	-4,33	1,69
112	363-24-6	-2,47	-3,74	-4,34	-3,41	-4,61	-4,03	1,56
113	3766-60-7	-3,90	-3,36	-3,08	-3,26	-3,05	-3,19	-0,71
114	37680-73-2	-7,89	-7,81	-7,70	-7,68	-7,59	-7,69	-0,20
115	3772-76-7	-2,54	-2,53	-2,67	-2,50	-2,78	-2,62	0,08
116	4248-19-5	0,10	-0,67	-1,64	-0,32	-1,64	-1,07	1,17
117	4248-20-8	-2,62	-2,58	-2,82	-2,54	-2,79	-2,68	0,06
118	4360-12-7	-2,82	-3,24	-2,15	-2,86	-2,20	-2,61	-0,21
119	439-14-5	-3,76	-4,94	-4,62	-5,03	-4,56	-4,79	1,03
120	4396-01-4	-0,45	-0,36	-0,25	0,06	-0,01	-0,14	-0,31
121	461-58-5	-0,31	-0,57	-0,43	-1,22	-0,37	-0,65	0,34
122	461-72-3	-0,40	-0,75	-0,51	-0,66	-0,50	-0,61	0,21
123	464-07-3	-0,62	-0,59	-1,43	-0,62	-1,36	-1,00	0,38
124	50-03-3	-4,30	-3,57	-3,35	-2,83	-3,74	-3,37	-0,93
125	50-06-6	-2,34	-2,84	-2,14	-2,40	-2,19	-2,39	0,05
126	50-11-3	-2,23	-1,32	-1,14	-1,32	-1,27	-1,26	-0,97
127	50-29-3	-8,08	-6,56	-6,94	-6,54	-6,82	-6,72	-1,36
128	50-53-3	-5,01	-4,53	-5,38	-4,73	-5,03	-4,92	-0,09
129	50-70-4	1,09	0,87	0,90	0,90	0,89	0,89	0,20
130	50-78-2	-1,61	-1,41	-1,56	-1,33	-1,63	-1,48	-0,13
131	513-36-0	-2,00	-2,06	-2,42	-2,18	-2,27	-2,23	0,23
132	514-10-3	-3,80	-5,54	-4,81	-6,05	-4,96	-5,34	1,54
133	514-17-0	-3,59	-3,73	-3,12	-2,85	-3,13	-3,21	-0,38
134	51481-61-9	-1,35	-2,80	-2,18	-1,34	-2,34	-2,16	0,81
135	51-67-2	-1,12	-1,29	-1,90	-1,12	-1,94	-1,56	0,44
136	519-05-1	-1,92	-2,14	-2,36	-1,85	-2,47	-2,20	0,28
137	51953-17-4	0,59	-0,83	-0,97	-0,98	-0,89	-0,92	1,51
138	52-21-1	-4,37	-3,72	-3,72	-2,96	-4,04	-3,61	-0,76
139	52-43-7	-2,06	-1,72	-1,94	-1,77	-1,84	-1,82	-0,24
140	52663-69-1	-7,92	-9,05	-8,95	-8,86	-8,86	-8,93	1,01
141	526-75-0	-1,43	-2,18	-1,75	-2,15	-1,90	-1,99	0,56
142	52712-05-7	-8,94	-8,92	-8,78	-8,69	-8,71	-8,77	-0,17
143	529-86-2	-4,73	-5,03	-5,17	-4,36	-5,14	-4,92	0,19
144	53808-87-0	-2,10	-4,06	-2,50	-3,11	-2,70	-3,09	0,99
145	539-86-6	-0,83	-2,27	-2,30	-1,47	-2,27	-2,08	1,25
146	54-31-9	-3,66	-3,37	-3,92	-2,65	-3,91	-3,46	-0,20
147	55-21-0	-0,96	-1,68	-1,64	-1,60	-1,59	-1,63	0,67
148	55215-17-3	-7,43	-7,49	-7,43	-7,50	-7,36	-7,45	0,02
149	552-63-6	-0,93	-1,15	-0,92	-1,05	-1,13	-1,06	0,13
150	55312-69-1	-7,21	-7,45	-7,40	-7,40	-7,34	-7,40	0,19
151	554-12-1	-0,14	-0,27	-0,16	-0,16	-0,21	-0,20	0,06
152	55-63-0	-2,22	-1,13	-0,86	-1,68	-0,95	-1,15	-1,07

Continua

Continuação

153	556-88-7	-1,37	-0,77	-0,79	-1,47	-0,73	-0,94	-0,43
154	5598-13-0	-4,82	-4,40	-4,08	-3,91	-4,08	-4,12	-0,70
155	5635-50-7	-4,43	-4,57	-4,46	-4,88	-4,70	-4,66	0,23
156	56-38-2	-4,29	-4,27	-3,93	-4,26	-4,18	-4,16	-0,13
157	56-81-5	1,12	0,58	0,38	0,81	0,38	0,54	0,58
158	569-31-3	-1,89	-2,28	-2,05	-1,64	-2,19	-2,04	0,15
159	57-41-0	-3,99	-4,71	-3,47	-3,88	-3,44	-3,87	-0,12
160	57-44-3	-1,39	-1,38	-1,27	-1,49	-1,27	-1,35	-0,04
161	57-74-9	-6,86	-8,71	-8,45	-8,42	-8,02	-8,40	1,54
162	57837-19-1	-1,60	-3,32	-2,07	-1,91	-2,25	-2,39	0,79
163	581-40-8	-4,72	-4,44	-3,89	-4,21	-4,08	-4,15	-0,57
164	581-42-0	-4,89	-4,36	-3,87	-4,18	-4,04	-4,11	-0,78
165	5819-08-9	-1,16	-2,21	-2,39	-1,69	-2,24	-2,13	0,97
166	58-22-0	-4,07	-4,67	-4,44	-4,20	-4,43	-4,44	0,37
167	58-55-9	-1,37	-1,93	-0,96	-1,26	-1,13	-1,32	-0,05
168	58-89-9	-4,60	-5,40	-5,87	-5,20	-5,41	-5,47	0,87
169	590-27-2	-3,22	-0,49	-0,52	-0,40	-0,48	-0,47	-2,75
170	590-36-3	-0,49	-1,07	-1,37	-0,37	-1,38	-1,05	0,56
171	592-35-8	-0,66	-1,21	-1,45	-1,15	-1,45	-1,31	0,65
172	59-31-4	-2,14	-2,03	-2,72	-2,60	-2,60	-2,48	0,34
173	59-46-1	-1,40	-2,07	-2,11	-1,73	-2,06	-1,99	0,59
174	595-33-5	-5,35	-4,44	-4,62	-4,93	-5,05	-4,76	-0,59
175	59-66-5	-2,36	-0,80	-0,96	-0,93	-0,96	-0,91	-1,45
176	600-36-2	-1,22	-1,49	-2,11	-1,40	-2,01	-1,75	0,53
177	60168-88-9	-4,38	-5,15	-5,43	-5,80	-5,13	-5,38	1,00
178	6032-29-7	-0,29	-1,10	-1,03	-0,71	-1,09	-0,98	0,69
179	60-35-5	1,58	0,06	-0,10	0,08	-0,07	-0,01	1,59
180	60-80-0	0,39	-2,56	-2,25	-2,50	-2,33	-2,41	2,80
181	608-21-9	-5,04	-2,59	-1,84	-1,99	-1,67	-2,02	-3,02
182	60-87-7	-4,26	-3,62	-4,28	-3,49	-3,75	-3,78	-0,48
183	613-13-8	-5,17	-5,21	-5,17	-5,00	-5,22	-5,15	-0,02
184	614-29-9	-2,52	-2,82	-2,62	-2,85	-2,83	-2,78	0,26
185	615-54-3	-4,50	-2,72	-1,96	-1,93	-1,76	-2,09	-2,41
186	615-77-0	-0,80	-1,26	-0,98	-0,82	-1,05	-1,03	0,23
187	618-51-9	-3,27	-1,68	-2,18	-1,94	-2,09	-1,97	-1,30
188	619-45-4	-1,59	-1,35	-2,06	-1,30	-1,96	-1,67	0,08
189	62-23-7	-2,80	-1,84	-2,41	-2,19	-2,35	-2,20	-0,61
190	622-45-7	-1,67	-2,02	-2,17	-1,94	-2,16	-2,07	0,40
191	625-23-0	-1,08	-1,52	-1,83	-0,84	-1,83	-1,50	0,42
192	625-29-6	-2,63	-2,66	-2,76	-2,93	-2,71	-2,77	0,14
193	628-41-1	-1,97	-2,08	-2,21	-2,35	-2,28	-2,23	0,26
194	629-05-0	-3,66	-2,56	-2,68	-2,83	-2,64	-2,68	-0,98
195	632-22-4	0,94	-0,63	-0,56	0,12	-0,57	-0,41	1,35
196	63-68-3	-0,42	-0,57	-0,96	-0,98	-0,97	-0,87	0,45
197	63-75-2	0,81	-0,50	-1,07	-0,74	-0,93	-0,81	1,62
198	638-42-6	-1,47	-1,67	-1,91	-1,61	-1,90	-1,77	0,30
199	64-77-7	-3,39	-3,41	-3,38	-3,32	-3,56	-3,42	0,03
200	65-71-4	-1,52	-0,89	-1,08	-1,19	-1,08	-1,06	-0,46
201	65-86-1	-1,93	-1,35	-0,88	-1,14	-0,76	-1,03	-0,90
202	664-95-9	-4,21	-4,25	-4,25	-4,41	-4,36	-4,32	0,11
203	67-20-9	-3,38	-2,12	-1,36	-1,76	-1,45	-1,67	-1,71
204	68-35-9	-3,51	-2,00	-1,94	-2,21	-1,91	-2,02	-1,49
205	693-54-9	-3,31	-3,00	-2,99	-3,44	-2,93	-3,09	-0,22
206	694-85-9	0,96	-1,12	-1,29	-0,92	-1,44	-1,19	2,15
207	6988-21-2	-1,57	-1,47	-1,27	-1,23	-1,36	-1,33	-0,24
208	70-47-3	-0,74	0,77	0,16	0,42	-0,04	0,33	-1,07
209	70-69-9	-2,63	-1,62	-1,92	-2,11	-1,98	-1,90	-0,73
210	71-43-2	-1,64	-1,51	-1,28	-1,73	-1,24	-1,44	-0,20
211	72-19-5	-0,09	0,56	0,29	0,03	0,11	0,25	-0,34
212	7286-84-2	-3,26	-3,05	-3,49	-2,60	-3,38	-3,13	-0,13

Continua

Continuação

213	7294-05-5	-1,27	-0,75	-1,30	-0,85	-1,19	-1,02	-0,25
214	738-70-5	-2,86	-4,24	-2,78	-3,50	-3,00	-3,38	0,52
215	751-97-3	-1,42	-2,56	-3,33	-4,28	-3,25	-3,35	1,93
216	75-25-2	-1,91	-0,47	-0,49	-0,37	-0,46	-0,45	-1,46
217	75-56-9	-0,59	-0,38	-0,23	-0,22	-0,21	-0,26	-0,33
218	75-85-4	0,08	-0,50	-0,75	0,21	-0,76	-0,45	0,53
219	76-01-7	-2,60	-1,47	-1,59	-1,43	-1,45	-1,49	-1,11
220	77-06-5	-1,84	-2,49	-3,68	-2,94	-3,90	-3,26	1,42
221	77-09-8	-2,90	-4,34	-5,84	-4,96	-5,92	-5,26	2,36
222	77-21-4	-2,34	-2,18	-2,40	-2,28	-2,40	-2,31	-0,03
223	77-74-7	-0,38	-0,85	-0,83	-0,15	-0,84	-0,67	0,29
224	77-92-9	0,51	1,02	1,49	0,90	1,52	1,23	-0,72
225	79-00-5	-1,48	-1,27	-1,36	-1,33	-1,22	-1,29	-0,19
226	79-34-5	-1,74	-1,32	-1,48	-1,30	-1,33	-1,36	-0,38
227	80-35-3	-3,28	-2,13	-2,62	-2,97	-2,64	-2,59	-0,69
228	81-25-4	-3,37	-4,60	-3,98	-3,17	-4,04	-3,95	0,57
229	83-34-1	-2,42	-2,88	-2,68	-2,01	-2,63	-2,55	0,13
230	83-43-2	-2,99	-4,10	-4,02	-3,42	-4,40	-3,98	0,99
231	83-49-8	-3,82	-5,42	-4,74	-4,43	-4,82	-4,85	1,03
232	846-49-1	-3,60	-5,64	-5,71	-6,45	-5,65	-5,87	2,27
233	85-44-9	-1,39	-1,77	-2,38	-1,84	-2,38	-2,09	0,70
234	85-68-7	-5,64	-4,23	-4,91	-4,38	-4,92	-4,61	-1,03
235	87-20-7	-3,16	-2,79	-3,46	-2,95	-3,43	-3,16	0,00
236	872-55-9	-2,59	-1,56	-1,08	-1,52	-1,18	-1,33	-1,26
237	87-61-6	-4,00	-3,75	-3,60	-3,76	-3,56	-3,67	-0,33
238	87-72-9	0,39	0,42	0,32	-0,30	0,42	0,21	0,18
239	88-13-1	-1,47	-0,69	-0,71	-0,81	-0,63	-0,71	-0,76
240	88-67-5	-2,73	-1,59	-2,02	-1,91	-1,97	-1,87	-0,86
241	90-02-8	-0,86	-1,41	-1,77	-1,58	-1,77	-1,63	0,77
242	90-11-9	-4,35	-4,04	-3,72	-3,62	-3,65	-3,76	-0,59
243	90-12-0	-3,70	-3,85	-3,94	-3,87	-3,95	-3,90	0,20
244	91-58-7	-4,14	-4,38	-4,12	-4,36	-4,13	-4,25	0,11
245	91-94-1	-4,92	-4,92	-5,49	-4,83	-5,39	-5,16	0,24
246	93-76-5	-2,96	-3,98	-3,40	-3,75	-3,40	-3,63	0,67
247	93-89-0	-2,32	-1,86	-2,06	-1,91	-2,11	-1,99	-0,33
248	94-09-7	-2,32	-1,58	-2,21	-1,62	-2,23	-1,91	-0,41
249	94-26-8	-2,72	-2,52	-3,10	-2,63	-3,12	-2,84	0,12
250	95-49-8	-3,52	-2,70	-2,42	-2,77	-2,43	-2,58	-0,94
251	95-95-4	-2,21	-3,57	-3,56	-3,41	-3,53	-3,52	1,31
252	96-14-0	-3,68	-2,45	-2,42	-2,71	-2,44	-2,50	-1,18
253	96-33-3	-0,22	-0,63	-0,50	-0,31	-0,40	-0,46	0,24
254	96-88-8	-1,55	-2,76	-2,85	-2,90	-2,67	-2,80	1,25
255	97-53-0	-1,56	-2,56	-2,53	-2,32	-2,63	-2,51	0,95
256	98-86-2	-1,28	-1,41	-1,70	-1,97	-1,68	-1,69	0,41
257	99-09-2	-2,19	-1,63	-2,04	-1,92	-1,97	-1,89	-0,30
258	99-24-1	-1,24	-0,91	-1,52	-0,70	-1,55	-1,17	-0,07
	$r_{pred}^{2[h]}$		0,82	0,80	0,83	0,80	0,83	

[a] Valor experimental de logS. [b] Valor predito pelo modelo 3. [c] Valor predito pelo modelo 5. [d] Valor predito pelo modelo 8. [e] Valor predito pelo modelo 10. [f] Predição consenso. [g] Resíduo: diferença entre o valor experimental e a predição consenso. [h] coeficiente de correlação entre o valor experimental e predito do conjunto teste

Os valores preditos para o conjunto teste através do método consenso se aproximam dos valores experimentais, desviando-se em média não mais do que 1 unidade logarítmica. Exceções são os compostos **7, 15, 34, 45, 111, 215 e 232**, que apresentam erro entre 1,32 e 2,27 unidades logarítmicas. O gráfico das relações entre

os valores experimentais e preditos de logS dos compostos dos conjuntos treinamento e teste é apresentado na Figura 4.16.

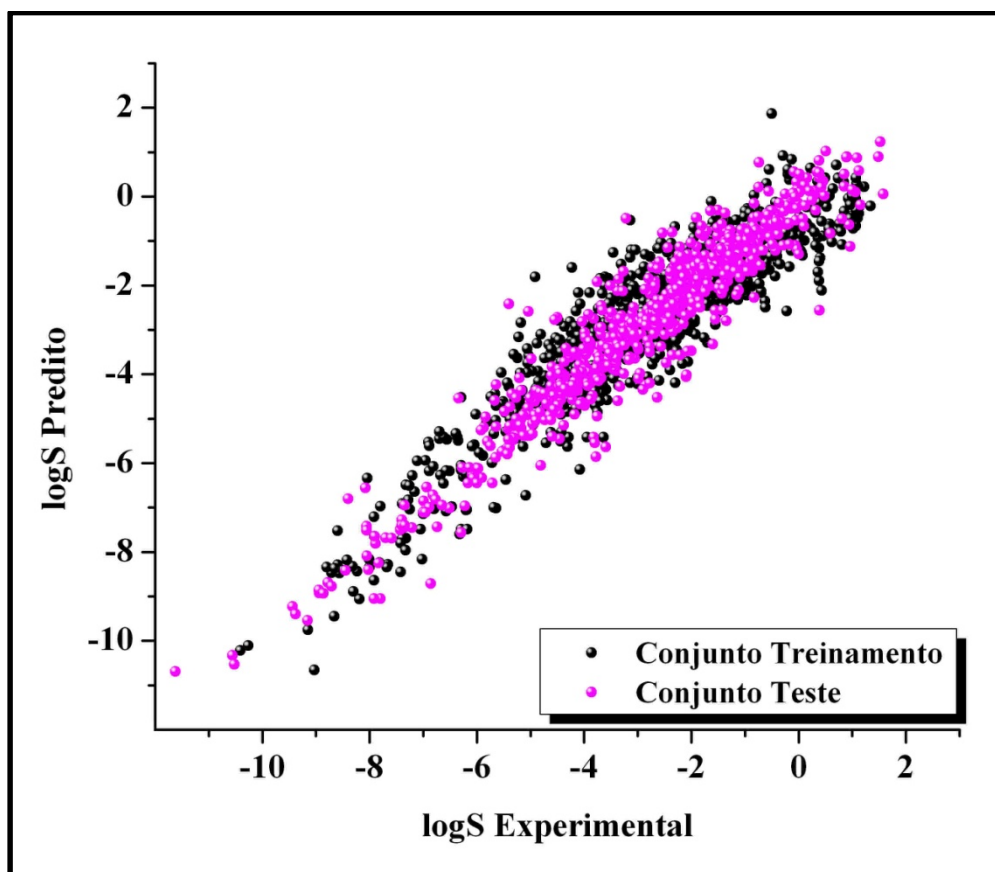


Figura 4.16 – Gráfico de HQSAR dos valores experimentais e preditos de solubilidade aquosa para os 1311 compostos do conjunto de dados.

Pode-se observar a boa concordância entre os valores preditos e experimentais, indicando que os modelos finais de HQSAR integrados pelo método consenso apresentam boa consistência interna e externa, com interessante capacidade preditiva para novos compostos orgânicos.

CAPÍTULO **5**

Base de Datos
PK/DB

5 PK/DB: Base de Dados para Propriedades Farmacocinéticas

5.1 Apresentação

Embora os estudos de propriedades farmacocinéticas sejam essenciais em P&D de fármacos, os pesquisadores que se dedicam a essas atividades, tanto na academia quanto na indústria farmacêutica, encontram grande dificuldade, pois há pouca disponibilidade de dados padronizados. Com o objetivo de disponibilizar informações farmacocinéticas à comunidade científica brasileira e mundial, a primeira base de dados brasileira para o estudo de propriedades farmacocinéticas, batizada de PK/DB (*Database for Pharmacokinetic Properties*), foi lançada por nosso grupo no Laboratório de Química Medicinal e Computacional (LQMC) sendo o desenvolvimento desta base parte do trabalho de doutorado apresentado nesta tese. A base PK/DB (<http://www.pkdb.ifsc.usp.br>) incorpora dados de alta qualidade de propriedades farmacocinéticas e físico-químicas de fármacos pertencendo a várias classes terapêuticas distintas, e compostos orgânicos ordinários com elevada diversidade química¹²⁶. O lançamento da base de dados foi um evento notável no Brasil, ganhando destaque na mídia nacional como pode ser observado na matéria do boletim FAPESP de 10/06/2008, entre outros meios de comunicação¹²⁷⁻¹²⁹.

5.2 Caracterização

Os dados de propriedades farmacocinéticas e físico-químicas foram coletados da literatura e de base de dados públicas, resultando em mais de 1,3 mil fármacos e 4,1 mil valores de propriedades como ilustrado na Tabela 5.1. Alguns dados são descritos na sessão **3.1** do capítulo **3**. Os compostos da base são identificados por sua estrutura,

fórmula molecular, peso molecular, representação SMILES, ação farmacológica e propriedades farmacocinéticas.

Tabela 5.1 – Número de medidas presentes na PK/DB para várias propriedades farmacocinéticas

Propriedade	Sigla	Medidas
Absorção Intestinal Humana	HIA	687
Biodisponibilidade Oral Humana	F	660
Ligação às Proteínas Plasmáticas	PPB	726
Permeabilidade da Barreira Hematoencefálica	BBB	200
Volume de Distribuição	V _d	738
Eliminação Renal	Cl	360
Tempo de Meia Vida	T _{1/2}	770
CYP-Metabolismo	CYP	295

5.3 Estrutura e Aplicações

PK/DB permite a busca de compostos por estrutura química, subestrutura, nome e fórmula molecular, bem como a busca por valor da propriedade farmacocinética específica ou por faixas de valores de propriedades.

A informação sobre a isoforma CYP responsável pelo metabolismo de diversos fármacos também é disponível. Os usuários podem ainda empregar uma combinação de diversos critérios (<http://www.pkdb.ifsc.usp.br/manual.pdf>) na busca de informações específicas. Para facilitar a análise, os resultados são mostrados em duas fases distintas. Na primeira fase, o usuário pode escolher o número de compostos a ser exibido no cabeçalho da seção *Search* e o resultado inicial pode ser organizado de acordo com vários parâmetros da base. A coluna de resultados da primeira fase apresenta a identificação PK/DB (MID), estrutura 2D, nome padrão dos compostos, representação SMILES, peso molecular e propriedades farmacocinéticas disponíveis, como mostrado na Figura 5.1. A segunda fase possibilita ao usuário, através de uma ligação (do inglês, *link*) direta no nome dos compostos, acesso a informações

detalhadas da base como estrutura 3D, propriedades farmacocinéticas, propriedades físico-químicas e ação farmacológica.

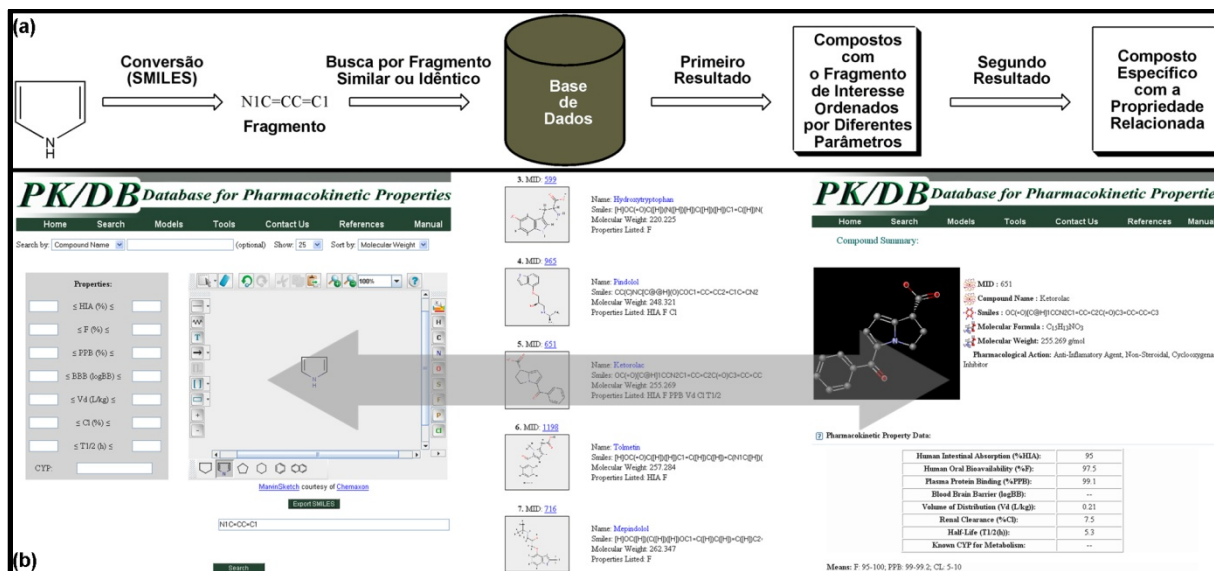


Figura 5.1 – PK/DB é a primeira base de dados de propriedades farmacocinéticas brasileira. (a) Uma visão geral da arquitetura do sistema da base PK/DB. (b) Esquema ilustrativo das etapas de busca e análise de resultados¹²⁶.

5.4 Predição de Propriedades Farmacocinéticas

A base PK/DB integra seis modelos *in silico* para a avaliação de propriedades farmacocinéticas e físico-químicas (*Predict It Service*), incluindo biodisponibilidade oral em humanos, ligação às proteínas plasmáticas, absorção intestinal humana, permeabilidade da barreira hematoencefálica, inibição da glicoproteína-P e solubilidade em água. Os modelos foram desenvolvidos e validados em nosso laboratório como descrito no capítulo 4, empregando a técnica de fragmentos moleculares especializados (HQSAR)^{47,88,121-122}. Pesquisadores do Brasil e do exterior podem testar suas moléculas em fase de desenvolvimento nos modelos da base PK/DB através do serviço *Predict It Service*, realizado mediante requisição.

Desde o lançamento da base, em maio de 2008, foram feitas predições *in silico* através do serviço *Predict It Service* para instituições como: National Cancer Institute – Division of Cancer Prevention – EUA, University of Pune – Índia, Department of

Biotechnology – Ministry of Science and Technology – Índia, Instituto de Química Orgânica – Consejo Superior de Investigaciones Científicas – Espanha, Faculdade de Farmácia – Universidade Federal do Rio de Janeiro – Brasil, Department of Zoology – Sri Venkateswara University – Índia, Laboratoire de Chimie Thérapeutique – Université de Nantes – França, Instituto Vita Nova – Brasil, entre outras.

Desde seu lançamento até o presente momento (julho de 2011) a base teve mais de 42 mil acessos refletindo a sua utilidade e importância no processo de pesquisa e desenvolvimento de moléculas bioativas com propriedades farmacocinéticas otimizadas. A evolução de acessos a base PK/DB desde o seu lançamento pode ser conferido no gráfico da Figura 5.2.

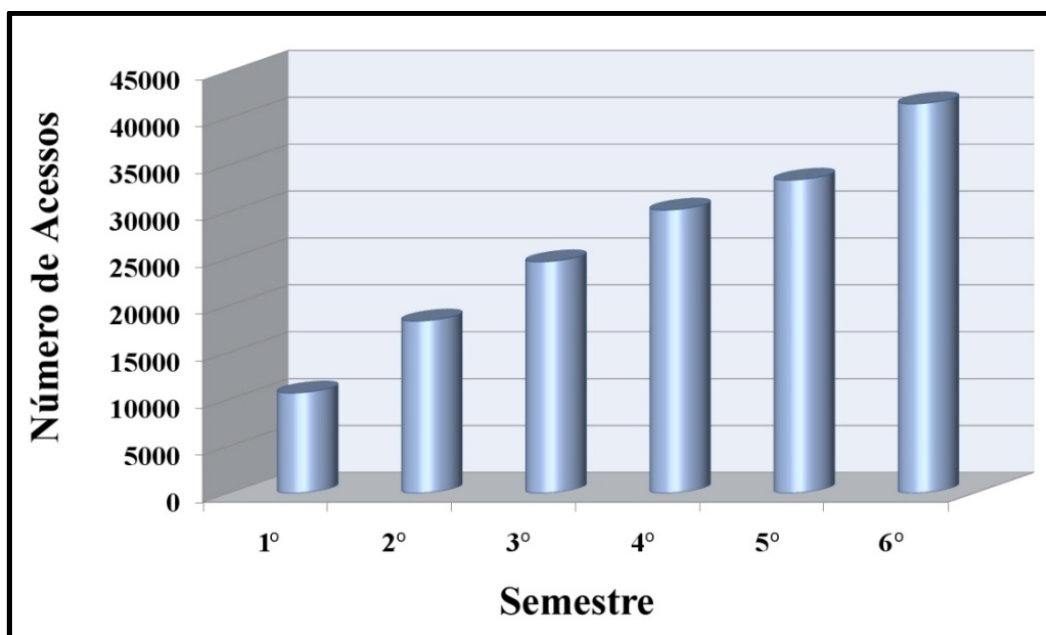


Figura 5.2 – Acessos a base PK/DB no período entre maio/2008 e julho/2011.

O desenvolvimento da primeira base de dados farmacocinéticos brasileira foi um trabalho pioneiro de ótima aceitação e repercussão. Aplicada com sucesso na busca por informações associadas às propriedades farmacocinéticas da base, e aos modelos *in silico*, PK/DB é uma ferramenta de grande utilidade no processo de pesquisa por candidatos a fármacos com propriedades otimizadas.

CAPÍTULO 6

**Estágio no
Exterior**

6 Estágio de Doutorado no Exterior

6.1 Estágio

Esta parte do trabalho foi desenvolvida entre maio/2009 e maio/2010 na *Eshelman School of Pharmacy* - Universidade da Carolina do Norte em Chapel Hill, EUA sob a orientação do Prof. Dr. Alexander Tropsha. O estágio foi realizado com apoio da agência governamental de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) através da bolsa de doutorado sanduíche do Programa de Doutorado no País com Estágio no Exterior (PDEE), com o objetivo de aprimorar e aprender novas técnicas de modelagem molecular de bases de dados complexas.

6.2 Bases de dados Complexas

São consideradas bases de dados complexas aquelas que se referem a sistemas biológicos complexos, tais como toxidez causada por agentes externos, receptores que atuam no combate ao câncer, interações entre sinalizadores de resposta, e receptores no sistema nervoso central (SNC), entre outros.

6.2.1 GPCRs

A família de receptores acoplados à proteína G (do inglês, *G protein-coupled receptors* – GPCRs) é uma das maiores codificada no genoma humano e representa uma significativa porção de alvos para fármacos. No sistema de sinalização celular do qual os receptores GPCRs fazem parte, as proteínas G são intermediários essenciais

para o mecanismo de transdução de sinal, ligando-se a diversos GPCRs, localizados na superfície celular¹³⁰. Os GPCRs podem ser ativados por ligantes como, por exemplo, hormônios, neurotransmissores e fatores de crescimento. Uma vez ativado o receptor, o sinal é transmitido pela proteína G e termina com a resposta mediada pela ação de moléculas efetoras¹³¹.

Os receptores serotoninérgicos (5-HT) fazem parte da família de GPCRs e são responsáveis pela modulação de desordens relacionadas aos transtornos comportamentais humanos como, por exemplo, depressão, ansiedade, síndrome do pânico, entre outros. Nessa série de transtornos também está listada a enxaqueca, uma doença neurológica caracterizada por fortes dores de cabeça comumente associadas a uma variedade de outros sintomas, por exemplo, náusea, vômito, fotofobia e fonofobia¹³². A única classe de medicamentos específica para o tratamento da enxaqueca é conhecida como triptanos, os quais são baseados na subestrutura do neurotransmissor natural serotonina (Figura 6.1). Entretanto, medicamentos de escolha como sumatriptano (Imigran[®]) e naratriptano (Naramig[®]) não são eficazes em todos os pacientes, sendo que em alguns casos, apresentam efeitos colaterais indesejados, pois também atuam sobre o receptor 5-HT_{1B} presentes nos vasos coronários onde promovem vasoconstrição do mesmo. Assim, triptanos são contra-indicados a pacientes que também sofrem de doenças cardiovasculares ou hipertensão arterial¹³³.

Descobertas recentes indicam que compostos que interagem simultaneamente como agonistas em relação ao triplo de receptores serotoninérgicos 5-HT_{1B}, 5-HT_{1D}, e 5-HT_{1F} e possuem diferentes estruturas químicas básicas quando comparados aos triptanos, podem ser agentes terapêuticos eficazes e seguros no tratamento de crises de enxaqueca¹³⁴. Esse é o caso do lasmiditano, composto baseado em uma série estruturalmente diferente dos triptanos, piridinoil-piperidina (Figura 6.1), e com alta afinidade e seletividade para o receptor 5-HT_{1F} tendo também afinidade pelos receptores 1B e 1D¹³⁵. O lasmiditano mostrou potência e eficácia quando administrado por via oral em modelos animais usados para testar compostos anti-enxaqueca e eficácia em aliviar sintomas da enxaqueca clinicamente¹³⁶. Esses resultados promissores, entre outros, são evidências crescentes sobre o importante papel que a ativação do receptor 5-HT_{1F} possui no tratamento contra a enxaqueca.

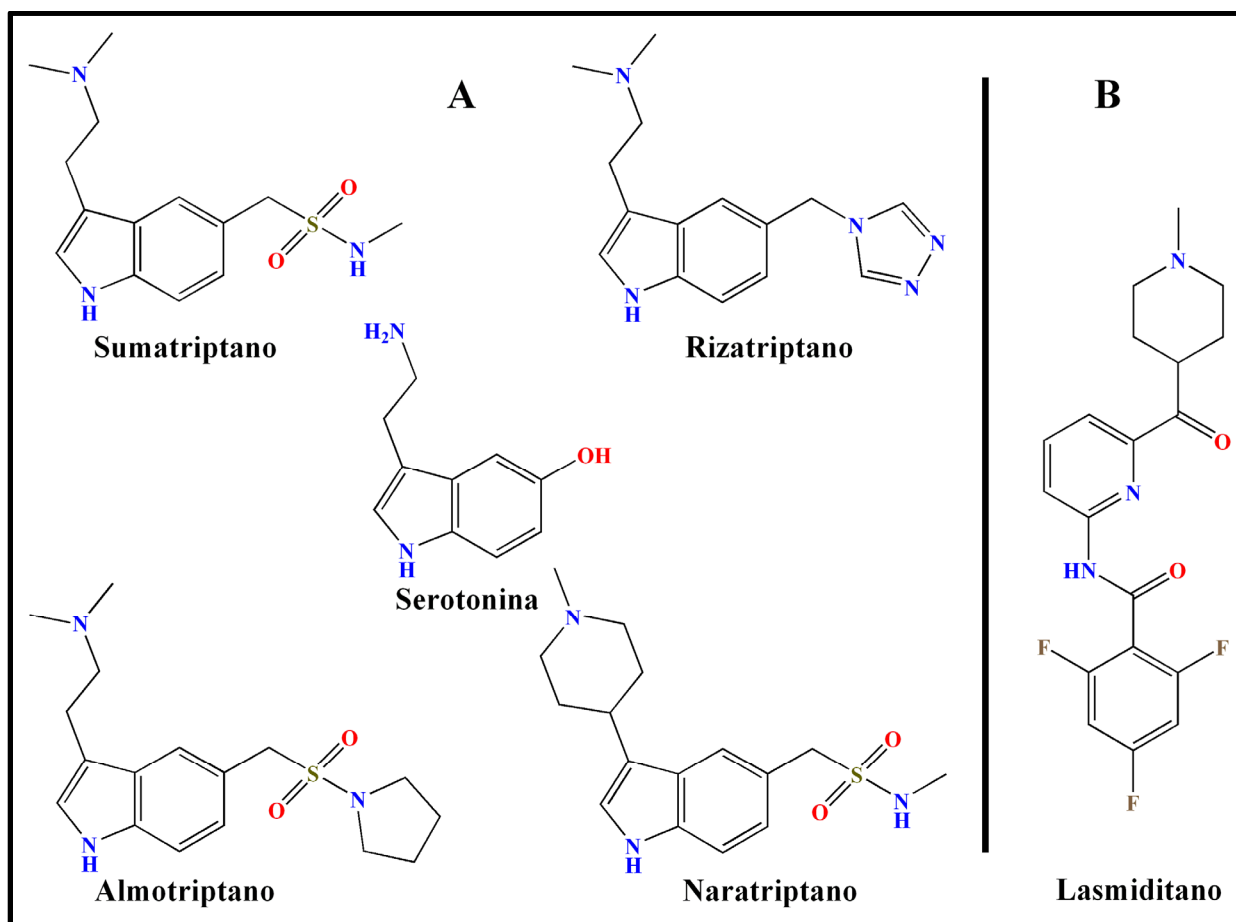


Figura 6.1 – (A) Fármacos baseados na estrutura básica da serotonina (triptanos). (B) Molécula com potencial atividade antiemigração não baseada na estrutura dos triptanos (lasmiditano).

A possibilidade de ação mútua (polifarmacologia) em três receptores diferentes no combate contra a enxaqueca originou o trabalho desenvolvido no estágio no exterior. Com o objetivo de coletar, organizar e investigar conjuntos de dados complexos ($5\text{-HT}_{1B/1D/1F}$) usando as metodologias de QSAR conhecidas como o método do vizinho mais próximo e máquinas de vetor de suporte, o trabalho apresentado a seguir descreve estudos sobre novos compostos que podem atuar no tratamento da enxaqueca.

6.3 Metodologia

6.3.1 Conjunto de dados

Foram criados três conjuntos de dados com 531 ligantes e não ligantes dos receptores 5-HT_{1B/1D/1F} (Tabela 6.1). Os dados foram coletados e organizados a partir da literatura e das bases de dados PDSP K_i (do inglês, *Psychoactive Drug Screening Program*), ChEMBL e Wombat¹³⁷⁻¹³⁹. Ligantes foram definidos como os compostos com valores de constante de afinidade (K_i) < 10 μ M, sendo considerados não ligantes os compostos com valores de $K_i \geq 10 \mu$ M.

Tabela 6.1 – Conjunto de dados dos receptores 5-HT

5-HT Subtipo	Ligantes ^[a]	Não Ligantes ^[b]	Total	Base de Dados
1B	91 (53%)	80 (47%)	171	PDSP
1D	87 (52%)	81 (48%)	168	PDSP
1F	132 (69%)	60 (31%)	192	PDSP, ChEMBL, Wombat

[a] Ligantes ($K_i < 10 \mu$ M). [b] Não Ligantes ($K_i \geq 10 \mu$ M).

A identificação de ligante ou não ligante foi atribuída a cada um dos compostos como zero (0) ou um (1), respectivamente, e foi usada como variável dependente no processo de modelagem molecular classificatória. Com exceção do conjunto de dados para o receptor 5-HT_{1F}, os outros dois conjuntos de receptores apresentam bom balanço em sua composição, com aproximadamente 50:50 de ligantes e não ligantes. A distribuição entre as categorias que compõem o conjunto de dados é uma questão importante nas discussões que envolvem métodos coerentes sobre modelagem^{98,140}. No caso do conjunto para o receptor 1F, foram atribuídos pesos para equilibrar a distribuição entre as classes.

Para a geração das estruturas químicas foi empregada a plataforma MOE v2007.09 (Chemical Computing Group Inc. Montreal, Canadá)¹⁴¹. As estruturas foram

normalizadas como descrito na sessão 3.1 do capítulo 3. Os descritores adimensionais (0D), unidimensionais (1D) e bidimensionais (2D) foram calculados no programa DRAGON 5.5 (Talete srl. Milano, Itália)¹⁴². Através do método *5-fold* de validação externa, os compostos dos três conjuntos de dados foram divididos randomicamente em cinco conjuntos para validação externa (20%) e cinco conjuntos para modelagem (80%) (Figura 6.2)¹⁴³. No segundo caso, os grupos foram posteriormente divididos em conjuntos treinamento e teste pelo método Esfera de Exclusão (do inglês, *Sphere Exclusion – SE*)¹⁴⁰.

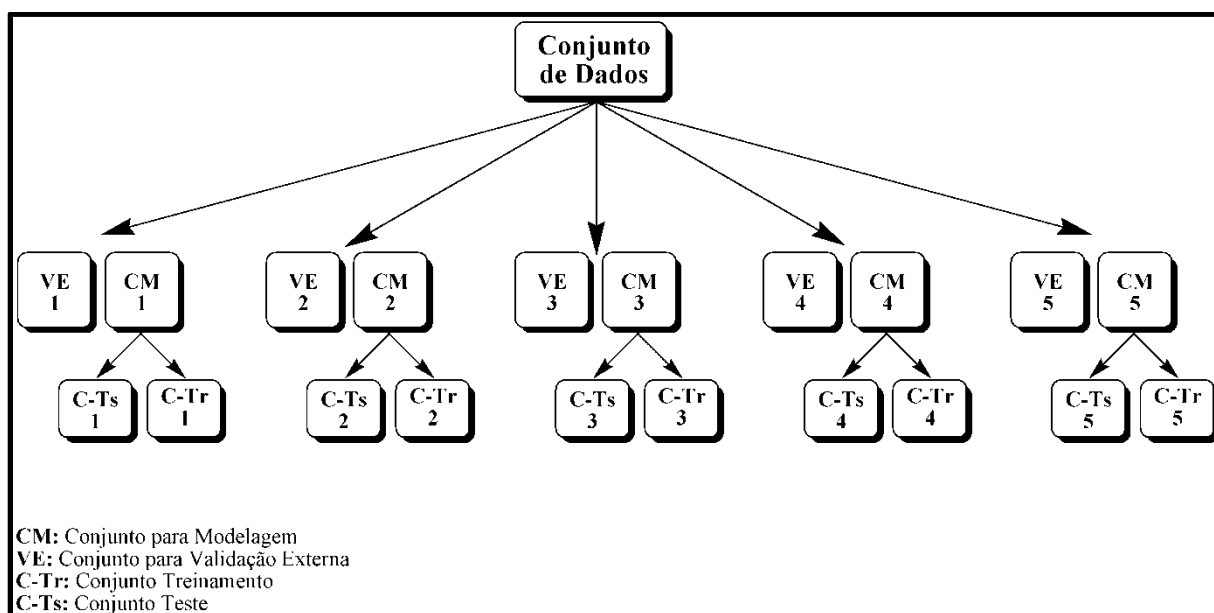


Figura 6.2 – Divisão randômica do conjunto de dados seguindo o princípio *5-fold* de validação externa.

6.3.2 Método do Vizinho Mais Próximo

O método do vizinho mais próximo (do inglês, *k-Nearest Neighbor – kNN*) é uma análise de reconhecimento de padrões comumente utilizada em processos classificatórios baseados em medidas indiretamente relacionadas com a propriedade estudada. O modo *kNN* faz a classificação de acordo com a similaridade de uma amostra desconhecida (conjunto teste) em relação a uma amostra conhecida (conjunto

treinamento). A similaridade é medida de acordo com a distância que separa dois pontos no espaço multidimensional, pontos esses que representam os compostos estudados no processo de modelagem molecular. Várias formas de mensurar distâncias podem ser usadas para expressar a similaridade entre compostos, mas a distância euclidiana (Equação. 6.1) é provavelmente a mais comum e foi empregada neste estudo^{140,144}:

$$d_{ij} = \sqrt{(\sum_{k=1,N}(d_{i,k} - d_{j,k})^2)} \quad \text{Eq (6.1)}$$

onde d_{ij} é a distância entre os pontos i e j no espaço de N dimensões. A classificação de uma amostra desconhecida é feita baseada no vizinho mais próximo da mesma, sendo assim a similaridade fruto da menor distância entre os pontos analisados. A Figura 6.3 traz a representação do processo classificatório k NN no espaço bidimensional.

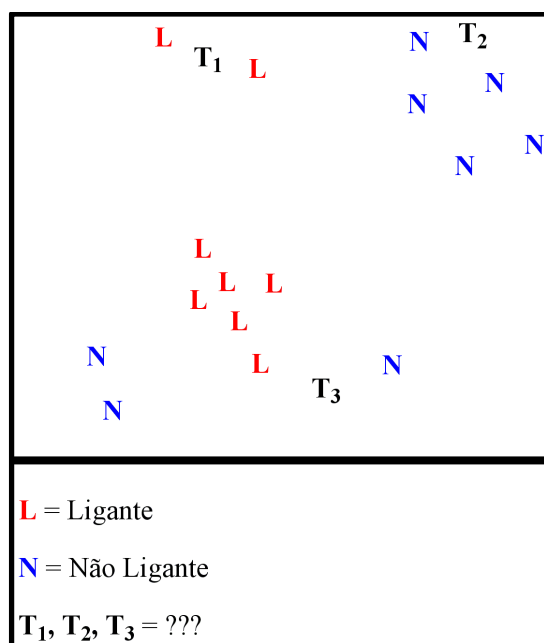


Figura 6.3 – Representação da técnica k NN no espaço bidimensional.

Os compostos do conjunto treinamento são representados pelas letras L e N (ligantes e não ligantes, respectivamente) e os do conjunto teste pelas letras T₁, T₂ e T₃. Os compostos T₁ e T₂ são facilmente classificados como ligantes e não ligantes pela comparação com os seus vizinhos mais próximos. Entretanto, a classificação do

composto T_3 não é tão clara, pois esse ponto parece estar equidistante a dois outros pontos pertencentes à compostos de categorias diferentes. Embora um dos vizinhos, provavelmente, esteja mais próximo do ponto analisado, existe uma incerteza nesta predição. Com a exceção de um vizinho, os outros pontos próximos a T_3 são ligantes o que leva a predição do mesmo como L^{145} . Esta ambiguidade mostra a importância da variação do número de vizinhos (k) para que se possam alcançar os modelos com maior poder de prever corretamente novos compostos com potencial ação anti-tenaxaeca.

6.3.3 Máquinas de Vetor de Suporte

As máquinas de vetor de suporte (do inglês, *support vector machines* – SVM) se constituem em uma técnica de aprendizado supervisionado aplicada a classificação de compostos de acordo com suas características químicas. O primeiro passo para a modelagem usando o método SVM é o treinamento dos modelos através de conjuntos de treinamento que representem o espaço químico-biológico estudado. O sistema de aprendizado SVM é baseado pela teoria de aprendizado estatístico que estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa capacidade de prever corretamente novos compostos no domínio em que ocorreu o aprendizado. Utilizando uma função Kernel, onde as mais comuns são a polinomial, gaussiano ou sigmoide, o espaço N dimensional é separado por hiperplanos onde os vetores de suporte são definidos a partir do produto escalar entre vetores (Figura 6.4). Dados não linearmente separáveis são redesenhados em um novo espaço linearmente separável. A este novo espaço é aplicado um hiperplano que separa duas classes, no caso da classificação binária^{146,147}. Na modelagem molecular essas duas classes podem conter informações como moléculas ligantes/não ligantes ou agonistas/antagonistas para um receptor específico.

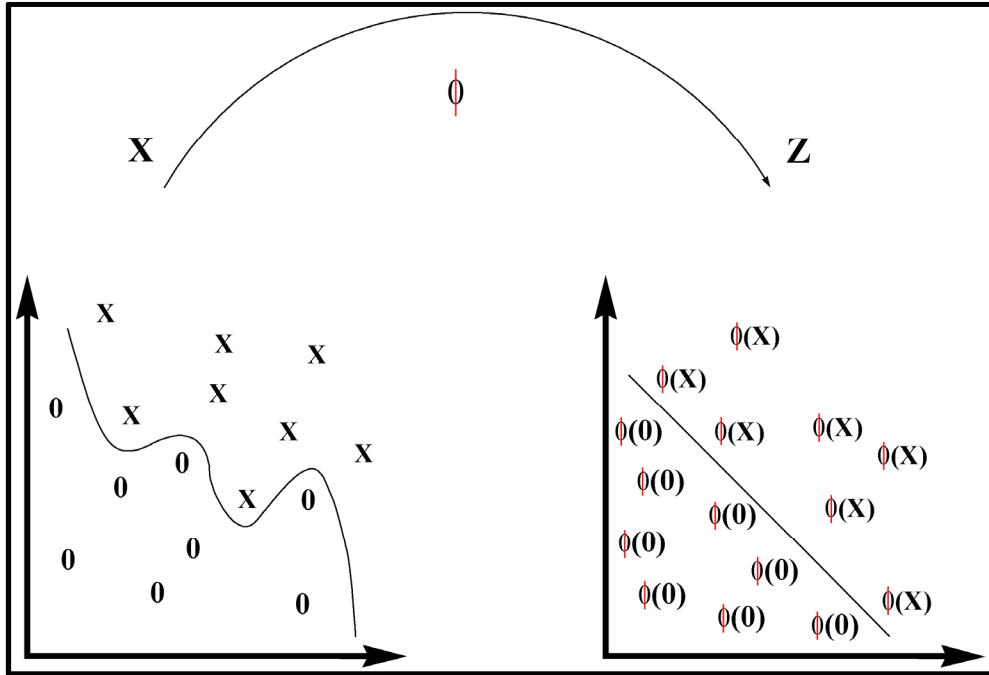


Figura 6.4 – Mapeamento do espaço de entrada via função Kernel.

A partir da transformação apresentada na Figura 6.4 uma classificação pode ser feita em Z . Embora a dimensão do espaço aumente em Z , a complexidade diminui, pois a classificação no espaço de entrada X só era possível utilizando métodos não lineares. Entretanto, no espaço de características Z a separação pode ser feita apenas com um hiperplano, ou um método linear.

A Figura 6.5 apresenta os vetores de características que são mais próximos das margens do hiperplano, também chamados de vetores suporte. Cada ponto no gráfico representa um vetor sendo que vetores fora da margem não interferem no hiperplano e na separação de classes. Por esse motivo os vetores encontrados nas margens do hiperplano são chamados de vetores suporte¹⁴⁶.

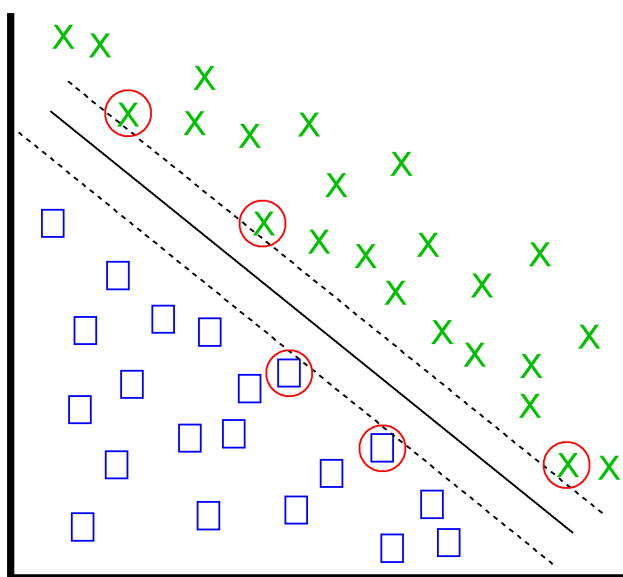


Figura 6.5 – Hiperplano com margens maximizadas e vetores suporte circundados em vermelho.

Os vários conjuntos treinamento empregados na modelagem molecular dos receptores serotoninérgicos tornam os modelos SVM desenvolvidos neste trabalho robustos e preditivos para novos compostos com potencial ação antiemética.

6.4 Cálculo dos Descritores

Um total de 2489 descritores 0D, 1D e 2D foram calculados para os 531 ligantes e não ligantes dos receptores 5-HTs coletados. Então, os descritores tiveram os coeficientes de correlação avaliados e um entre dois descritores com $r^2 \geq 0.95$ foi excluído. Após a eliminação dos descritores redundantes, 457, 458 e 395 descritores químicos relevantes remaneceram para os conjuntos 5-HT_{1B}, 5-HT_{1D} e 5-HT_{1F}, respectivamente. Com o objetivo de prevenir disparidades entre os descritores, os seus valores foram normalizados entre 0 e 1, baseados no valor máximo e mínimo de cada descritor encontrado no conjunto de dados. O uso da normalização evita que pesos desproporcionais sejam atribuídos a descritores que possuem larga variância na escala numérica quando comparados a outros no espaço multidimensional que não possuem tal variância (i.e., massa molar vs. número de doadores de hidrogênio).

6.5 Índice de Classificações Corretas

Os modelos de QSAR para predição de ligantes dos receptores serotoninérgicos 5HTs foram desenvolvidos usando as técnicas *k*NN e SVM. Os modelos *in silico* foram validados pelo método do índice de classificações corretas (do inglês, *Correct Classification Rate* – CCR), representado na Equação 6.2.

$$\text{CCR} = \frac{1}{2} \left(\frac{N_{1\text{Corr}}}{N_{1\text{Tot}}} + \frac{N_{2\text{Corr}}}{N_{2\text{Tot}}} \right) \quad \text{Eq (6.2)}$$

onde $N_{1\text{Corr}}$ é o número de ligantes preditos corretamente pelo modelo, $N_{1\text{Tot}}$ o número total de ligantes encontrados no conjunto teste, $N_{2\text{Corr}}$ o número de não ligantes preditos corretamente pelo modelo, $N_{2\text{Tot}}$ o número total de não ligantes no conjunto teste^{140,144}. Diferentes índices foram usados na avaliação dos modelos: $\text{CCR}_{\text{Trein}}$ para avaliar os modelos desenvolvidos validados apenas pelo método LOO, $\text{CCR}_{\text{Teste}}$ para os modelos através da predição dos conjuntos testes e $\text{CCR}_{5\text{-fold}}$ para os modelos através da predição de conjuntos externos.

6.6 Domínio de Aplicação

Idealmente os modelos de QSAR podem prever a propriedade alvo para qualquer composto desde que cada descritor químico possa ser calculado. Entretanto, modelos de QSAR preveem a atividade de novas moléculas a partir da interpolação dos compostos do conjunto treinamento. Assim, o domínio de aplicação (do inglês, *Applicability Domain* – AD), ou limite de similaridade, deve ser aplicado para impedir extrapolações extremas por parte dos modelos quando esses preveem compostos que são dissimilares dos compostos que deram origem aos modelos¹⁴⁸. Em suma, a

distribuição das distâncias entre os compostos no conjunto treinamento são calculadas para produzir o limite do domínio de aplicação (D_T) calculado como:

$$D_T = \bar{y} + Z\sigma \quad \text{Eq (6.3)}$$

onde \bar{y} é a média da distância euclidiana dos k vizinhos mais próximos para cada composto pertencente ao conjunto treinamento, σ é o desvio padrão dessas distâncias euclidianas, e Z é um parâmetro arbitrário para controle do nível de significância e flexibilidade da extrapolação realizada pelos modelos na predição de novas moléculas. Baseado em estudos anteriores, o valor de Z foi definido como 0,5; valor que introduz o limite para o qual cada composto será predito. Esse valor representa até a metade do desvio padrão mencionado acima. Assim, se a distância de um composto externo até pelo menos um dos seus vizinhos no conjunto treinamento exceder esse limite, a predição é considerada inviável, ou não confiável^{149,150}.

6.7 Modelos QSAR k NN e SVM

O desenvolvimento dos modelos k NN para o receptor 5-HT_{1B} empregou 457 descritores DRAGON como variável independente, derivados de 171 ligantes e não ligantes divididos em 136 compostos nos conjunto treinamento e teste, e 35 compostos para a validação externa. O número de vizinhos mais próximos foi variado de 1 a 5. No total foram gerados 13.803 modelos k NN, sendo que, após a aplicação dos filtros CCR_{train} e CCR_{teste} de valores iguais ou superiores a 0,85, houve 2.578 modelos restantes para a validação externa (Tabela 6.2). Como parte da estratégia de modelagem combinatória, o método SVM foi aplicado sobre o mesmo conjunto de dados e descritores para o desenvolvimento de novos modelos *in silico*. Esses modelos adicionais tendem a aumentar as chances de sucesso na busca por novos ligantes frente ao receptor 5-HT_{1B}. No total 5.775 modelos foram obtidos com o método SVM, desses, 107 tiveram os parâmetros CCR_{train} e $CCR_{\text{teste}} \geq 0,85$ sendo selecionados para a

avaliação externa (Tabela 6.1). As cinco divisões resultantes do método de validação *5-fold* fazem com que a integração dos modelos finais seja necessária para a real avaliação do poder de predição dos modelos através da validação externa. Assim, a Tabela 6.2 apresenta os resultados individuais para os vários modelos gerados a partir de diferentes conjuntos treinamento e teste, e selecionados através dos valores de CCRs como descrito acima. Também é apresentado o resultado para o consenso entre as cinco divisões $[(1 + 2 + 3 + 4 + 5)/5]$, pelo qual podemos avaliar a coerência da predição dos modelos integrados, e ter a avaliação correta do poder de predição do sistema de modelos através do consenso para o CCR_{5-fold}

Tabela 6.2 – Parâmetros estatísticos para os modelos *k*NN e SVM 5-HT_{1B}

Método	Divisão	Nº de Modelos	Sens. ^[a] (%)	Esp. ^[b] (%)	CCR _{5-fold}
<i>k</i> NN	1	636	68	100	0,84
	2	442	77	95	0,86
	3	354	100	80	0,90
	4	747	95	64	0,80
	5	399	93	80	0,86
	Consenso		2578	87	84
SVM	1	27	84	69	0,76
	2	23	83	81	0,82
	3	23	100	81	0,91
	4	15	88	69	0,79
	5	19	88	81	0,85
	Consenso		107	87	76

[a] Sensibilidade: N_{1Corr}/N_{1Tot} . [b] Especificidade: N_{2Corr}/N_{2Tot} .

No caso do receptor 5-HT_{1D}, 458 descritores DRAGON foram identificados com potencial para modelagem a partir do conjunto de dados formado por 168 compostos, que posteriormente foram divididos em 134 compostos para conjunto treinamento e teste, e 34 para validação externa. Para o receptor 5-HT_{1F}, 395 descritores foram derivados de 192 moléculas que foram divididas em 153 para os conjuntos treinamento e teste, e 39 para validação externa. Em ambos os casos centenas de modelos foram obtidos pelos métodos *k*NN e SVM. Após o emprego dos filtros $CCRs \geq 0,85$ o número total de modelos com potencial preditivo foi reduzido, sendo a nova quantidade apresentada nas Tabelas 6.3 e 6.4. Os modelos restantes foram então submetidos ao processo de validação externa.

Tabela 6.3 – Parâmetros estatísticos para os modelos k NN e SVM 5-HT_{1D}

Método	Divisão	Nº de Modelos	Sens. ^[a] (%)	Esp. ^[b] (%)	CCR _{5-fold}
k NN	1	1812	100	74	0,87
	2	554	100	77	0,88
	3	791	89	86	0,88
	4	2809	91	77	0,84
	5	3530	90	85	0,87
	Consenso		9496	94	80
SVM	1	17	100	75	0,88
	2	15	83	81	0,82
	3	13	100	75	0,88
	4	15	100	87	0,94
	5	26	82	87	0,85
	Consenso		102	93	81

[a] Sensibilidade: $N_{1\text{Corr}}/N_{1\text{Tot}}$. [b] Especificidade: $N_{2\text{Corr}}/N_{2\text{Tot}}$.

Tabela 6.4 – Parâmetros estatísticos para os modelos k NN e SVM 5-HT_{1F}

Método	Divisão	Nº de Modelos	Sens. ^[a] (%)	Esp. ^[b] (%)	CCR _{5-fold}
k NN	1	547	93	71	0,82
	2	1052	96	70	0,83
	3	482	90	100	0,95
	4	702	92	67	0,80
	5	924	100	67	0,83
	Consenso		3707	94	75
SVM	1	8	96	83	0,90
	2	10	96	75	0,86
	3	12	96	100	0,98
	4	12	96	83	0,90
	5	10	89	75	0,82
	Consenso		52	95	83

[a] Sensibilidade: $N_{1\text{Corr}}/N_{1\text{Tot}}$. [b] Especificidade: $N_{2\text{Corr}}/N_{2\text{Tot}}$.

Os resultados da validação externa sugerem que existe atividade intrínseca relacionada aos ligantes dos receptores 5-HTs e que esta é razoavelmente bem descrita pelos métodos k NN e SVM usando descritores DRAGON como variável independente no processo de modelagem molecular.

6.8 Triagem Virtual em Paralelo

O primeiro passo para a triagem virtual em paralelo (Figura 6.6) foi a filtragem de 281.410 compostos através do domínio de aplicação dos conjuntos de dados 5HT_{1B/1D/1F} usados na modelagem molecular^{140,147}. Os compostos são provenientes de bases de dados das empresas Maybridge, Asinex e TimTec, juntamente com 1.200 compostos testados e aprovados para uso clínico da base de dados Prestwick¹⁵¹⁻¹⁵⁴. Esse processo resultou na seleção de 12.350, 10.660 e 12.822 compostos para as bases 5HT_{1B}, 5HT_{1D} e 5HT_{1F}, respectivamente (Figura 6.7). As moléculas selecionadas podem ser previstas pelos modelos, como ligantes ou não ligantes, com margem de segurança aceitável, pois suas estruturas químicas estão dentro do domínio de aplicação dos modelos desenvolvidos. Em seguida foi realizada a predição consenso para as moléculas selecionadas usando todos os seis diferentes tipos de modelos, que somam 15.781 *k*NN e 245 SVM.

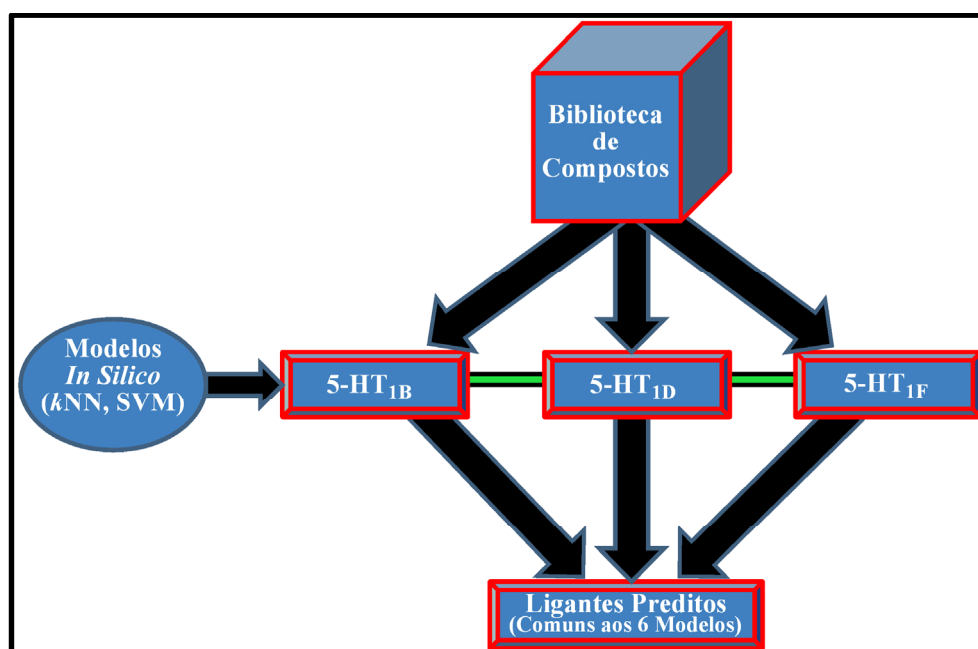


Figura 6.6 – Esquema ilustrativo da triagem virtual em paralelo.

Após a predição consenso, 3.421 compostos foram classificados como ligantes nos diferentes modelos *in silico*. Pelo fato dos modelos atuarem na predição da mesma base de dados muitos desses compostos são sobrepostos, e, são justamente os

compostos sobrepostos preditos por todos os modelos *k*NN e SVM o alvo da triagem virtual em paralelo. Esse consenso aumenta a chance de serem encontrados compostos moduladores da tríade de receptores 5-HTs. Assim, 132 compostos foram preditos como ligantes pelos seis diferentes tipos de modelos *in silico* (Figura 6.6). O processo integral da triagem virtual em paralelo é mostrado na Figura 6.7. Entre os 132 compostos selecionados, 10 foram priorizados para o ensaio *in vitro* baseado na diversidade química, custo e disponibilidade de aquisição.

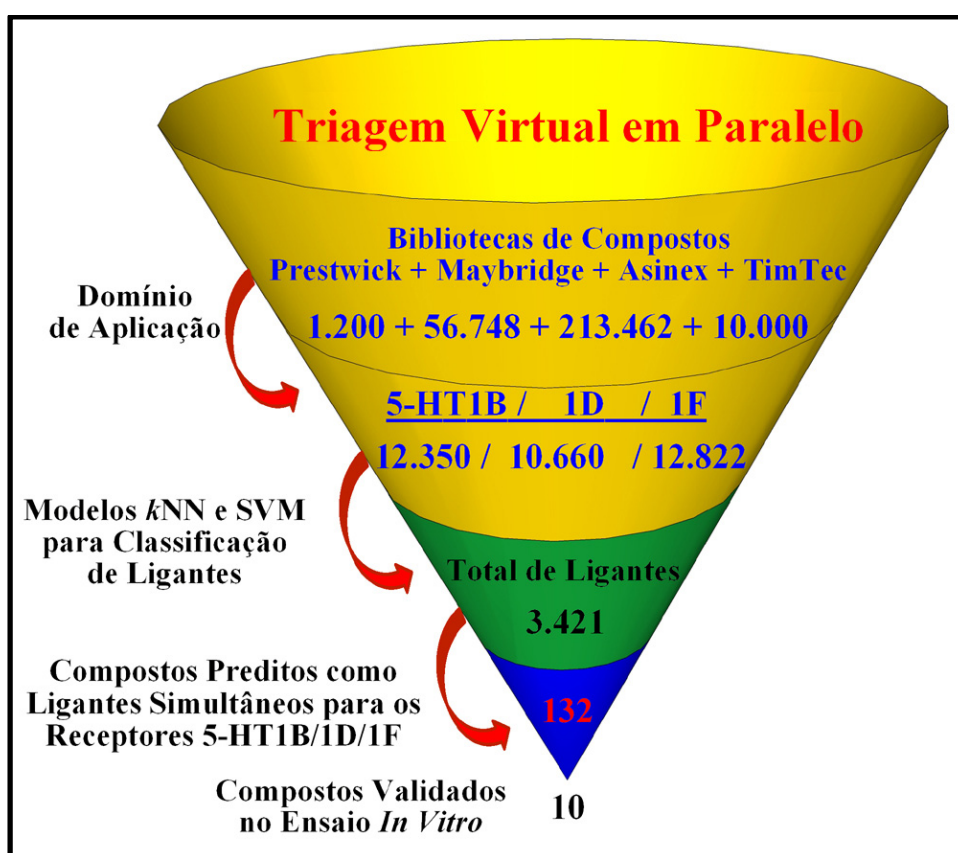


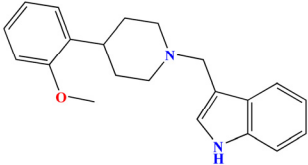
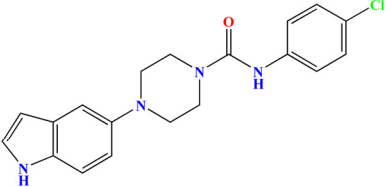
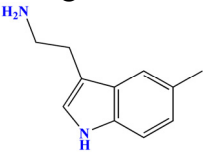
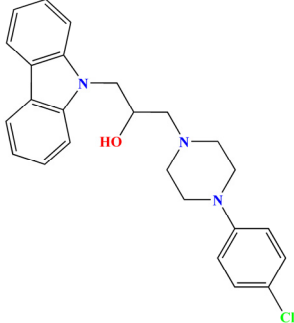
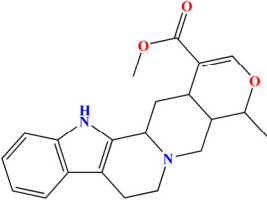
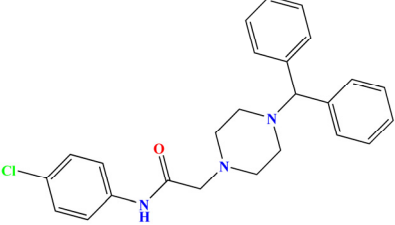
Figura 6.7 – Processo usado para identificação de novos ligantes 5-HTs.

6.9 Validação Experimental

Os ensaios *in vitro* foram realizados no programa PDSP do Instituto Nacional de Saúde Mental na Universidade da Carolina do Norte, Chapel Hill – NC, EUA sob a supervisão do Prof. Dr. Bryan L. Roth¹⁵⁵.

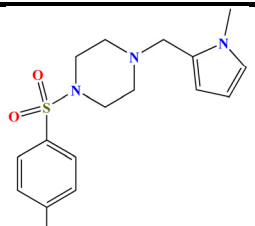
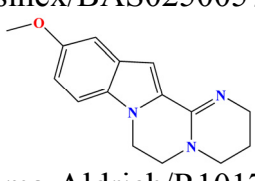
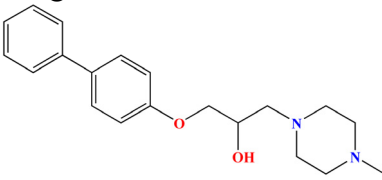
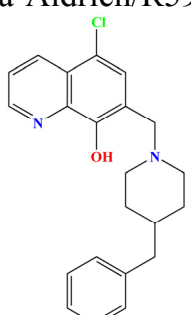
Dez compostos estruturalmente diversos (Tabela 6.5) foram selecionados a partir da triagem virtual em paralelo para validação experimental. Como resultado, três compostos foram confirmados como inibidores da ligação entre o ligante utilizado como padrão (*radioligant binding assay*) e o receptor 5-HT_{1B}. Da mesma forma, três compostos foram confirmados como ligantes para o receptor 5-HT_{1D}, o que significa uma taxa de sucesso de 30% em ambos os casos. Os valores para a constante de afinidade (K_i) variaram entre 0,02 e 5,11 μM para 5-HT_{1B} e entre 0,02 e 3,63 μM para 5-HT_{1D}, sendo que dois compostos mostraram valores de $K_i < 1 \mu\text{M}$ (**1** e **3**, Tabela 6.5). No primeiro caso, o composto **1** apresentou K_i de 0,39 para o receptor 5-HT_{1D}, sendo que no segundo caso o composto **3** apresentou $K_i < 1 \mu\text{M}$ para os dois receptores 1B e 1D, ($K_i = 0,02 \mu\text{M}$). É importante notar que nenhum desses compostos estava incluído na base de dados original que treinou os modelos preditivos.

Tabela 6.5 – Resultado da validação experimental para os 10 compostos selecionados como ligantes 5-HT_{1B/1D/1F}

ID	Estrutura Química Companhia/Código	Predição Consenso			Experimental K_i (μ M)		
		5-HT _{1B}	5-HT _{1D}	5-HT _{1F}	5-HT _{1B}	5-HT _{1D}	5-HT _{1F}
1	 Maybridge/HTS04089	0,16	0,18	0,08	5,11	0,39	ND
2	 Maybridge/BTB14566	0,12	0,11	0,15	> 10,00	3,63	ND
3	 Maybridge/AC15867	0,12	0,06	0,04	0,02	0,02	ND
4	 Asinex/BAS00898679	0,06	0,07	0,04	> 10,00	ND	ND
5	 Asinex/BAS00225301	0,07	0,23	0,02	1,00	ND	ND
6	 Asinex/BAS02070678	0,10	0,09	0,05	> 10,00	> 10,00	ND

Continua

Continuação

7		0,15	0,08	0,10	> 10,00	> 10,00	ND
	Asinex/BAS02500576						
8		0,12	0,12	0,06	> 10,00	> 10,00	ND
	Sigma-Aldrich/R101796						
9		0,15	0,19	0,06	> 10,00	> 10,00	ND
	Sigma-Aldrich/R593222						
10		0,13	0,15	0,08	> 10,00	ND	ND
	TimTec/ST014827						

*ND: indisponível. Predição Consenso: [(predições *k*NN + predições SVM)/2].

Como mencionado acima, o composto **1** apresentou afinidade para os receptores 5HT_{1B/1D}, bem como o composto **3**. Esses resultados cumprem parte do objetivo principal deste projeto de encontrar ligantes que possam atuar no trio de receptores 5HT_{1B/1D/1F}. Ensaio adicionais devem ser realizados para os compostos **4**, **5**, e **10** para que seja definido o K_i frente ao receptor 1D. Tendo em vista os ensaios realizados, é grande a probabilidade do composto **5** apresentar afinidade na região de baixo micromolar para o receptor 1D, elevando a taxa de sucesso para compostos que atuam nos receptores 5HT_{1B/1D} simultaneamente.

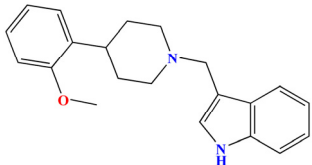
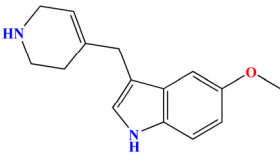
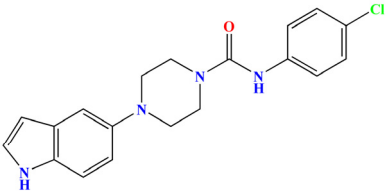
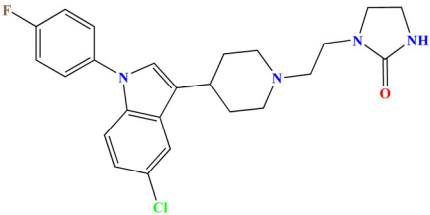
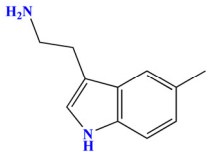
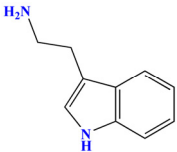
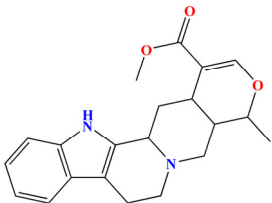
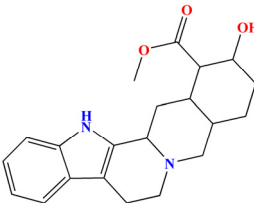
Também devem ser realizados ensaios adicionais para os dez compostos e assim determinar a afinidade desses perante o receptor 5-HT_{1F}. Entretanto, os ensaios já realizados indicam que há grande chance dos dois compostos com confirmada afinidade pelos receptores 1B e 1D (**1**, **3**; Tabela 6.5) também mostrarem o mesmo

comportamento quando ensaiados contra o receptor 1F. O mesmo deve ocorrer para o composto **5**, com afinidade comprovada para o receptor 1B.

Ensaios funcionais foram realizados para determinar a natureza agonista/antagonista da ligação entre os receptores e os compostos testados¹⁵⁵. Todos os compostos selecionados na triagem virtual em paralelo com atividade perante os receptores 5-HTs (**1, 2, 3, 5**; Tabela 6.5) ,mostraram natureza agonista em relação aos respectivos receptores.

A diversidade química dos compostos validados experimentalmente em relação aos compostos dos conjuntos de dados foi avaliada através de uma simples busca por similaridade. Os cálculos foram realizados com uma combinação de descritores de similaridade métrica: chaves estruturais MACCS de 16 bits e o coeficiente de Tanimoto (T_C)^{140,144}. Baseado nessa combinação, os compostos vizinhos mais próximos vindos das bases de dados que originaram os modelos, e os quatro compostos validados como ligantes experimentalmente, são relacionados na Tabela 6.6.

Tabela 6.6 – Vizinhos mais próximos (conjunto treinamento) dos 4 compostos validados experimentalmente como ativos

ID	Composto Proposto pela Triagem Virtual	Vizinho mais Próximo (Conjunto 5-HT _{1B/1D})	Similaridade (T _c)
1			82%
2			85%
3			97%
5			77%

*Similaridade: Baseada nas chaves MACCS e no coeficiente de Tanimoto.

Os resultados obtidos mostram a validade da triagem virtual com consenso entre modelos integrados na priorização de compostos providos por coleções virtuais, sendo possível alcançar boas taxas de sucesso na identificação de novos ligantes para os receptores serotoninérgicos 5-HT. As atividades obtidas na faixa de baixo micromolar, aliadas à natureza agonista da interação dos compostos selecionados, reforçam a validade dos modelos *in silico* e ampliam a possibilidade de serem encontrados compostos com potencial atividade antiemessa.

CAPÍTULO 7

**Conclusões e
Perspectivas**

7 Conclusões e Perspectivas

Nesta tese de doutorado foram apresentados os resultados obtidos no decorrer de quatro anos de projeto. Várias técnicas e estratégias foram empregadas para a modelagem molecular das propriedades farmacocinéticas absorção intestinal humana, metabolismo de primeira passagem, permeabilidade da barreira hematoencefálica, inibição da glicoproteína-P, solubilidade e também para os receptores serotoninérgicos 5-HT_{1B/1D/1F}.

Diversos modelos *in silico* foram desenvolvidos e validados para as propriedades farmacocinéticas estudadas. Parâmetros estatísticos relevantes relacionados à modelagem foram variados com o objetivo de serem alcançados os modelos mais robustos e de maior poder preditivo. A diversidade química e a distribuição dos valores da propriedade são pontos atrativos dos conjuntos de dados criados, proporcionando excelente cobertura do espaço químico-biológico a ser explorado.

A gama de modelos *in silico* obtidos é extremamente positiva no sentido de prever a possibilidade dos fármacos serem administrados por via oral, atuarem em conjunto aumentando a eficiência de quimioterápicos (através da inibição do efluxo promovido pela glicoproteína-P) e também a possibilidade de atuarem no sistema nervoso central. As propriedades estudadas permitem as principais caracterizações farmacocinéticas pelas quais novos compostos bioativos devem passar.

Os modelos desenvolvidos nesta tese de doutorado, aliados aos modelos de biodisponibilidade e ligação às proteínas plasmáticas, desenvolvidos anteriormente, durante o mestrado de Tiago L. Moda, dão ao Laboratório de Química Medicinal e Computacional - IFSC/USP um portfólio completo de predições das propriedades farmacocinéticas, desde a administração até a distribuição de NCEs pelo corpo humano. Assim, foi desenvolvida uma ferramenta de grande utilidade no processo de pesquisa por candidatos a fármacos com propriedades farmacocinéticas otimizadas. Também durante o período desta tese de doutorado, foi desenvolvida a base de dados PK/DB, um trabalho genuinamente brasileiro, e que se transformou em uma

ferramenta muito útil para pesquisadores do mundo todo. O sucesso da base é comprovado pelo número crescente de acessos, desde seu lançamento, até a conclusão desta tese (julho de 2011), tivemos mais de 42.000 visitas. Entre os serviços mais procurados estão os de busca por propriedades farmacocinéticas usando subestruturas químicas ou nome de compostos, e o serviço de predição “*Predict It Service*” realizado mediante requerimento.

Por fim, durante o período de doutorado sanduíche na *Eshelman School of Pharmacy* - Universidade da Carolina do Norte em Chapel Hill, EUA sob a orientação do Prof. Dr. Alexander Tropsha, foi possível aprender diversas técnicas e métodos de modelagem molecular aplicadas nesta tese na pesquisa por compostos com potencial aplicação antiemessa, podendo também esses métodos serem aplicados a qualquer finalidade farmacodinâmica ou farmacocinética. Entre as técnicas destacam-se as de busca por similaridade, domínio de aplicação e triagem virtual em paralelo. A oportunidade deste estágio no exterior e o aprendizado dos métodos, incluindo *kNN* e *SVM*, permitiu uma visão mais ampla e concreta da modelagem molecular, bem como o desenvolvimento futuro de modelos com elevado crivo científico.

Publicações

No decorrer do doutorado houve a publicação ou submissão de artigos e um capítulo de livro, descritos a seguir:

Um artigo de revisão foi elaborado com o intuito de divulgar as propriedades farmacocinéticas importantes para o sucesso de NCEs nas pesquisas clínicas e descrever métodos de determinação, tanto *in silico* quanto *in vitro*. O artigo também apresenta novas perspectivas quanto à integração de modelos para a otimização de propriedades ADME.

Esse artigo intitulado “*Pharmacokinetic properties: an old problem with new perspectives*” de autoria de Káthia M. Honório, Tiago L. Moda e Adriano D. Andricopulo foi submetido para publicação no *Current Medicinal Chemistry* em julho de 2011 (Apêndice A).

Entre os modelos desenvolvidos neste doutorado o de HIA foi escolhido para ser publicado em primeiro lugar devido à importância dessa propriedade na P&D de fármacos. O artigo descreve os resultados obtidos na modelagem da absorção intestinal humana usando o método HQSAR e as perspectivas quanto à utilização dos modelos *in silico* na seleção de moléculas com HIA apropriada.

Esse artigo intitulado “*Consensus hologram QSAR modeling for the prediction of human intestinal absorption*” de autoria de Tiago L. Moda e Adriano D. Andricopulo foi submetido para publicação no *ChemMedChem* em julho de 2011 (Apêndice B).

O estágio de doutorado no exterior também rendeu um manuscrito onde são descritos os trabalhos de modelagem, realizados sobre os receptores serotoninérgicos 5-HT_{1B/1D/1F}, na busca por compostos com potencial atividade antiemessa usando os métodos *k*NN e SVM.

Esse manuscrito intitulado “*Parallel screening for novel antimigraine therapeutics that act on 5-HT_{1B}, 5-HT_{1D} and 5-HT_{1F} serotonin receptors using QSAR models and experimental validation*” de autoria de Tiago L. Moda, Xiang S. Wang,

Adriano D. Andricopulo, Tudor I. Oprea, Alexander Tropsha esta em fase final de preparação (Apêndice C).

Além de artigos, foi possível participar da elaboração de um capítulo de livro abordando os fundamentos das principais propriedades farmacocinéticas, usando os conhecimentos adquiridos durante o mestrado e doutorado. O capítulo também descreve as principais propriedades moleculares que influenciam os processos de ADME e a modelagem *in silico* da absorção intestinal humana, biodisponibilidade e metabolismo, entre outras. A escrita desse capítulo foi realizada em conjunto com a Profa. Dra. Káthia M. Honório da Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP) e com o Prof. Dr. Adriano D. Andricopulo (IFSC-USP).

Este capítulo intitulado “Propriedades farmacocinéticas e o planejamento de novas substâncias bioativas” de autoria de Tiago L. Moda, Káthia M. Honório e Adriano D. Andricopulo foi publicado no livro *Química Medicinal: Métodos e Fundamentos em Planejamento de Fármacos*. São Paulo: EdUSP, 2011. p. 515-537, editado por Carlos A. Montanari (Apêndice D).

Um artigo no formato comunicação foi elaborado para descrever e introduzir a base de dados PK/DB na comunidade científica internacional. O artigo mostra as propriedades farmacocinéticas presentes na base e também apresenta todas as possibilidades de busca e organização dos dados contidos nela. Essa publicação foi uma ótima oportunidade para o contato com cientistas internacionais permitindo o estabelecimento de diversas colaborações, descritas na sessão 5.4 do capítulo 5, através da predição do perfil farmacocinético de novas moléculas.

Essa comunicação intitulada “*PK/DB: database for pharmacokinetic properties and predictive in silico ADME models*” de autoria de Tiago L. Moda, Leonardo G. Torres, Alexandre E. Carrara e Adriano D. Andricopulo foi publicada no *Bioinformatics*, v. 24, n. 19, p. 2270-2271, 2008 (Apêndice E).

Referências

- 1 MUNOS, B. Lessons from 60 years of pharmaceutical innovation. *Nature Review Drug Discovery*, v. 8, n. 12, p. 959-968, 2009.
- 2 CAPRINO, L.; RUSSO, P. Developing a paradigm of drug innovation: an evaluation algorithm. *Drug Discovery Today*, v. 11, n. 11, p. 999-1006, 2006.
- 3 YAGO, G.; AMRAM, M.; MAGULA, T. Financial innovations for funding early-stage. *Drug Discovery & Development*, v.9, n. 12, p. 22-28, 2006.
- 4 RUSSO, E. Chemistry plans a structural overhaul. *Nature Jobs*, v. 419, n. 6903, p. 4-7, 2002.
- 5 JÓNSDÓTTIR, S. Ó.; JORGENSEN, F. S.; BRUNAK, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drugs candidates. *Bioinformatics*, v. 21, n. 10, 2145-2160, 2005.
- 6 ANDRICOPULO, A. D.; MONTANARI, C. A. Structure-Activity relationships for the design of small-molecule inhibitors. *Mini-Reviews in Medicinal Chemistry*, v. 5, n. 6, p. 585-593, 2005.
- 7 BERNDT, E. R.; GOTTSCHALK, A. H.; PHILIPSON, T. J.; STROBECK, M. W. Industry funding of the FDA: effects of PDUFA on approval times and withdrawal rates. *Nature Reviews Drug Discovery*, v. 4, n. 7, p. 545-554, 2005.
- 8 LAMATTINA, J. L. *Drug truths - dispelling the myths about pharma R&D*. New Jersey: Wiley, 2009. p. 24-31.
- 9 LOMBARDINO, J. G.; LOWE III, J. A. The role of the medicinal chemist in drug discovery – then and now. *Nature Reviews Drug Discovery*, v. 3, n. 10, p. 853-862, 2004.
- 10 TOLLMAN, P.; GUY, P.; ALTSHULER, J.; FLANAGAN, A.; STEINER, M. A revolution in R&D: how genomics and genetics are transforming the biopharmaceutical industry. *Boston Consulting Group*, v. 1, n. 1, p. 1-59, 2001.

- 11 NETTERWALD, J. Will the sun come out tomorrow? *Drug Discovery & Development*, v. 06, n. 12, p. 17-20, 2006.
- 12 COUZIN, J. The brains behind blockbusters. *Science*, v. 309, n. 5735, p. 728-730, 2005.
- 13 FEE, R. The cost of clinical trials. *Drug Discovery. Development*, v. 10, n. 3, p. 32, 2007.
- 14 MULLIN, R. Do-or-die time. *Chemical & Engineering News*, v. 89, n. 8, p. 12-16, 2011.
- 15 Pfizer 2010 Financial Report. Disponível em: <<http://www.pfizer.com/files/annualreport/2010/financial/financial2010.pdf/>>. Acesso em: 16 mai. 2011.
- 16 JARVIS, L, M. Pfizer reveals more R&D cuts. *Chemical & Engineering News*, v. 89, n. 6, p. 5, 2011.
- 17 JAARVIS, L, M. Pharma's next transformation. *Chemical & Engineering News*, v. 89, n. 9, p. 36-38, 2011.
- 18 MCGEE, P. Clinical trials on the move. *Drug Discovery. Development*, v. 9, n. 6, p. 16-22, 2006.
- 19 VAN DE WATERBEEMD, H.; GIFFORD, E. ADMET in silico modelling: towards prediction paradise? *Nature Reviews Drug Discovery*, v. 2, p. 192-204, 2003.
- 20 FISCHER, S. Outlook 2007: we heard an awful lot about the advent of biomarkers. *Drug Discovery & Development*, v. 9, n. 12, p. 30-32, 2006.
- 21 ASHBURN, T. T.; THOR, K. B. Identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, v. 3, p. 673-683, 2004.

- 22 May, M. Structured drug design. *Drug Discovery & Development*, v. 13, n. 6, p. 8-10, 2010.
- 23 CAMPBELL, S. L.; GAULTON, A.; MARSHALL, J.; BICHKO, D.; MARTIN, S.; BROUWER, C.; HARLAND, L. Visualizing the drug target landscape. *Drug Discovery Today*, v. 15, n. 1-2, p. 3-15, 2010.
- 24 LU, W.; LIU, X.; CAO, X.; XUE, M.; LIU, K.; ZHAO, Z.; SHEN, X.; JIANG, H.; XU, Y.; HUANG, J.; LI, H. SHAFTS: A hybrid approach for 3d molecular similarity calculation. 2. prospective case study in the discovery of diverse p90 ribosomal s6 protein kinase 2 inhibitors to suppress cell migration. *Journal of Medicinal Chemistry*, v. 51, n. 10, p. 3564-3574, 2011.
- 25 HOVDE, M.; WEINER, D.; MARIER, J. F. Do you compute? *Drug Discovery & Development*, v. 10, n. 9, p. 34-38, 2007.
- 26 MAY, M. Simulating with software. *Drug Discovery & Development*, v. 14, n. 2, p. 22-23, 2011.
- 27 BERESFORD, A. P.; SELICK, H. E.; TARBIT, M. H. The emerging importance of predictive ADME simulation in drug discovery. *Drug Discovery Today*, v. 7, n. 2, p. 109-116, 2002.
- 28 BARREIRO, E. J.; FRAGA, C. A. M. *Química medicinal: as bases moleculares da ação dos fármacos*. Porto Alegre: ArtMed Ltda, 2001.
- 29 MERVIS, J. Inside the pipeline: pharma goes to work. *Science*, v. 309, n. 5735, p. 721, 2005.
- 30 GLEESON, M. P.; HERSEY, A.; MONTANARI, D.; OVERINGTON, J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nature Reviews Drug Discovery*, v. 10, n. 3 p. 197-208, 2011.
- 31 EKINS, S. HONEYCUTT, J. D.; METZ, J. T. Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discovery Today*, v. 15, n. 11-12, p. 451-460, 2010.

- 32 KUTTLER, A.; DIMKE, T.; KERN, S.; HELMLINGER, G.; STANSKI, D.; FINELLI, L. A. Understanding pharmacokinetics using realistic computational models of fluid dynamics: biosimulation of drug distribution within CSF space for intrathecal drugs. *Journal of Pharmacokinetics and Pharmacodynamics*, v. 37, n. 6, p. 629-644, 2010.
- 33 CHEN, H.; WINIWARTER, S.; FRIDÉN, M.; ANTONSSON, M.; ENGVIST, O. In silico prediction of unbound brain-to-plasma concentration ratio machine learning algorithms. *Journal of Molecular Graphics and Modelling*, v. 29, n. 8, p. 985-995, 2011.
- 34 LEESON, P. D.; SPRINGTHORPE, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery*, v. 6, n. 11 p. 881-890, 2007.
- 35 RITCHIE, T. J.; ERTL, P.; LEWIS, R. The graphical representation of ADME-related molecule properties for medicinal chemistry. *Drug Discovery Today*, v. 16, n. 1-2, p. 65-72, 2011.
- 36 LI, A. P. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today*, v. 6, n. 7, p. 357-366, 2001.
- 37 VAN DE WATERBEEMD, H.; LENNERNAS, H.; ARTURSSON, P. *Drug bioavailability: estimation of solubility, permeability, absorption and bioavailability*. Weinheim: Wiley-VCH, 2001. (Methods and principles in medicinal chemistry).
- 38 GARDNER, C. R.; WALSH, C. T.; ALMARSSON, O. Drugs as materials: valuing physical form in drug discovery. *Nature Reviews Drug Discovery*, v. 3, n. 11, p. 926-934, 2004.
- 39 LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Discovery Reviews*, v. 46, n. 1-3, p. 3-26, 1997.
- 40 MARTIN, Y. C. A bioavailability score. *Journal of Medicinal Chemistry*, v. 48, n. 9, p. 3164-3170, 2005.

- 41 VEBER, D. F.; JOHNSON, S. R.; CHENG, H. Y.; SMITH, B. R.; WARD, K. W.; KOPPLE, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, v. 45, n. 12, p. 2615-2623, 2002.
- 42 VIETH, M.; SIEGEL, M. G.; HIGGS, R. E.; WATSON, I. A.; ROBERTSON, D. H.; SAVIN, K. A.; DURST, G. L.; HIPSKIND, P. A. Characteristic physical properties and structural fragments of marketed oral drugs. *Journal of Medicinal Chemistry*, v. 47, n. 1, p. 224-232, 2004.
- 43 VAN DE WATERBEEMD, H. Improving compound quality through in silico physicochemical profiling. *Chemistry & Biodiversity*, v. 6, n. 11, p. 1760-1766, 2009.
- 44 HANSCH, C.; LEO, A.; MEKAPATI, S. B.; KURUP, A. QSAR and ADME. *Bioorganic Medicinal Chemistry*, v. 12, n. 12, p. 3391-3400, 2004.
- 45 SMITH, D. A.; VAN DE WATERBEEMD, H.; WALKER, D. K. *Pharmacokinetics and metabolism in drug design*. Germany: Wiley-VCH, 2001. (Methods and principles in medicinal chemistry, vol. 13).
- 46 AURELI, L.; CRUCIANI, G.; MARICONI, A. Predicting human serum albumin affinity of interleukin-8 (CXCL8) inhibitors by 3D-QSPR approach. *Journal of Medicinal Chemistry*, v. 48, n. 7, p. 2469-2479, 2005.
- 47 MODA, T. L.; MONTANARI, C. A.; ANDRICOPULO, A. D. In silico prediction of human plasma protein binding using hologram QSAR. *Letters in Drug Design & Discovery*, v. 4, n. 7, p. 502-509, 2007.
- 48 SHEPPARD, G. S.; BOUSKA, J. J. Why optimize cancer drugs for ADMET? *Drug Discovery Today*, v. 2, n. 4, p. 343-349, 2005.
- 49 KRATOCHWIL, N. A.; GERBER, P. R. Predicting plasma protein binding of drugs: a new approach. *Biochemical Pharmacology*, v. 64, n. 9, p. 1355-1374, 2002.
- 50 WERMUTH, C. G. *The Practice of Medicinal Chemistry*. 2th ed. Great Britain: Academic Press, 2003.

- 51 RITSCHHELIS, W. A.; KEARNSIS, G. L. *Handbook of basic pharmacokinetics -- including clinical applications*. 7th ed. EUA: American Pharmacists Association, 2009. p. 131-132.
- 52 HARDMAN, J. G.; LIMBIRD, L. E.; GILMAN, A. G. *Goodman & Gilman's the pharmacological basis of therapeutics*. 10th ed. EUA: McGraw-Hill, 2001.
- 53 DE GRAAF, C.; VERMEULEN, N. P. E.; FEENSTRA, K. A. Cytochrome P450 in silico: an integrative modeling approach. *Journal of Medicinal Chemistry*, v. 48, n. 8, p. 2725-2755, 2005.
- 54 WILLIAMS, A. D.; LEMKE, T. L. *Foye's principles of medicinal chemistry*. 5th ed. EUA: Lippincott Williams & Wilkins, 2002.
- 55 TULSI, B. SAR revitalized by huge compound repositories. *Drug Discovery & Development*, v. 6, p. 41-44, 2003.
- 56 DE GROOT, M. J. Designing better drugs: predicting cytochrome P450 metabolism. *Drug Discovery Today*, v 11, n. 13-14 p. 601-606, 2006.
- 57 YAP, C. W.; CHEN, Y. Z. Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. *Journal of Chemical Information and Modeling*, v. 45, n. 4, p. 982-992, 2005.
- 58 STOLL, F.; GÖLLER, A. H.; HILLISCH, A. Utility of protein structures in overcoming ADMET –related issues of drug-like compounds. *Drug Discovery Today*, v. 16, n. 11-12, p. 530-538, 2011.
- 59 BU, H. Z. A literature review of enzyme kinetic parameters for CYP3A4-mediated metabolic reactions of 113 drugs in human liver microsomes: structure-kinetics relationship assessment. *Current. Drug Metabolism*, v. 7, n. 3, p. 231-249, 2006.
- 60 *Physicians' Desk Reference (PDR)*. 59th ed. EUA: Thomson, 2005.

- 61 ZARETZKI, J.; BERGERON, C.; RYDBERG, P.; HUANG, T. W.; BENNETT, K. P.; BRENNEMAN, C. M. RS-Predictor: a new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. *Journal of Chemical Information and Modeling*, 2011. doi: 10.1021/ci2000488.
- 62 ZHANG, T.; CHEN, Q.; LI, L.; LIU, L. A.; WEI, D. Q. In silico prediction of cytochrome P450-mediated drug metabolism. *Combinatorial Chemistry & High Throughput Screening*, v. 14, n. 5, p. 388-395, 2011.
- 63 TARCSAY, Á. KESERU, G. M. In silico site of metabolism prediction of cytochrome P450-mediated biotransformations. *Expert Opinion on Drug Metabolism & Toxicology*, v. 7, n. 3, p. 299-312, 2001.
- 64 ROSSATO, G.; ERNST, B.; SMIESKO, M.; SPREAFICO, M.; VEDANI, A. Probing small-molecule binding to cytochrome P450 2D6 and 2C9: an in silico protocol for generating toxicity alerts. *ChemMedChem*, v. 5 n. 12, p. 2088-2101, 2010.
- 65 BRYAN, J. Crossing the blood-brain barrier: drug delivery to the brain is still elusive. *The Pharmaceutical Journal*, v. 273, n. 10, p. 475-476, 2004.
- 66 CECHELLI, R.; BEREZOWSKI, V.; LUNDQUIST, S.; CULOT, M.; RENFTEL, M.; DEHOUCK, M. P.; FENART, L. Modelling of the blood-brain barrier in drug discovery and development. *Nature Reviews Drug Discovery*, v. 6, n. 8, p. 650-661, 2007.
- 67 PARDRIDGE W. M. Drug targeting to the brain. *Pharmaceutical Research*, v. 24, n. 9, p. 1733-1744, 2007.
- 68 HILL, A. P.; YOUNG, R. J. Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discovery Today*, v 15, n. 15-16 p. 648-655, 2010.
- 69 VAN DE WATERBEEMD, H.; SMITH, D. A.; JONES, B. C. Lipophilicity in PK design: methyl, ethyl, futile. *Journal of Computer-aided Molecular Design*, v. 15, n. 3, p. 273-286, 2001.

- 70 HUANG, J.; MA, G.; MUHAMMAD, I.; CHENG, Y. Identifying P-glycoprotein substrates using support vector machine optimized by a particle swarm. *Journal of Chemical Information and Modeling*, v. 47, n. 4, p. 1638-1647, 2007.
- 71 DAVIS, A. M.; TEAGUE, S. J.; KLEYWEGT, G. J. Application and limitation of x-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie*, v. 42, n. 24, p. 2718-2736, 2003.
- 72 LI, W. X.; LI, L.; EKSTEROWICZ, J.; LING, X. B.; CARDOZO, M. Significance analysis and multiple pharmacophore models for differentiating p-glycoprotein substrates. *Journal of Chemical Information and Modeling*, v. 47, n. 6, p. 2429-2438, 2007.
- 73 RAMACHANDRA, M.; AMBUDKAR, S. V.; CHEN, D.; HRYCYNA, C. A.; DEY, S.; GOTTESMAN, M. M.; PASTAN, I. Human P-glycoprotein exhibits reduced affinity for substrates during catalytic transition state. *Biochemistry*, v. 37, n.4, p. 5010-5019, 1998.
- 74 LAM, F. C.; LIU, R.; LU, P.; SHAPIRO, A. B.; RENOIR, J. M.; SHAROM, F. J.; REINER, P. B. beta-Amyloide efflux mediated by p-glycoprotein. *Journal of Neurochemistry*. v. 76, n. 4, p. 1121-1128, 2001.
- 75 AMBUDKAR, S. V.; KIMCHI-SARFATY, C.; SAUNA, Z. E.; GOTTESMAN, M. M. P-Glycoprotein: from genomics to mechanism. *Oncogene*, v. 22, n. 47, p. 7468-7485, 2003.
- 76 TANDON, T. L. R.; KAPOOR, B.; BANO, G.; GUPTA, S.; GILLANI, Z.; KOUR, D. P-glycoprotein: pharmacological relevance. *Indian Journal of Pharmacology*, v. 38, n. 1, p. 13-24, 2006.
- 77 EVERTS, S. A cell's vacuum cleaner. *Chemical & Engineering News*, v. 87, n. 13, p. 10, 2009.
- 78 ALLER, S. G.; CHANG, G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, v. 323, n. 5922, p. 1718-1722, 2009.

- 79 KRUIJTZER, C. M.; BEIJNEN, J. H.; ROSING, H.; TEN BOKKEL HUININK, W. W.; SCHOT, M.; JEWELL, R. C.; PAUL, E. M.; SCHELLENS, J. H. Increased oral bioavailability of topotecan in combination with the breast cancer resistance protein and P-glycoprotein inhibitor GF120918. *Journal of Clinical Oncology*, v. 20, n. 13, p. 2943-2950, 2002.
- 80 SEGALL, M.; CHAMPNESS, E.; OBREZANOVA, O.; LEEDING, C. Beyond profiling: using ADMET models to guide decisions. *Chemistry & Biodiversity*, v. 6, n. 11, p. 2144-2151, 2009.
- 81 MODI, S. Positioning ADMET in silico tools in drug discovery. *Drug Discovery Today*, v. 9, n. 1, p. 14-15, 2004.
- 82 EKINS S, WILLIAMS A. J. Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development. *Lab on a Chip*, v. 10, n. 1, p. 13-22, 2010.
- 83 JAMEI, M.; MARCINIAK, S.; FENG, K.; BARNETT, A.; TUCKER, G.; ROSTAMI-HODJEGAN, A. The Simcyp population-based ADME simulator. *Expert Opinion on Drug Metabolism & Toxicology*, v. 5, n. 2, p. 211-23, 2009.
- 84 WOMBAT-PK - Database for clinical pharmacokinetics. Disponível em: <<http://www.sunsetmolecular.com/>>. Acesso em: 20 mai. 2011.
- 85 PharmGKB - The pharmacogenetics and pharmacogenomics knowledge base. Disponível em: <<http://www.pharmgkb.org/>>. Acesso em: 20 mai. 2011.
- 86 PK/DB - Database for pharmacokinetic properties. Disponível em: <<http://www.pkdb.ifsc.usp.br/>>. Acesso em: 20 mai. 2011.
- 87 DE BENEDETTI, P. G.; FANELLI, F. Computational quantum chemistry and adaptative ligand modeling in mechanistic QSAR. *Drug Discovery Today*, v. 15, n. 859-866, p. 19-20, 2010.
- 88 MODA, T. L.; MONTANARI, C. A.; ANDRICOPULO, A. D. Hologram QSAR model for the prediction of human oral bioavailability. *Bioorganic Medicinal Chemistry*, v. 15, n. 24, p. 7738-7745, 2007.

- 89 TIAN, S.; LI, Y.; WANG, J.; ZHANG, J.; HOU, T. ADME evaluation in drug discovery. 9. prediction of oral bioavailability in humans based on molecular properties and structural fingerprints. *Molecular Pharmaceutics*, v. 8, n. 3, p. 841-851, 2011.
- 90 CHEN, L.; LI, Y.; ZHAO, Q.; PENG, H.; HOU, T. ADME evaluation in drug discovery. 10. predictions of p-glycoprotein inhibitors using recursive partitioning and naive bayesian classification techniques. *Molecular Pharmaceutics*, v. 8 n. 3, p. 889-900, 2011.
- 91 WANG, J.; KRUDY, G.; XIE, X. Q.; WU, C.; HOLLAND, G. Genetic algorithm-optimized QSPR models for bioavailability, protein binding, and urinary excretion. *Journal of Chemical Information and Modeling*, v. 46, n. 6, p. 2674-2683, 2006.
- 92 NIELSEN, P. A.; ANDERSSON, O.; HANSEN, S. H.; SIMONSEN, K. B.; ANDERSSON, G. Models for predicting blood-brain barrier permeation. *Drug Discovery Today*, v. 16, n. 11-12, p. 472-475, 2011.
- 93 PAINE, S. W. A rapid computational filter for predicting the rate of human renal clearance. *Journal of Molecular Graphics & Modelling*, v. 29, n. 4, p. 529-37, 2010.
- 94 ABRAHAM, D. J. *Burger's medicinal chemistry and drug discovery*. 6th ed. EUA: John & Sons Inc, 2003. v.1.
- 95 GASTEIGER, J. The central role of chemoinformatics. *Chemometrics and Intelligent Laboratory Systems*, v. 82, n. 1-2, p. 200-209, 2006.
- 96 HANSCH, C.; KURUP, A.; GARG, R.; GAO, H. Chem-Bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chemical Reviews*, v. 101, n. 3, p. 619-672, 2001.
- 97 GLEESON, M. P.; HERSEY, A.; HANNONGBUA, S. In-silico ADME models: a general assessment of their utility in drug discovery applications. *Current Topics in Medicinal Chemistry*, v. 11, n. 4, p. 358-381, 2011.

- 98 FOURCHES, D.; MURATOV, E.; TROPSHA A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, v. 50, n. 7, p. 1189-1204, 2010.
- 99 PALM, K.; STENBERG, P.; LUTHMAN, K.; ARTURSSON, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharmaceutical Research*, v. 14, n. 5, p. 568-571, 1997.
- 100 WESSEL, M. D.; JURIS, P. C.; TOLAN, J. W.; MUSKAL, S. M. Prediction of human intestinal absorption of drug compounds from molecular structure. *Journal of Chemical Information and Computer Sciences*, v. 38, n. 4, p. 726-735, 1998.
- 101 KANSY, M.; SENNER, F.; GUBERNATOR, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *Journal of Medicinal Chemistry*, v. 41, n. 7, p. 1007-1010, 1998.
- 102 OPREA, T.; GOTTFRIES, J. Toward minimalistic modeling of oral drug absorption. *Journal of Molecular Graphics & Modelling*, v. 17, n. 5-6, p. 261-274, 1999.
- 103 ABRAHAM, M. H.; ZHAO, Y. H.; LE, J.; HERSEY, A.; LUSCOMBE, C. N.; REYNOLDS, D. P.; BECK, G.; SHERBORNE, B.; COOPER, I. On the mechanism human intestinal absorption. *European Journal of Medicinal Chemistry*, v. 37, n. 7, p. 595-605, 2002.
- 104 KLOPMAN, G.; STEFAN, L. R.; SAIKHOV, R. D. ADME evaluation 2. a computer model for prediction of intestinal absorption in humans. *European Journal of Pharmaceutical Sciences*, v. 17, n. 4-5, p. 253-263, 2002.
- 105 ZHAO, Y. H.; ABRAHAM, M. H.; LE, J.; HERSEY, A.; LUSCOMBE, C. N.; BECK, G.; SHERBORNE, B.; COOPER, I. Evaluation of rat intestinal absorption data and correlation with human intestinal absorption. *European Journal of Medicinal Chemistry*, v. 38, n. 3, p. 233-243, 2003.

- 106 HOU, T.; WANG, J.; ZHANG, W.; XU, X. ADME evaluation in drug discovery. 7. prediction of oral absorption by correlation and classification. *Journal of Chemical Information and Modeling*, v. 47, n. 1, p. 208-218, 2007.
- 107 HOU, T. J.; XU, X. J. ADME evaluation in drug discovery. 3. modeling blood-brain barrier partitioning using simple molecular descriptors. *Journal of Chemical Information and Computer Sciences*, v. 43, n. 6, p. 2137-2152, 2003.
- 108 AL-FAHEMI, J. H. A.; COOPER, D. L.; ALLAN, N. L. Investigating the utility of momentum-space descriptors for predicting blood-brain barrier penetration. *Journal of Molecular Graphics & Modelling*, v. 26, n. 3, p. 607-612, 2007.
- 109 WINKLER, D. A.; BURDEN, F. R. Modelling blood-brain barrier partitioning using. *Journal of Molecular Graphics & Modelling*, v. 22, n. 6, p. 499-505, 2004.
- 110 PLATTS, J. A.; ABRAHAM, M. H.; ZHAO, Y. H.; HERSEY, A.; IJAZ, L.; BUTINA, D. Correlation and prediction of a large blood-brain barrier distribution data set – an LFER study. *European Journal of Medicinal Chemistry*, v. 36, n. 9, p. 719-730, 2001.
- 111 SUBRAMANIAN, G.; KITCHEN, D. B. Computational models to predict blood-brain barrier permeation and CNS activity. *Journal of Computer Aided Molecular Design*, v. 17, n. 10, p. 643-664, 2003.
- 112 HUTTER, M. C. Prediction of blood-brain barrier permeation using quantum chemically derived information. *Journal of Computer Aided Molecular Design*, v. 17, n. 7, p. 415-433, 2003.
- 113 LOMBARDO, F.; BLAKE, J. F.; CURATOLO, W. J. Computation of brain-blood partitioning of organic solutes via free energy calculations. *Journal of Medicinal Chemistry*, v. 39, n. 24, p. 4750-4755, 1996.
- 114 ROSE, K.; HALL, L. H. Modeling blood-brain barrier partitioning using the electrotopological state. *Journal of Chemical Information and Computer Sciences*, v. 42, n. 3, p. 651-666, 2002.

- 115 NORINDER, U.; SJÖBERG, P.; ÖSTERBERG, T. Theoretical calculation and prediction of brain-blood partitioning of organic solutes using MolSurf parametrization and PLS statistics. *Journal of Pharmaceutical Science*, v. 87, n. 8, p. 952-959, 1998.
- 116 PAN, D.; IYER, M.; LIU, J.; LI, Y.; HOPFINGER, A. J. Constructing optimum blood brain barrier QSAR models using a combination of 4D-molecular similarity measures and cluster analysis. *Journal of Chemical Information and Computer Sciences*, v. 44, n. 6, p. 2083-2098, 2004.
- 117 LESSIGIARSKA, I.; PAJEVA, I.; CRONIN, M. T.; WORTH, A. P. 3D QSAR investigation of the blood-brain barrier penetration of chemical compounds. *SAR and QSAR in Environmental Research*, v. 16, n. 1-2, p. 79-91, 2005.
- 118 KAISER, D.; TERFLOTH, L.; KOPP, S.; SCHULZ, J.; DE LAET, R.; CHIBA, P.; ECKER, G. F.; GASTEIGER, J. Self-Organizing maps for identification of new inhibitors of p-glycoprotein. *Journal of Medicinal Chemistry*, v. 50, n. 7, p. 1698-1702, 2007.
- 119 COPELAND, R. A. *Evaluation of enzyme inhibition in drug discovery: a guide for medicinal chemistry and pharmacologists*. EUA: Wiley, 2005.
- 120 WANG, J.; KRUDY, G.; HOU, T.; ZHANG, W.; HOLLAND, G.; XU, X. Development of reliable aqueous solubility models and their application in druglike analysis. *Journal of Chemical Information and Modeling*, v. 47, n. 4, p. 1395-1404, 2007.
- 121 HQSARTM, *User's Manual*. St. Louis: Tripos Inc., 2006.
- 122 LOWIS, D. R. *HQSAR: a new, Highly Predictive QSAR Technique*. Tripos Technical Notes, 1, 1997.
- 123 CRIVORI, P.; CRUCIANI, G.; CARRUPT, P. A.; TESTA, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *Journal of Medicinal Chemistry*, v. 43, n. 11, p. 2204-2216, 2000.

- 124 CRUCIANI, G.; CRIVORI, P.; CARRUPT, P. A.; TESTA, B. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *-Journal Molecular Structure-theochem*, v. 503, n. 1-2, p. 17-30, 2000.
- 125 BORCHHARDT, D. M.; CASTILHO, M. S.; ANDRICOPULO, A. D. Classical and fragment-based hologram structure-activity relationships for a series of analgesic cyclic imides. *Letter in Drug Design & Discovery*, v. 5, n. 1, p. 57-64, 2008.
- 126 MODA, T. L.; TORRES, L. G.; CARRARA, A. E.; ANDRICOPULO, A. D. PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. *Bioinformatics* v. 24, n. 19, p. 2270-2271, 2008.
- 127 Agência FAPESP - No encaço das moléculas. Disponível em: <<http://agencia.fapesp.br/8951/>>. Acesso em: 3 jul. 2011.
- 128 Jornal da Ciência - No encaço das moléculas. Disponível em: <<http://www.jornaldaciencia.org.br/Detalhe.jsp?id=56605/>>. Acesso em: 3 jul. 2011.
- 129 InovaBrasil - Lançada a primeira base de dados brasileira de propriedades farmacocinéticas. Disponível em: <<http://inovabrasil.blogspot.com/2008/06/lanada-primeira-base-de-dados.html/>>. Acesso em: 3 jul. 2011.
- 130 ARMBRUSTER, B. N.; ROTH, B. L. Mining the receptorome. *The Journal of Biological Chemistry*, v. 280, n. 7, p. 5129-5132, 2005.
- 131 BERGER, M.; GRAY, J. A.; ROTH, B. L. The Expanded biology of serotonin. *Annual Review of Medicine*, v. 60, p. 355-366, 2009.
- 132 I.H.S. – International Headache Society. Disponível em: <http://www.i-h-s.org/frame_non_members.asp/>. Acesso em: 4 jul. 2011.
- 133 GOADSBY, P. J. Can we develop neurally acting drugs for the treatment of migraine? *Nature Reviews Drug Discovery*, v. 4, n. 9, p. 741-750, 2005.

- 134 GOADSBY, P. J.; CLASSEY, J. D. Evidence for serotonin (5-HT)_{1B}, 5-HT_{1D} and 5-HT_{1F} receptor inhibitory effects on trigeminal neurons with craniovascular input. *Neuroscience*, v. 122, n. 2 p. 491-498, 2003.
- 135 NELSON, D. L.; PHEBUS, L. A.; JOHNSON, K. W.; WAINSCOTT, D. B.; COHEN, M. L.; CALLIGARO, D. O.; XU, Y-C. Preclinical pharmacological profile of the selective 5-HT_{1F} receptor agonist lasmiditan. *Cephalalgia*, v. 30, n. 10, p. 1159-1169, 2010.
- 136 FERRARI, M. D.; FARKKILA, M.; REUTER, U.; PILGRIM, A.; DAVIS, C.; KRAUSS, M.; DIENER, H-C. Acute treatment of migraine with the selective 5-HT_{1F} receptor agonist lasmiditan – A randomized proof-of-concept trial. *Cephalalgia*, v. 30, n. 10, p. 1170-1178, 2010.
- 137 PDSP - Psychoactive Drug Screening Program. Disponível em: <<http://pdsp.med.unc.edu/>>. Acesso em: 4 jul. 2011.
- 138 ChEMBL - Database of bioactive drug-like small molecules. Disponível em: <<https://www.ebi.ac.uk/chembl/db/>>. Acesso em: 4 jul. 2011.
- 139 Wombat - World of Molecular Bioactivity. Disponível em: <<http://www.sunsetmolecular.com/>>. Acesso em: 4 jul. 2011.
- 140 HAJJO, R.; GRULKE, C. M.; GOLBRAIKH, A.; SETOLA, V.; HUANG, X-P.; ROTH, B. L.; TROPSHA, A. Development, validation, and use of quantitative structure-activity relationship models of 5-hydroxytryptamine (2b) receptor ligands to identify novel receptor binders and putative valvulopathic compounds among common drugs. *Journal of Medicinal Chemistry*, v. 53, n. 21, p. 7573-7586, 2010.
- 141 MOE - Molecular Operating Environment. Disponível em: <<http://www.chemcomp.com/>>. Acesso em: 4 jul. 2011.
- 142 Talete SRL. - DRAGON for descriptor calculation. Disponível em: <<http://www.talete.mi.it/>>. Acesso em: 4 jul. 2011.

- 143 FOURCHES, D.; BARNES, J. C.; DAY, N. C.; BRADLEY, P.; REED, J. Z.; TROPSHA, A. Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. *Chemical Research in Toxicology*, v. 23, n. 1, p. 171-83, 2010.
- 144 HSIEH, J. H.; WANG, X. S.; TEOTICO, D.; GOLBRAIKH, A.; TROPSHA, A. Differentiation of AmpC beta-lactamase binders vs. decoys using classification kNN QSAR modeling and application of the QSAR classifier to virtual screening. *Journal of Computer-aided Molecular Design*, v. 22, n. 9, p.593-609, 2008.
- 145 LIVINGSTONE, D. *Data analysis for chemists: applications to QSAR and chemical product design*. Oxford: Oxford University Press, 1995. p. 90-95.
- 146 CRISTIANINI, N.; TAYLOR, J. S. *An introduction to support vector machines: and other kernel-based learning methods*. Cambridge: Cambridge University Press, 2000.
- 147 TROPSHA, A.; GOLBRAIKH, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Current Pharmaceutical Design*, v. 13, n. 34, p. 3494-3504, 2007.
- 148 ZHENG, W. F.; TROPSHA, A. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. *Journal of Chemical Information and Computer Sciences*, v. 40, n. 1, p. 185-194, 2000.
- 149 TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, v. 22, n. 1, p. 69-77, 2003.
- 150 WOLD, S.; ERIKSSON, L.; CLEMENTI, S. *Statistical validation of QSAR results in chemometrics methods in molecular design*. Weinheim: VCH, 1995. p. 309-318.
- 151 Maybridge - Screening collection. Disponível em: <http://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__146/DesktopDefault.aspx/>. Acesso em: 3 jul. 2011.

- 152 Asinex - Gold and platinum collections. Disponível em: <<http://www.asinex.com/library-gold.html/>>. Acesso em: 3 jul. 2011.
- 153 TimTec - MyriaScreen diversity collection. Disponível em: <<http://timtec.net/myriascreen-diversity-collection.html/>>. Acesso em: 3 jul. 2011.
- 154 Prestwick - Prestwick chemical library. Disponível em: <<http://www.prestwickchemical.com/index.php?pa=26/>>. Acesso em: 3 jul. 2011.
- 155 PDSP - Assay protocol book. Disponível em: <<http://pdsp.med.unc.edu/UNC-CHProtocolBook.pdf/>>. Acesso em: 3 jul. 2011.

Apêndices

APÊNDICE A

Pharmacokinetic Properties:

An Old Problem with New Perspectives

Káthia M. Honório^{1,2}, Tiago Moda³ and Adriano D. Andricopulo^{3*}

¹Universidade Federal do ABC, Santo André-SP, Brazil; ²Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Av. Arlindo Bettio 1000, 03828-000, São Paulo-SP, Brazil; ³Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil

*Address correspondence to this author at the Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil; Tel: + 55 16 3373-8095; Fax: +55 16 3373-9881; E-mail: aandrico@if.sc.usp.br

Abstract: The discovery and development of a new drug is a lengthy, complex and expensive process. This complex process has evolved from classical methods into an integration of modern technologies and innovative strategies addressed to the design of new chemical entities for a variety of biological targets. The combination of available knowledge of several 3D protein structures with hundreds of thousands of small-molecule drug candidates have attracted the attention of scientists from all over the world for the application of structure- and ligand-based drug design approaches. The investigation of new potential biological targets is often limited by the absence of initial compounds that bind to a specific site in the protein, possessing reasonable chemical and biological properties for further lead optimization. Biological screening requires testing a large number of compounds in standard models to assess potency, selectivity and affinity. In this context, it is very important to study the pharmacokinetic profile of the compounds under investigation. Recent advances have been made in the collection of data and development of models to assess and predict pharmacokinetic properties (absorption, distribution, metabolism, and excretion – ADME) in the early stages of drug discovery projects. This review provides a brief perspective on the evolution and current status of *in silico* ADME tools, highlighting the present challenges, limitations and new opportunities in medicinal chemistry.

Keywords: Medicinal chemistry, drug design, pharmacokinetics, ADME.

APÊNDICE B

Consensus Hologram QSAR Modeling for the Prediction of Human Intestinal Absorption

Tiago Moda¹ and Adriano D. Andricopulo^{1*}

¹Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil

*Address correspondence to this author at the Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil; Tel: + 55 16 3373-8095; Fax: +55 16 3373-9881; E-mail: aandrico@if.sc.usp.br

Abstract: Over the past decades, advances in combinatorial chemistry and high-throughput screening have resulted in a dramatic shift in the drug discovery landscape. Common use of sophisticated automation and other new technologies have enabled large libraries of compounds to be synthesized and screened for biologic activity in a rapid and efficient manner. Thus, modern approaches to previous lead identification of drug-like properties, such as human intestinal absorption are required. A diversity of *in silico* pharmacokinetics models have been developed since ADME (absorption, distribution, metabolism and excretion) properties were identified as one of the major problem in clinical trials. In the present work, hologram quantitative structure – activity relationships (HQSAR) were performed on a training set of 638 structurally diverse molecules with known human intestinal absorption. The most significant HQSAR models (q^2 ranging from 0.63 to 0.71) were selected for posterior validation. The predictive ability of the models were evaluated by an external test set containing 128 molecules not included in the training set, and the predicted values using the consensus score were in good agreement with the experimental values. The consensus HQSAR models should be useful for the selection of leads having acceptable human intestinal absorption by virtual screening of large libraries of compounds.

Keywords: Medicinal chemistry, human intestinal absorption, HQSAR, ADME.

APÊNDICE C

Parallel Screening for Novel Antimigraine Therapeutics that Act on 5-HT_{1B}, 5-HT_{1D} and 5-HT_{1F} Serotonin Receptors using QSAR Models and Experimental Validation

Tiago L. Moda^{1,2}, Xiang S. Wang^{*,3}, Adriano D. Andricopulo²,
Tudor I. Oprea⁴, Alexander Tropsha^{*,1}

¹Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products and Carolina Exploratory Center for Cheminformatics Research, School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina, 27599, USA.

²Laboratory of Computational and Medicinal Chemistry, Physics Institute of Sao Carlos, University of Sao Paulo, Sao Carlos, Sao Paulo, 13560, Brazil; ³Laboratory of Cheminformatics and Drug Design, Department of Pharmaceutical Sciences, School of Pharmacy, Howard University, Washington, District of Columbia 20059, USA; ⁴Departments of Biochemistry and Molecular Biology and Cell Biology and Physiology, School of Medicine, University of New Mexico, Albuquerque, New Mexico, USA.

*Correspondence should be addressed to X.S.W. (xiang.wang@howard.edu) or A.T. (alex_tropsha@unc.edu), CB #7360, Beard Hall, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC27599-7360, Tel. 919-966-2955, Fax: 919-966-0204

Abstract: Migraine is a recurring, episodic neurological disorder characterized as unilateral, throbbing headache. The only class of marketed drugs to treat migraine specifically is a group of triptans, e.g., sumatriptan and naratriptan, but they are not effective in all patients. Recent findings indicate that compounds that can simultaneously target a triplet of serotonin 5-HT receptors, i.e., 5-HT_{1B}, 5-HT_{1D}, and 5HT_{1F}, may be very effective and safe medications. We have developed multiple and rigorously validated quantitative structure-activity relationship (QSAR) models for 531 structurally diverse molecules with known affinity for 5-HT_{1B}, 5-HT_{1D} or 5-HT_{1F} using variable selection *k* nearest neighbor (*k*NN) and support vector machines (SVM). These in silico models were then used to mine in parallel over 281 000 chemicals, yielding 132 diverse computational hits. Ten of them were tested in radioligand binding assays and two of them were found active for 5-HT_{1B} and 5-HT_{1D} simultaneously. All active compounds were then tested in functional assays, and all compounds were identified as true 5-HT_{1B} and 5-HT_{1D} agonist. Results shown that the QSAR models developed in this study allied to parallel virtual screening could be used as reliable tools to improve the chance of finding candidates that are likely potential atimigraine.

Keywords: Migraine, serotonin receptor, *k*NN, SVM, virtual screening.

APÊNDICE D

15

PROPRIEDADES FARMACOCINÉTICAS E O
PLANEJAMENTO DE NOVAS SUBSTÂNCIAS BIOATIVASTIAGO L. MODA
KÁTHIA M. HONÓRIO
ADRIANO D. ANDRICOPULO

15.1 INTRODUÇÃO

A indústria farmacêutica vem passando por importantes transformações decorrentes sobretudo dos notáveis avanços científicos e tecnológicos fundamentados em novos paradigmas de natureza intrinsecamente multidisciplinar. A descoberta de novos fármacos de diversas classes terapêuticas tem proporcionado melhorias significativas na qualidade de vida das diversas populações no mundo¹. O papel relevante da química medicinal nos processos inovativos da gênese planejada de fármacos é preservado ao longo dos anos nas várias interfaces e conexões entre química e biologia. Em particular, os métodos em química e bioinformática têm sido amplamente empregados no processo de descoberta de novos fármacos, encontrando grande aplicação nas etapas de pesquisa básica, incluindo a identificação, seleção e otimização de moléculas candidatas a novas entidades químicas (do inglês, new chemical entities – NCEs)². A integração de métodos computacionais avançados ao trabalho químico e biológico experimental é essencial para a geração de

1. L. Caprino e P. Russo, "Developing a Paradigm of Drug Innovation: An Evaluation Algorithm", *Drug Discov. Today*, 11: 999-1006, 2006.

G. Yago, M. Amram e T. Magula, "Financial Innovations for Funding Early-Stage", *Drug Discov. Dev.*, 12: 22-28, 2006.

2. E. Russo, "Chemistry Plans a Structural Overhaul", *Nature*, 419: 4-7, 2002.

15.7.6 BARREIRA HEMATOENCEFÁLICA

Modelos de permeabilidade da barreira hematoencefálica (BBB) foram desenvolvidos baseados em dezenove descritores moleculares calculados a partir do programa Algorithm Builder e uma matriz de fragmentos moleculares⁸⁸. Os modelos foram empregados na avaliação das propriedades de 1.593 compostos e os resultados atribuíram maior importância ao descritor de ligação de hidrogênio no processo de permeabilidade da BBB.

Baseado no método de redes neurais artificiais (do inglês, artificial neural network – ANN) um modelo preditivo de BBB foi obtido empregando-se parâmetros moleculares e estruturais, e considerando o transporte ativo mediado pela P-gp no efluxo de certas moléculas. Um total de 191 compostos foi empregado neste estudo, sendo que o peso molecular e a TPSA mostraram maior importância, entre outros descritores, na predição da BBB. O modelo final apresenta boa utilidade na avaliação de compostos que possam atuar no sistema nervoso central (SNC)⁸⁹.

15.8 CONCLUSÕES E PERSPECTIVAS

O estudo de propriedades farmacocinéticas tem enorme valor no processo de P&D de fármacos. O desenvolvimento de modelos preditivos *in silico* de propriedades de ADME ganhou maior importância a partir da década de 1990 e as perspectivas apontam para um futuro ainda mais promissor. A aplicação em análises de larga escala é outra vantagem clara das técnicas *in silico*, em comparação com métodos *in vitro* e *in vivo*, empregados na determinação de propriedades de ADME. O gerenciamento da informação é outro componente extremamente vantajoso das técnicas *in silico*, especialmente no tratamento de grandes coleções de compostos e na organização de bases de dados dirigidas.

O grande desafio que se traduz no principal benefício dos métodos *in silico* está em sua capacidade de prever propriedades farmacocinéticas de compostos ainda não sintetizados, auxiliando intensamente o planejamento de NCEs com propriedades adequadas de ADME. O acúmulo de informações experimentais (*in vitro* e *in vivo*) tem contribuído para a criação de bases de dados mais robustas para os estudos *in silico*, favorecendo assim, o desenvolvimento de modelos baseados em estratégias diferentes⁹⁰.

AGRADECIMENTOS

Os autores agradecem a Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) e CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo suporte ao desenvolvimento das pesquisas do grupo.

88. Y. H. Zhao, M. H. Abraham, A. Ibrahim, P. V. Fish, S. Cole, M. L. Lewis, M. J. Groot e D. P. Reynolds, "Predicting Penetration across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes", *J. Chem. Inf. Model.*, 47: 170-175, 2007.

89. P. Garg e J. Verma, "In silico Prediction of Blood-Brain Barrier Permeability: An Artificial Neural Network Model", *J. Chem. Inf. Model.*, 46: 289-297, 2006.

90. I. V. Tetko, P. Bruneau, H. W. Mewes, D. C. Rohrer e G. I. Poda, "Can We Estimate the Accuracy of ADME/Tox Predictions?", *Drug Discov. Today*, 11: 700-707, 2006.

Databases and ontologies

PK/DB: database for pharmacokinetic properties and predictive *in silico* ADME models

Tiago L. Moda, Leonardo G. Torres, Alexandre E. Carrara and Adriano D. Andricopulo*

Laboratory of Computational and Medicinal Chemistry, Center for Structural Molecular Biotechnology, Institute of Physics of São Carlos, University of São Paulo, São Carlos-SP, 13566-970, Brazil

Received on May 6, 2008; revised on August 1, 2008; accepted on August 3, 2008

Advance Access publication August 6, 2008

Associate Editor: Jonathan Wren

ABSTRACT

Summary: The study of pharmacokinetic properties (PK) is of great importance in drug discovery and development. In the present work, PK/DB (a new freely available database for PK) was designed with the aim of creating robust databases for pharmacokinetic studies and *in silico* absorption, distribution, metabolism and excretion (ADME) prediction. Comprehensive, web-based and easy to access, PK/DB manages 1203 compounds which represent 2973 pharmacokinetic measurements, including five models for *in silico* ADME prediction (human intestinal absorption, human oral bioavailability, plasma protein binding, blood–brain barrier and water solubility).

Availability: <http://www.pkdb.ifsc.usp.br>

Contact: aandrico@ifsc.usp.br

1 INTRODUCTION

The challenges facing the pharmaceutical industry are tremendous at every step of the drug discovery and development process. Technology-based discovery is certainly an important element to increase R&D productivity (Guido *et al.*, 2008; Jónsdóttir *et al.*, 2005). A drug intended for use in humans should have an ideal balance of efficacy and safety, as well as good pharmacokinetic properties (PK) (Moda *et al.*, 2007a). Problems with drug candidates' absorption, distribution, metabolism and excretion (ADME), however, have been identified as a major cause of drug candidate failure in late stages of the drug development process. Therefore, it is critical to accurately predict these qualities earlier in the investigation of lead candidates (Moda *et al.*, 2007a; Norinder and Bergström, 2006). Computational methods have emerged as a powerful strategy for the prediction of human PK. In this regard, a variety of useful *in silico* ADME models have been developed with different levels of complexity for the screening of large data sets of compounds, creating tools that are faster, simpler and more cost-effective than traditional experimental procedures (Canavan, 2007).

In an effort to make high quality pharmacokinetic data and predictive models available to a worldwide scientific community, PK/DB (a freely available database for PK) was designed by our research group incorporating high quality databases of structurally diverse drug-like and lead-like molecules for a variety of PK.

2 DATABASE CONTENT

The chemical and pharmacokinetic data were collected both from public databases and from the literature (http://www.pkdb.ifsc.usp.br/pkdb/literature_src.php), resulting in a total of 1203 compounds with 2973 property values grouped and organized as shown in Table 1.

3 INTERFACE AND DATA MANAGEMENT

In PK/DB, a web-based query tool incorporating a molecular drawing interface enables the database to be searched by chemical structure or standard name, substructure or molecular fragment, molecular formula or by an exact or range of a specific pharmacokinetic property (Table 1). Also available for searching is the information on human CYP-mediated drug metabolism for a number of compounds. The user can also employ a combination of criteria as a useful way for database searching. To facilitate the analysis, the results of such searches are showed in two steps. In the first step, the user can choose the number of compounds at the top of the search session (e.g. 10, 25, 50, 100 or all) and sort it by molecular weight, compound name, HIA, F, PPB, BBB, Vd, Cl and T1/2. The column field presents the compounds, showing the PK/DB identification (MID), 2D structure, standard name, SMILES, molecular weight and a list of properties available in PK/DB, as shown in Figure 1. The second step allows access to more detailed information using some hyperlinks: 3D structures, PK and pharmacological action extracted from PubChem (<http://pubchem.ncbi.nlm.nih.gov/>).

Table 1. Number of measurements in PK/DB for a number of pharmacokinetic properties

Acronym	Property	Measurements
HIA	Human intestinal absorption	677
F	Human oral bioavailability	660
PPB	Plasma protein binding	440
BBB	Blood Brain Barrier	200
Vd	Volume of distribution	291
Cl	Renal clearance	360
T1/2	Half-life	355

*To whom correspondence should be addressed.

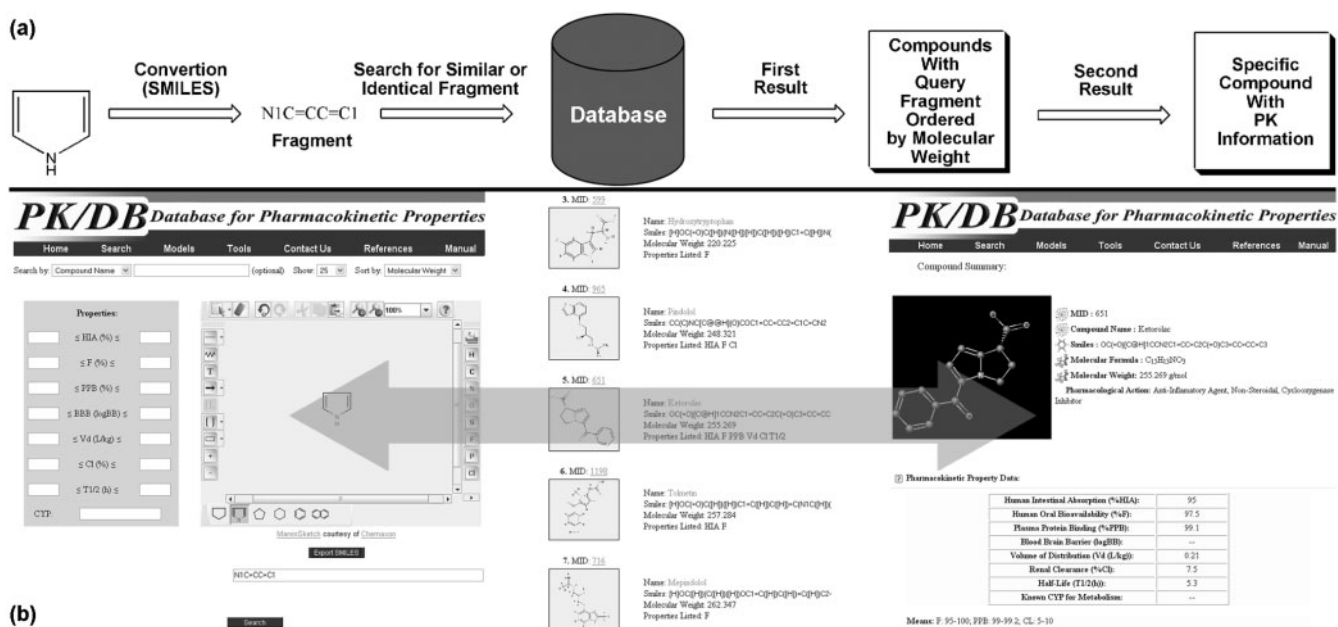


Fig. 1. PK/DB allows searching of PK for a large number of compounds using an easy to access interface. (a) Overview of the PK/DB system architecture. (b) This screenshot shows the results of a search based on a specific molecular fragment.

4 IN SILICO PREDICTIVE ADME MODELS

PK/DB presents five *in silico* predictive models available for the evaluation of ADME properties, including human oral bioavailability, plasma protein binding, human intestinal absorption, blood–brain barrier permeation and water solubility. These predictive models are statistically robust and have both good internal and external consistency. In addition, the models have been validated by external test sets of compounds which were not considered for model generation. The predictive models were developed by our research group based on specialized molecular fragments using the hologram quantitative structure–activity relationships (HQSAR) method (Castilho *et al.*, 2007; Moda *et al.*, 2007a, b).

5 IMPLEMENTATION

An integrated data structure and a variety of querying logics were developed to allow easy and efficient retrieval of pharmacokinetic data. The PK/DB system is installed on Red Hat Enterprise Linux workstations and employs the PostgreSQL v8.2 (<http://www.postgresql.org>) as a relational database management system (RDBMS) and Apache 2.0 (<http://www.apache.org>) server as a web server platform. Its web interface is implemented using PHP (<http://www.php.net>), Javascript and DHTML. For the flexible integration of the information present in PK/DB, the server interface was implemented using PHP and the search logic implemented in C++ using the OEChem TK (OpenEye Inc., Santa Fe, NM, USA), a programming library of functions that properly handle the details of working with molecules. Towards more user-friendly searching and retrieval systems, PK/DB provides interactive web interfaces, including the graphic structure editor MarvinSketch applet (ChemAxon Inc., Budapest, Hungary).

6 CONCLUSIONS

The use of computational models in the prediction of PK of active compounds is growing rapidly in drug discovery due to the benefits they provide in throughput and early application in the design of new drug candidates. PK/DB is a new database that provides useful information on a variety of important PK, as well as access to predictive *in silico* ADME models. The PK/DB suite is designed to be utilized by all researchers in the drug discovery field, and will be continuously updated and upgraded as new information becomes available.

Funding: FAPESP (The State of São Paulo Research Foundation); CNPq (The National Council for Scientific and Technological Development), Brazil.

Conflict of Interest: none declared.

REFERENCES

- Canavan, N. (2007) FDA and drug companies alike want ADME-tox testing performed earlier and earlier in a drug's life cycle. *Drug Discov. Dev.*, **10**, 34–36.
- Castilho, M.S. *et al.* (2007) Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. *Lett. Drug Des. Discov.*, **4**, 106–113.
- Guido, R.V.C. *et al.* (2008) Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, **15**, 37–46.
- Jónsdóttir, S.O. *et al.* (2005) Prediction methods and databases within chemoinformatics: emphasis on drug and drug candidates. *Bioinformatics*, **21**, 2145–2160.
- Moda, T.L. *et al.* (2007a) Hologram QSAR model for the prediction of human oral bioavailability. *Bioorg. Med. Chem.*, **15**, 7738–7745.
- Moda, T.L. *et al.* (2007b) *In silico* prediction of human plasma protein binding using hologram QSAR. *Lett. Drug. Des. Discov.*, **4**, 502–509.
- Norinder, U. and Bergström, C.A.S. (2006) Prediction of ADMET properties. *ChemMedChem.*, **1**, 920–937.



Hologram QSAR model for the prediction of human oral bioavailability

Tiago L. Moda,^a Carlos A. Montanari^b and Adriano D. Andricopulo^{a,*}

^aLaboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, 13566-970 São Carlos, SP, Brazil

^bGrupo de Química Medicinal de Produtos Naturais, Instituto de Química de São Carlos, Universidade de São Paulo, 13566-970 São Carlos, SP, Brazil

Received 2 June 2007; revised 22 August 2007; accepted 28 August 2007
Available online 1 September 2007

Abstract—A drug intended for use in humans should have an ideal balance of pharmacokinetics and safety, as well as potency and selectivity. Unfavorable pharmacokinetics can negatively affect the clinical development of many otherwise promising drug candidates. A variety of *in silico* ADME (absorption, distribution, metabolism, and excretion) models are receiving increased attention due to a better appreciation that pharmacokinetic properties should be considered in early phases of the drug discovery process. Human oral bioavailability is an important pharmacokinetic property, which is directly related to the amount of drug available in the systemic circulation to exert pharmacological and therapeutic effects. In the present work, hologram quantitative structure–activity relationships (HQSAR) were performed on a training set of 250 structurally diverse molecules with known human oral bioavailability. The most significant HQSAR model ($q^2 = 0.70$, $r^2 = 0.93$) was obtained using atoms, bond, connection, and chirality as fragment distinction. The predictive ability of the model was evaluated by an external test set containing 52 molecules not included in the training set, and the predicted values were in good agreement with the experimental values. The HQSAR model should be useful for the design of new drug candidates having increased bioavailability as well as in the process of chemical library design, virtual screening, and high-throughput screening.

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

The challenges facing the pharmaceutical industry are tremendous at every step of the drug discovery and development process. The high number of compounds emerging from combinatorial chemistry and high-throughput medicinal chemistry programs is increasing the demand for new compounds that need to be screened in a wide range of biological assays.¹ The classical strategy to screen thousands of compounds solely for potency and selectivity brought the pharmaceutical industry to face the reality of disproportionate attrition in advanced stages of clinical development, because many attractive compounds do not possess the required pharmacokinetic properties of a drug. New chemical entities (NCEs) expected to advance into clinical trials should have an ideal balance of pharmacodynamic and

pharmacokinetic properties.² Problems with absorption, distribution, metabolism, and excretion (ADME) have been identified as a major cause of drug candidate failure in late stages of the pharmaceutical R&D process.^{3,4}

The early evaluation of ADME properties in drug research has driven the need for large-scale screening methods. *In vitro* and *in vivo* ADME assays (e.g., Caco-2, PAMPA, MDCK, assessment of absorption, metabolism studies) are lengthy, complex, and relatively expensive in terms of resources, reagents, and detection techniques.⁵ Computational methods have emerged during the past decade as a powerful strategy for the prediction of human pharmacokinetics.⁶ In this regard, a variety of useful *in silico* ADME models has been developed with different levels of complexity for the screening of large data sets of compounds, creating tools that are faster, simpler, and more cost-effective than traditional experimental procedures.^{6–9}

Quantitative structure–activity relationships (QSAR) employing both classical and modern technologies have

Keywords: ADME; Pharmacokinetics; *In silico* modeling; QSAR; Drug design.

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881; e-mail: aandrico@if.sc.usp.br

proven useful in a large number of settings.^{10,11} QSAR is a technology that generates descriptors based on molecular structures and uses computational algorithms to relate the key descriptors to the dependent property value of interest.¹² It is therefore possible to match these unique characteristics to pharmacokinetic properties, mainly due to the accumulation of suitable ADME data for model generation.^{13,14} Human oral bioavailability, which is defined as the fraction of an administered dose of drug that reaches the systemic circulation, is a critical property to be considered during the early stages of discovery. Several reports in the literature indicate that there is a significant scientific and practical need for new tools for early prediction of oral bioavailability as well as other important pharmacokinetic properties.^{15,16} In the present work, we have organized a data set of 302 structurally diverse molecules with known human oral bioavailability, and used the data to create predictive 2D QSAR models, employing the hologram QSAR (HQSAR) method.^{17–19} The results of modeling this data set are reported herein.

2. Materials and methods

2.1. Computational approach

The QSAR analyses, calculations, and visualizations for HQSAR were performed using the SYBYL 7.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Enterprise Linux workstations. A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) employing the complete linkage clustering method with no standardization as previously described.^{18,19}

2.2. Data set

The data set of 302 structurally diverse molecules used in the QSAR analyses was collected from the literature.^{20–22} For convenience, the list of compounds along with the corresponding human oral bioavailability data are shown in Table 3. The 3D structures of the molecules employed in this work were constructed using CONCORD and standard geometric parameters available in the SYBYL 7.2 molecular modeling package.

2.3. Hologram QSAR

Statistical HQSAR modeling was carried out as previously described using the standard parameters implemented in SYBYL 7.2.^{17–19,23,24} Briefly, HQSAR requires only 2D structures and the property value as input. In this method, each molecule in the training set is broken down into several unique structural fragments, which are then arranged to form a molecular hologram, an extended form of fingerprint that encodes all possible molecular fragments (e.g., linear, branched, and overlapping) and maintains a count of the number of occurrences of each fragment. Incorporation of information about each fragment, and each of its constituent sub-fragments, implicitly encodes 3D structural information, such as hybridization and chirality.^{17,19,24} With the

transformation of the chemical representation of a molecule into its corresponding molecular hologram, this method requires no explicit 3D information (e.g., 3D structures, putative binding conformations, and molecular structural alignment).^{17–19,25} HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. Several combinations of fragment distinction were considered during the QSAR modeling runs. HQSAR analyses were performed by screening the 12 default series of hologram lengths. The influence of different fragment sizes was also investigated. The patterns of fragment counts from the training set molecules were then related to the experimental oral bioavailability data using full cross-validated r^2 (q^2) partial least squares (PLS) leave-one-out (LOO), and leave-many-out methods.^{17–19,22,23} The predictive ability of the models was assessed by their q^2 values.

3. Results and discussion

3.1. Data set characterization

The data set of 302 compounds employed in this work is structurally diverse, containing enzyme inhibitors, receptor agonists and antagonists as well as other biologically active agents of several important therapeutic classes, including antibiotics, analgesics, antivirals, anticancers, antibacterials, antifungals, antidepressants, antiepileptics, antihypertensives, anti-inflammatories, antiparasitics, anxiolytics, antipsychotics, antispasmodics, among others. An example of the high chemical and pharmacological diversity of the data set is shown in Figure 1. Human oral bioavailability is represented as the percentage of an administered dose of a drug that reaches the systemic circulation after oral administration. When two or more values of oral bioavailability for the same compound were available or a value had a range, the average value was employed. In cases where large differences (>20%) between two or more values were found, the compounds were not incorporated in the data set. Training set compounds containing one asymmetric (chiral) center, for which the corresponding oral bioavailability was determined for the racemate, were considered as the individual enantiomers and modeled accordingly.

The distribution of the human oral bioavailability data for the complete data set is presented in form of a histogram in Figure 2. Although weighted toward the high-bioavailability end of the spectrum, values of bioavailability (%) are acceptably distributed across the range of values. Basically, three distinct groups can be seen in Figure 2, indicated as ‘low’ ($\leq 40\%$, 74 compounds), ‘intermediate’ (41–80%, 127 compounds) and ‘high’ (>80%, 101 compounds) bioavailability values. These compounds were grouped in such a way to provide reasonable ranges for comparison reasons.

The selection of appropriate training and test sets is of critical importance in the generation of robust QSAR models. A hierarchical cluster analysis was carried out

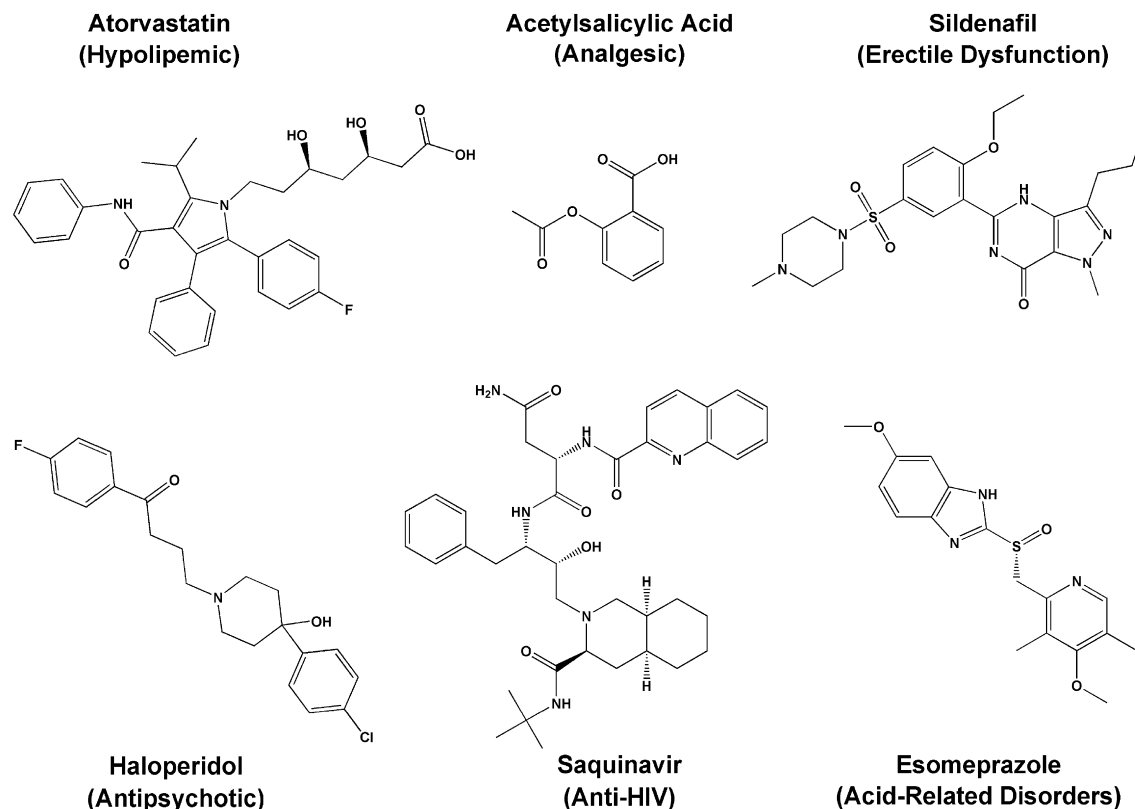


Figure 1. Chemical structure and therapeutic class of representative drugs of the data set.

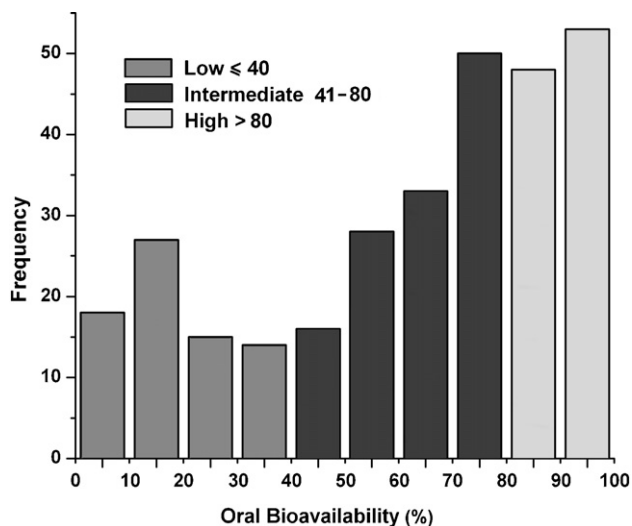


Figure 2. Histogram representation of the distribution of human oral bioavailability for the 302 data set compounds.

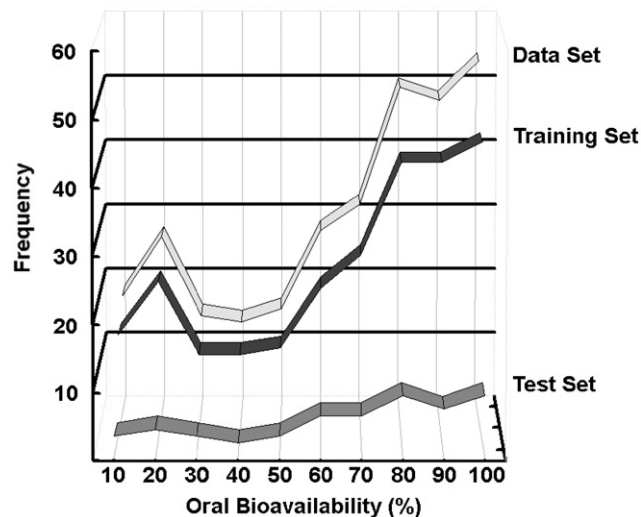


Figure 3. Data set, training set, and test set composition.

with the assistance of the Tsar 3D program, as depicted in Figure 3. In this study, the original data set of 302 compounds was arranged into training (001–250, Table 3) and test (251–302, Table 3) sets such that structurally diverse molecules possessing property values of a wide range were included in both sets. For example, the 302 data set drugs of different therapeutic classes were included in the final training and test sets, respectively, as follows (shown in percentages): analgesics (9.4 and 7.8); antibacterials (9.9 and 13.5); anticancers (5.6 and

2.0); antidepressants (4.2 and 1.5); antiemetics (3.3 and 2.0); antihypertensives (11.8 and 15.6); anti-inflammatories (6.1 and 7.8); antilipemics (2.0 and 5.9); contraceptives (1.9 and 1.9); immunosuppressives (1.9 and 1.9); and others (43.9 and 40.1). Thus, the data set is suitable for HQSAR model development in terms of structural diversity, distribution of drug classes and property values. The training set was then used to generate the HQSAR models, while the test set was held out for the process of model external validation.

3.2. HQSAR modeling

HQSAR investigations require selecting values for parameters that specify hologram length, as well as the size and type of fragments that are to be encoded.^{17,19} Holograms were created using the standard parameters implemented in SYBYL 7.2. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor atoms (DA). Several combinations of these parameters were considered using the fragment size default (4–7), as follows: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/H, A/B/C/Ch, A/B/C/Ch/DA, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/H/Ch/DA, A/B/C/H/DA, and A/B/Ch/DA. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins, using the fragment size default (4–7). The statistical results from the PLS analyses for the 250 training set compounds using the several fragment distinction combinations are summarized in Table 1.

The best statistical results were obtained using either A/B/C ($q^2 = 0.52$, $r^2 = 0.89$) or A/B/C/Ch ($q^2 = 0.53$, $r^2 = 0.90$) as distinction information (models 2 or 6, respectively). The influence of different fragment sizes on the statistical parameters was further investigated for the HQSAR generated using A/B/C/Ch. Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint, therefore, a fundamental aspect is to be considered regarding the statistical generation of QSAR models.^{17–19} These parameters can be adjusted to include larger or smaller fragments in the molecular holograms. As mentioned previously, the fragment size default (4–7), which has proved useful in a number of situations, was first used to derive the HQSAR models. The statistical results for the sequence of different fragment sizes evaluated (including, 2–5, 3–6, 4–7, 5–8, 6–9, 7–10, 8–11, and 9–12) are summarized in Table 2.

Table 1. Results of HQSAR analyses for various fragment distinctions on the key statistical parameters using the fragment size default (4–7)

Model	Fragment distinction	Statistical parameters			
		q^2	r^2	N	HL
1	A/B	0.43	0.60	5	53
2	A/B/C	0.52	0.89	8	353
3	A/B/C/H	0.50	0.81	8	257
4	A/B/C/H/Ch	0.48	0.79	7	257
5	A/B/H	0.46	0.75	7	401
6	A/B/C/Ch	0.53	0.90	8	353
7	A/B/C/Ch/DA	0.47	0.87	7	353
8	A/B/DA	0.41	0.73	8	353
9	A/B/C/DA	0.52	0.81	6	307
10	A/B/H/DA	0.38	0.70	7	401
11	A/B/C/H/Ch/DA	0.40	0.75	7	353
12	A/B/C/H/DA	0.44	0.76	6	353
13	A/B/Ch/DA	0.42	0.69	6	401

q^2 , cross-validated correlation coefficient. r^2 , noncross-validated correlation coefficient. SEE, noncross-validated standard error. N , optimal number of components. HL, hologram length.

Table 2. HQSAR analysis for the influence of various fragment sizes on the key statistical parameters using the fragment distinction A/B/C/Ch

Fragment size	Statistical parameters				
	q^2	r^2	SEE	N	HL
2–5	0.42	0.70	16.11	6	307
3–6	0.47	0.81	12.81	6	353
4–7	0.53	0.90	9.14	8	353
5–8	0.42	0.82	12.37	6	353
6–9	0.47	0.81	12.60	7	151
7–10	0.70	0.93	7.60	8	199
8–11	0.44	0.83	11.95	7	199
9–12	0.35	0.64	17.47	4	199

q^2 , cross-validated correlation coefficient. r^2 , noncross-validated correlation coefficient. SEE, noncross-validated standard error. N , optimal number of components. HL, hologram length.

The variation of fragment size led to the generation of a substantially better HQSAR model compared to that derived for the fragment size default (4–7). As it can be seen, the best statistical results ($q^2 = 0.70$, $r^2 = 0.93$) among all models were obtained with the fragment size 7–10. A similar procedure was applied for the evaluation of the fragment distinction A/B/C of model 2 (results not shown), but the results were not any better than those described for the HQSAR model 6. The stability of the best model (fragment distinction A/B/C/Ch, fragment size 7–10) was tested using 10 different data set splits, varying the relative distribution of the groups low–med–high. Values of q^2 between 0.66 and 0.70 were obtained indicating the high stability of the final model (results not shown).

In terms of validation of a QSAR model, a measure of internal consistency is available in the form of q^2 (Tables 1 and 2). However, the most important test of a QSAR model is its ability to predict the property value for new compounds. As the structure encoded in a 2D fingerprint is directly related to the human oral bioavailability of the training set molecules, HQSAR models should be able to predict the bioavailability of new structurally diverse compounds from its fingerprint. Thus, the predictive power of the best HQSAR model derived using the 250 training set molecules (fragment distinction: A/B/C/Ch; fragment size: 7–10, Table 2) was assessed by predicting bioavailability values for 52 test set molecules, which were not included in the training set for model generation. What is needed is a measure of how close the actual data is to the predicted data of an external data set. The results obtained are presented in Table 3, while the graphic representation for the experimental versus predicted oral bioavailability values of both training (model generation) and test (external evaluation) sets are displayed in Figure 4.

The results show that the test set compounds, which represent the different structural features incorporated within the training set, are reasonably well predicted. The good agreement between experimental and predicted bioavailability values indicates the robustness of the HQSAR model. The predictive power ($r^2_{\text{pred}} = 0.85$) of the model generated is remarkable given the inherent complexity of the target property and the large chemical universe of

Table 3. Experimental, predicted, and residual values of human oral bioavailability for both training and test set compounds for the final HQSAR model

No.	Compound	Exp ^a	Pred ^b	Res ^c	No.	Compound	Exp ^a	Pred ^b	Res ^c
<i>Training set</i>									
001	Abacavir ²²	86.00	79.66	6.34	002	Acetaminophen ²²	88.00	75.20	12.80
003	Acetylsalicylic acid ²²	68.00	74.15	-6.15	004	Albuterol ²² (S)	71.00	70.88	0.12
005	Almotriptan ²¹	70.00	69.11	0.89	006	Alosetron ²¹	55.00	64.95	-9.95
007	Anastrozole ²²	84.00	86.12	-2.12	008	Aprepitant ²¹	62.50	74.29	-11.79
009	Atomoxetine ²¹	63.00	77.19	-14.19	010	Atorvastatin ²¹	14.00	19.60	-5.60
011	Atovaquone ²¹	23.00	17.38	5.62	012	Bepridil ²³ (R)	60.00	63.56	-3.56
013	Bosentan ²¹	50.00	47.62	2.38	014	Bromocriptine ²²	4.50	-2.28	6.78
015	Budesonide ²³	11.00	21.23	-10.23	016	Bufuralol ²³ (R)	46.00	61.48	-15.48
017	Bufuralol ²³ (S)	46.00	58.41	-12.41	018	Bumetanide ²²	81.00	72.65	8.35
019	Bupropion ²³ (R)	70.00	76.94	-6.94	020	Bupropion ²³ (S)	70.00	64.66	5.34
021	Busulfan ²¹	80.00	71.51	8.49	022	Calcitriol ²²	61.00	61.44	-0.44
023	Candesartan ²¹ (R)	15.00	2.45	12.55	024	Candesartan ²¹ (S)	15.00	9.60	5.40
025	Carbamazepine ²²	70.00	69.62	0.38	026	Cefaclor ²³	90.00	88.57	1.43
027	Cefazolin ²²	90.00	81.51	8.49	028	Cephalexin ²²	90.00	96.70	-6.70
029	Chlorambucil ²²	87.00	78.84	8.16	030	Chloramphenicol ²³	69.00	79.49	-10.49
031	Chloramphenicol palmitate ²³	80.00	75.36	4.64	032	Chloroquine ²² (S)	80.00	79.12	0.88
033	Chlorpromazine ²²	32.00	24.71	7.29	034	Chlortetracycline ²³	27.50	19.59	7.91
035	Chlorthalidone ²² (R)	64.00	62.22	1.78	036	Ciclopriol ²³ (R)	100.00	92.67	7.33
037	Ciclopriol ²³ (S)	100.00	95.18	4.82	038	Cimetropium bromide ²³	2.00	22.22	-20.22
039	Ciprofloxacin ²¹	70.00	83.34	-13.34	040	Citalopram ²² (R)	80.00	87.92	-7.92
041	Citalopram ²² (S)	80.00	68.42	11.58	042	Clonidine ²²	60.00	74.50	-14.50
043	Clorazepate ²² (R)	91.00	91.27	-0.27	044	Clorazepate ²² (S)	91.00	92.96	-1.96
045	Chloroquine ²² (R)	80.00	77.87	2.13	046	Clozapine ²²	55.00	65.38	-10.38
047	Cyclophosphamide ²²	74.00	72.82	1.18	048	Dapsone ²³	93.00	90.53	2.47
049	Delavirdine ²¹	96.00	92.66	3.34	050	Diazepam ²²	100.00	87.63	12.37
051	Diflunisal ²³	100.00	90.31	9.69	052	Digitoxin ²³	88.50	86.82	1.68
053	Digoxin ²¹	95.00	90.23	4.77	054	Dihydroergosine ²³	10.00	6.35	3.65
055	Diltiazem ²¹	40.00	47.61	-7.61	056	Diphenhydramine ²²	72.00	81.68	-9.68
057	Disopiramide ²² (R)	83.00	70.21	12.79	058	Disopiramide ²² (S)	83.00	95.43	-12.43
059	Dixyrazine ²³ (S)	10.00	5.69	4.31	060	Domperidone ²³	14.00	14.34	-0.34
061	Doxapram ²³ (R)	61.00	63.03	-2.03	062	Doxapram ²³ (S)	61.00	45.34	15.66
063	Doxazosin ²³	65.00	62.11	2.89	064	Doxycycline ²²	93.00	90.60	2.40
065	Drospirenone ²¹	76.00	76.08	-0.08	066	Dutasteride ²¹	60.00	73.93	-13.93
067	Eletriptan ²¹	62.50	62.38	0.12	068	Emtricitabine ²¹	93.00	82.78	10.22
069	Enalapril ²²	41.00	45.59	-4.59	070	Endralazine ²³	75.00	78.38	-3.38
071	Eproxidine ²³	70.00	66.72	3.28	072	Erythromycin ²²	35.00	40.74	-5.74
073	Estradiol valerate ²³	3.00	13.19	-10.19	074	Ethambutol ²²	77.00	82.66	-5.66
075	Ethinyl estradiol ²¹	40.00	38.26	1.74	076	Esomeprazole ²¹	90.00	96.50	-6.50
077	Famciclovir ²¹	77.00	70.43	6.57	078	Felodipine ²² (R)	15.00	6.13	8.87
079	Felodipine ²² (S)	15.00	4.82	10.18	080	Fenflumizole ²³	50.00	49.62	0.38
081	Fenfluramine ²³ (R)	89.00	83.38	5.62	082	Fenoprofen ²³ (R)	80.00	89.82	-9.82
083	Flecainide ²³ (R)	95.00	82.15	12.85	084	Flecainide ²³ (S)	95.00	83.34	11.66
085	Flucloxacillin ²³	49.00	38.61	10.39	086	Flucytosine ²¹	83.50	81.52	1.98
087	Flunisolide ²³	20.00	23.63	-3.63	088	Fluocortolone ²³	83.50	77.39	6.11
089	Flurbiprofen ²² (R)	92.00	85.47	6.53	090	Fluticasone ²¹	1.00	0.06	0.94
091	Gatifloxacin ²² (R)	96.00	98.47	-2.47	092	Gefitinib ²¹	60.00	61.81	-1.81
093	Gemifloxacin ²¹ (S)	71.00	72.86	-1.86	094	Glipizide ²²	95.00	92.13	2.87
095	Glyburide ²²	95.00	96.27	-1.27	096	Granisetron ²²	60.00	69.74	-9.74
097	Haloperidol ²²	60.00	59.64	0.36	098	Hydrochlorothiazide ²²	71.00	84.86	-13.86
099	Hydromorphone ²¹	24.00	21.00	3.00	100	Ibuprofen ²² (S)	80.00	72.21	7.79
101	Idarubicin ²²	28.00	34.18	-6.18	102	Imipramine ²²	42.00	53.33	-11.33
103	Irbesartan ²¹	70.00	77.05	-7.05	104	Isradipine ²² (R)	19.50	14.26	5.24
105	Itraconazole ²¹	55.00	54.72	0.28	106	Ketoprofen ²³ (S)	85.00	89.63	-4.63
107	Lamivudine ²²	86.00	86.69	-0.69	108	Lansoprazole ²² (R)	80.00	83.57	-3.57
109	Lansoprazole ²² (S)	80.00	83.57	-3.57	110	Letrozole ²²	99.90	98.34	1.56
111	LAAM ²²	47.00	50.76	-3.76	112	Levonorgestrel ²²	87.00	93.83	-6.83
113	Linezolid ²¹	100.00	95.28	4.72	114	Lormetazepam ²³ (R)	75.00	85.19	-10.19
115	Losartan ²¹	33.00	42.71	-9.71	116	Lovastatin ²²	5.00	15.83	-10.83
117	Melphalan ²² (R)	71.00	72.46	-1.46	118	Melphalan ²² (S)	71.00	64.79	6.21
119	Mepindolol ²³ (R)	82.00	82.63	-0.63	120	Mepindolol ²³ (S)	82.00	84.74	-2.74
121	Metergoline ²³	23.00	23.30	-0.30	122	Methacycline ²³	58.00	60.17	-2.17
123	Methadone ²² (R)	92.00	84.31	7.69	124	Methadone ²² (S)	92.00	88.91	3.09
125	Methimazole ²³	93.00	79.60	13.40	126	Methotrexate ²² (R)	70.00	64.41	5.59

Table 3 (continued)

No.	Compound	Exp ^a	Pred ^b	Res ^c	No.	Compound	Exp ^a	Pred ^b	Res ^c
127	Methylphenobarbital ²³ (R)	73.00	74.12	-1.12	128	Methylphenobarbital ²³ (S)	73.00	75.48	-2.48
129	Methylprednisolone ²²	82.00	82.15	-0.15	130	Methysergide ²³	13.00	8.24	4.76
131	Metoclopramide ²²	76.00	84.05	-8.05	132	Metopimazine ²³	19.00	20.14	-1.14
133	Mexiletine ²² (R)	87.00	79.35	7.65	134	Mianserin ²³ (S)	20.00	28.41	-8.41
135	Midalcipran ²³	84.00	79.19	4.81	136	Midodrine ²¹ (R)	93.00	81.10	11.90
137	Midodrine ²¹ (S)	93.00	80.82	12.18	138	Milrinone ²³	92.00	80.02	11.98
139	Minocycline ²²	97.50	99.46	-1.96	140	Moexipril ²¹	13.00	10.74	2.26
141	Montelukast ²¹	73.00	62.17	10.83	142	Morphine ²¹	40.00	47.75	-7.75
143	Moxifloxacin ²¹	90.00	93.29	-3.29	144	Mycophenolate ²¹	72.00	74.82	-2.82
145	Nadolol ²³ (R)	34.00	45.13	-11.13	146	Nadolol ²³ (S)	34.00	47.68	-13.68
147	Nalbuphine ²²	11.00	4.75	6.25	148	Naloxone ²²	2.00	3.53	-1.53
149	Naltrexone ²³	20.00	24.33	-4.33	150	Naratriptan ²²	67.50	53.78	13.72
151	Nevirapine ²¹	93.00	86.92	6.08	152	Nifedipine ²²	50.00	58.17	-8.17
153	Nimodipine ²¹ (S)	13.00	26.35	-13.35	154	Nitrendipine ²³ (S)	16.00	30.91	-14.91
155	Nitrofurantoin ²²	87.00	81.34	5.66	156	Nizatidine ²¹	70.00	77.26	-7.26
157	Nomifensine ²³ (S)	27.00	33.66	-6.66	158	Norfenfluramine ²³ (R)	85.00	84.41	0.59
159	Norfenfluramine ²³ (S)	85.00	84.02	0.98	160	Norzimelidine ²³	66.00	77.27	-11.27
161	Ofloxacin ²² (R)	97.50	95.15	2.35	162	Ofloxacin ²² (S)	97.50	96.90	0.60
163	Olmesartan ²¹	26.00	27.64	-1.64	164	Ondansetron ²² (R)	62.00	63.33	-1.33
165	Ondansetron ²² (S)	62.00	64.69	-2.69	166	Oseltamivir ²¹	75.00	83.14	-8.14
167	Oxytetracycline ²³	58.00	60.40	-2.40	168	Pantoprazole ²¹ (R)	77.00	74.51	2.49
169	Pantoprazole ²¹ (S)	77.00	74.51	2.49	170	Penicillin G ²³	22.50	31.35	-8.85
171	Pentazocine ²³	18.00	23.28	-5.28	172	Phencyclidine ²³	72.00	61.14	10.86
173	Phenobarbital ²³	96.00	92.26	3.74	174	Phenylethylmalonamide ²³	91.00	91.64	-0.64
175	Phenytoin ²²	90.00	92.95	-2.95	176	Physostigmine ²³	6.00	15.33	-9.33
177	Pimozide ²²	50.00	39.30	10.70	178	Pinacidil ²³ (R)	57.00	71.58	-14.58
179	Pinacidil ²³ (S)	57.00	69.34	-12.34	180	Pipotiazine ²³	26.00	23.90	2.10
181	Piroxicam ²³	100.00	96.44	3.56	182	Pravastatin ²¹	17.00	1.88	15.12
183	Prednisolone phosphate ²²	82.00	83.75	-1.75	184	Prednisone ²²	80.00	66.84	13.16
185	Primaquine ²³ (R)	96.00	84.76	11.24	186	Primaquine ²³ (S)	96.00	88.22	7.78
187	Primidone ²³	92.00	96.90	-4.90	188	Procyclidine ²³ (R)	75.00	77.83	-2.83
189	Procyclidine ²³ (S)	75.00	75.21	-0.21	190	Promethazine ²³ (S)	25.00	30.67	-5.67
191	Propiomazine ²³ (R)	33.00	36.15	-3.15	192	Propylthiouracil ²³	78.00	75.20	2.80
193	Proscillaridin ²³	7.00	6.59	0.41	194	Protriptyline ²³	85.00	84.93	0.07
195	Proxyphylline ²³ (R)	100.00	92.14	7.86	196	Proxyphylline ²³ (S)	100.00	94.69	5.31
197	Quinidine ²²	75.00	63.57	11.43	198	Raloxifene ²¹	2.00	-5.10	7.10
199	Repaglinide ²¹	56.00	57.45	-1.45	200	Ribavirin ²¹	64.00	65.56	-1.56
201	Rifampin ²¹	88.80	87.38	1.42	202	Rizatriptan ²¹	45.00	57.70	-12.70
203	Rofecoxib ²¹	93.00	94.59	-1.59	204	Saccharin ²³	84.00	82.37	1.63
205	Saquinavir ²¹	4.00	-0.83	4.83	206	Sildenafil ²¹	40.00	34.79	5.21
207	Sirolimus ²¹	14.00	6.18	7.82	208	Sobrerol ²³ (trans)	72.00	69.4	2.60
209	Sotalol ²³ (R)	60.00	62.28	-2.28	210	Sotalol ²³ (S)	60.00	58.41	1.59
211	Stavudine ²²	82.00	79.30	2.70	212	Suprofen ²³ (R)	92.00	80.85	11.15
213	Suprofen ²³ (S)	92.00	78.06	13.94	214	Tacrolimus ²¹	17.00	15.47	1.53
215	Telenzepine ²³	54.00	51.91	2.09	216	Telithromycin ²¹	57.00	48.81	8.19
217	Testosterone ²³	7.00	0.67	6.33	218	Theophylline ²¹	98.80	86.12	12.68
219	Tiagabine ²¹	90.00	88.68	1.32	220	Tiapamil ²³	22.00	23.94	-1.94
221	Tocainide ²² (R)	89.00	79.35	9.65	222	Tocainide ²² (S)	89.00	79.16	9.84
223	Tolbutamide ²²	85.00	85.90	-0.90	224	Toliprolol ²³ (S)	90.00	80.19	9.81
225	Topiramate ²²	70.00	71.56	-1.56	226	Topotecan ²²	32.00	32.55	-0.55
227	Torasemide ²³	91.00	89.60	1.40	228	Tramadol ²¹	75.00	70.63	4.37
229	Trandolapril ²¹	10.00	7.92	2.08	230	Triamterene ²²	51.00	59.44	-8.44
231	Triazolam ²²	44.00	44.55	-0.55	232	Trimethoprim ²²	63.00	70.08	-7.08
233	Tropium ²¹	9.60	6.38	3.22	234	Valganciclovir ²¹ (R)	59.40	75.30	-15.90
235	Valganciclovir ²¹ (S)	59.40	71.01	-11.61	236	Valproic acid ²¹	90.00	78.64	11.36
237	Venlafaxine ²² (S)	45.00	42.54	2.46	238	Verapamil ²¹ (S)	13.50	16.39	-2.89
239	Verdenafil ²¹	15.00	15.07	-0.07	240	Viloxazine ²³ (R)	85.00	83.58	1.42
241	Viloxazine ²³ (S)	85.00	82.07	2.93	242	Warfarin ²² (R)	93.00	93.35	-0.35
243	Warfarin ²² (S)	93.00	90.07	2.93	244	Zaleplon ²¹	30.00	38.45	-8.45
245	Zidovudine ²²	64.00	75.48	-11.48	246	Ziprasidone ²¹	60.00	74.32	-14.32
247	Zolmitriptan ²¹	40.00	54.00	-14.00	248	Zolpiclone ²³ (R)	80.00	79.36	0.64
249	Zolpiclone ²³ (S)	80.00	90.18	-10.18	250	Zolpidem ²²	72.00	73.19	-1.19
<i>Test set</i>									
251	Amiodarone ²²	50.00	55.77	-5.77	252	Bepridil ²³ (S)	60.00	49.36	10.64
253	Caffeine ²³	100.00	94.56	5.44	254	Chlorthalidone ²² (S)	64.00	48.81	15.19

(continued on next page)

Table 3 (continued)

No.	Compound	Exp ^a	Pred ^b	Res ^c	No.	Compound	Exp ^a	Pred ^b	Res ^c
255	Cimetidine ²²	60.00	77.71	-17.71	256	Clarithromycin ²¹	50.00	44.82	5.18
257	Clavulanate ²²	75.00	63.22	11.78	258	Clofibrate ²²	95.00	79.76	15.24
259	Clonazepam ²¹	90.00	79.95	10.05	260	Cloxacillin ²³	37.00	35.60	1.40
261	Codeine ²³	55.00	45.49	9.51	262	Dixyrazine ²³ (R)	10.00	12.33	-2.33
263	Dronabinol ²¹	15.00	5.68	9.32	264	Escitalopram ²¹ (S)	80.00	68.42	11.58
265	Fenfluramine ²³ (S)	89.00	81.39	7.61	266	Fenopropfen ²³ (S)	80.00	88.32	-8.32
267	Fenoximone ²³	53.00	63.32	-10.32	268	Flecainide ²² (R)	70.00	82.15	-12.15
269	Fluphenazine ²²	2.70	19.42	-16.72	270	Flurbiprofen ²² (S)	92.00	78.71	13.29
271	Fluvastatin ²¹	24.00	30.79	-6.79	272	Gabapentin ²¹	60.00	73.38	-13.38
273	Gatifloxacin ²² (S)	96.00	104.69	-8.69	274	Gemifloxacin ²¹	71.00	66.47	4.53
275	Glimepiride ²²	100.00	107.33	-7.33	276	Ibuprofen ²² (R)	80.00	70.43	9.57
277	Isradipine ²² (S)	19.50	10.74	8.76	278	Ketoprofen ²³ (R)	85.00	96.79	-11.79
279	Levofloxacin ²²	99.00	96.89	2.11	280	Lorazepam ²² (R)	93.00	83.31	9.69
281	Lorazepam ²² (S)	93.00	78.53	14.47	282	Lormetazepam ²³ (S)	75.00	81.87	-6.87
283	Methotrexate ²² (S)	70.00	59.45	10.55	284	Mexiletine ²² (S)	87.00	76.76	10.24
285	Midazolam ²³	40.50	54.81	-14.31	286	Nateglinide ²¹	73.00	60.01	12.99
287	Nimodipine ²¹	13.00	28.34	-15.34	288	Nitrazepam ²³	78.00	87.02	-9.02
289	Nitrendipine ²³ (R)	16.00	31.80	-15.80	290	Nomifensine ²³ (R)	27.00	43.55	-16.55
291	Norethindrone ²³	64.00	77.49	-13.49	292	Pirazolac ²³	93.50	101.81	-8.31
293	Pramipexole ²¹	90.00	81.72	8.28	294	Prazosin ²²	68.00	74.90	-6.90
295	Procainamide ²²	83.00	70.91	12.09	296	Risperidone ²¹	70.00	80.11	-10.11
297	Ropinirole ²¹	55.00	68.84	-13.84	298	Simvastatin ²²	5.00	4.90	0.10
299	Spirolactone ²³	25.00	18.91	6.09	300	Tetracycline ²²	77.00	94.25	-17.25
301	Toliprolol ²³ (R)	90.00	77.18	12.82	302	Venlafaxine ²² (R)	45.00	47.45	-2.45

^a Experimental.

^b Predicted.

^c Residual, the difference between experimental and predicted values.

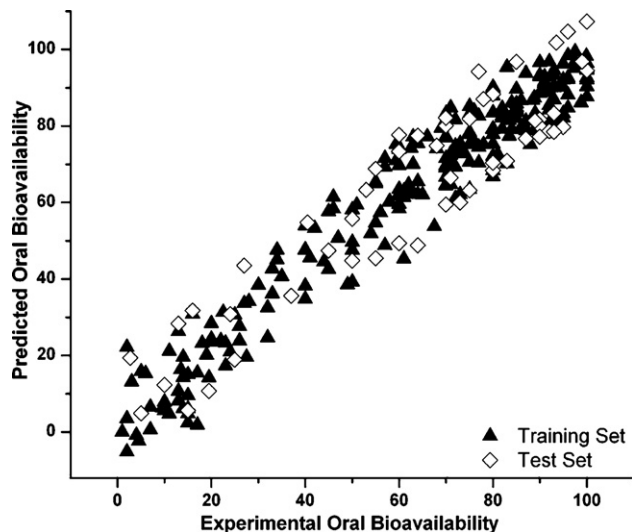


Figure 4. Plot of predicted versus experimental human oral bioavailability for training (250 molecules, model generation) and test sets (52 molecules, external validation).

structural types from which the model was derived. This model can be employed to assist the processes of chemical library design, virtual screening, and high-throughput screening. Compound libraries usually possess broad chemical diversity, therefore, in silico ADME models that are needed to screen these libraries should inevitably be global models designed to cover a broad scope of the chemical space. This was achieved by training the model with compounds from numerous chemical classes. The level of predictability is also adequate for the goal at this stage, given the chemical diversity and number of com-

pounds involved. However, the model possesses some limitations that should be considered when interpreting the results. Considering that the majority of the data set members are drugs, those compounds which do not follow Lipinski's Rule of 5²⁶ would not have significant representation in the data set. Poorly soluble and not orally bioavailable drugs are not included in the data set, therefore, the model is not suitable in this case. In addition, in silico ADME approaches face considerable challenges in drug design. For instance, human oral bioavailability is a difficult property to measure because of the complexity of the underlying biological processes. These include variability of the experimental data, determination of the standard error, and validation of the massive amount of data, among others. In general, in silico ADME models require constant refinement and updating to ensure their applicability to developing medicinal chemistry in a way to continuously improve their quality and predictive power.

4. Conclusions

Drug discovery programs generally focus on the development of orally administered drugs for reasons of convenience, acquiescence, and market perspectives. The use of computational models in the prediction of pharmacokinetic properties of compounds is growing rapidly in drug discovery due to the benefits they provide in throughput and early application in the design of new drug candidates. The use of in silico ADME models has progressed with significant improvements in predictability and simplification, among other relevant aspects with the advance of new technologies. The HQSAR

model described herein shows both good internal and external consistency, and should be useful for the design of new drug candidates having improved oral bioavailability. Although such a model cannot completely replace in vitro and in vivo testing, it is a useful tool for rapidly screening compounds for their probable behavior, allowing early elimination of unfavorable candidates.

Acknowledgments

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation), CNPq (The National Council for Scientific and Technological Development) and CAPES (The Brazilian Coordination for the Improvement of Higher Education Personnel), Brazil.

References and notes

- O'Brien, S. E.; de Groot, M. J. *J. Med. Chem.* **2005**, *48*, 1287.
- Mager, D. E. *Adv. Drug Deliv. Rev.* **2006**, *58*, 1326.
- van de Waterbeemd, H.; Gifford, E. *Nat. Rev. Drug Discovery* **2003**, *2*, 192.
- Lombardo, F.; Gifford, E.; Shalaeva, M. *Mini-Rev. Med. Chem.* **2003**, *3*, 861.
- Saunders, K. C. *Drug Discovery Today: Technol.* **2004**, *1*, 373.
- Ekins, S.; Nikolsky, Y.; Nikolskaya, T. *Trends Pharmacol. Sci.* **2005**, *26*, 202.
- Weiner, D. *Drug Discovery Dev.* **2006**, *12*, 54.
- Hou, T.; Wang, J.; Zhang, W.; Xu, X. *J. Chem. Inf. Model.* **2007**, *47*, 208.
- Jensen, B. F.; Vind, C.; Padkjaer, S. B.; Brockhoff, P. B.; Refsgaard, H. H. *J. Med. Chem.* **2007**, *50*, 501.
- Leitão, A.; Andricopulo, A. D.; Oliva, G.; Pupo, M. T.; de Marchi, A. A.; Vieira, P. C.; da Silva, M. F. G. F.; Ferreira, V. F.; de Souza, M. C. B. V.; Sa, M. M.; Moraes, V. R. S.; Montanari, C. A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2199.
- Yoshida, F.; Topliss, J. G. *J. Med. Chem.* **2000**, *43*, 2575.
- Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619.
- Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. *Bioorg. Med. Chem.* **2004**, *12*, 3391.
- Jónsdóttir, S. Ó.; Jorgensen, F. S.; Brunak, S. *Bioinformatics* **2005**, *21*, 2145.
- Beresford, A. P.; Selick, H. E.; Tarbit, M. H. *Drug Discovery Today* **2002**, *7*, 109.
- Tetko, I. V.; Bruneau, P.; Mewes, H. W.; Rohrer, D. C.; Poda, G. I. *Drug Discovery Today* **2006**, *11*, 700.
- Castilho, M. S.; Postigo, M. P.; de Paula, C. B. V.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D. *Bioorg. Med. Chem.* **2006**, *14*, 516.
- Salum, L. B.; Polikarpov, I.; Andricopulo, A. D. *J. Mol. Graphics Modell.* **2007**, *25*, 434.
- Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119.
- Physicians' Desk Reference (PDR), 59th ed.; Thomson, 2005.
- Hardman, J. G.; Limbird, L. E.; Gilman, A. G. *Pharmacokinetic Data*, 10th ed. In *Goodman's The Pharmacological Basis of Therapeutics*; McGraw-Hill: United States of America, 2001, p. 1924.
- Sietsema, W. K. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1989**, *27*, 179.
- Honorio, K. M.; Garratt, R. C.; Polikarpov, I.; Andricopulo, A. D. *Lett. Drug Des. Discovery* **2006**, *3*, 261.
- HQSAR™ Manual, SYBYL 7.2, Tripos Inc., St. Louis, MO, 2003.
- Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D. *Lett. Drug Des. Discovery* **2007**, *4*, 106.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeny, P. J.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.

In Silico Prediction of Human Plasma Protein Binding Using Hologram QSAR

Tiago L. Moda^a, Carlos A. Montanari^b and Adriano D. Andricopulo^{a,*}

^aLaboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970, São Carlos-SP, Brazil

^bGrupo de Química Medicinal de Produtos Naturais, Instituto de Química de São Carlos, Universidade de São Paulo, 13560-970, São Carlos-SP, Brazil

Received May 18, 2007; Revised June 20, 2007; Accepted July 01, 2007

Abstract: Human plasma protein binding (PPB) is an important pharmacokinetic property in drug design. Hologram quantitative structure-activity relationships (HQSAR) were conducted on a series of structurally diverse molecules with known PPB. The best statistical model ($q^2 = 0.72$, $r^2 = 0.91$) was used to predict the PPB of 62 test set compounds, and the predicted values were in good agreement with the experimental results.

Keywords: ADME, Drug design, QSAR, Plasma protein binding, Pharmacokinetics.

Modern drug discovery tools such as combinatorial chemistry, high-throughput screening and molecular docking techniques have allowed the high-speed generation of leads as promising candidates for further drug development [1]. Researchers in the pharmaceutical and biotech industries have been developing tools over the past years to maximize the efficacy of drugs while minimizing toxicity. A drug intended for use in humans should have an ideal balance of efficacy and safety, as well as good pharmacokinetic properties [2]. Problems with drug candidates' absorption, distribution, metabolism and excretion (ADME), however, have been identified as a major cause of drug candidate failure in late stages of the drug development process [3, 4]. A variety of *in silico* ADME models is currently available due to a better appreciation that pharmacokinetic properties should be considered early in drug discovery programs. *In vitro* and *in vivo* ADME assays (e.g., Caco-2, PAMPA, MDCK, assessment of absorption, metabolism studies) are relatively expensive in terms of resources, reagents and detection techniques [5]. Therefore, *in silico* methods have emerged during the past decade as a powerful strategy for the modeling and prediction of human ADME properties [6].

One major ADME property is the absorption of a drug candidate into the body. The absorption characteristic of an orally-administered drug is typically related to the drug's permeability — its ability to cross intestinal cell barriers before being distributed throughout the body toward the site of action. Plasma protein binding (PPB) plays a crucial role in drug design, determining the fraction of bioavailable free drug distributed through the various tissues [7]. A number of acidic drugs have high affinity for sites on blood proteins, particularly albumin, whereas the main binding protein for many basic drugs is alpha-1-acid glycoprotein (AGP) [8].

PPB of drugs is expressed as percentage of total drug that is bound. Prediction of the extent to which a molecule binds to plasma proteins is a critical phase of drug development because it influences human metabolism, clearance, and drug safety [9].

Quantitative structure-activity relationships (QSAR) employing both classical and modern technologies have proved useful in a large number of settings [10, 11]. QSAR is a technology that generates descriptors based on molecular structures and uses computational algorithms to relate the key descriptors to the dependent property value of interest [12]. It is therefore possible to match these unique characteristics to pharmacokinetic properties due to large data sets of relevant properties that have become available over the past few years [13, 14]. In the present work, we have organized a data set of 312 structurally diverse molecules with known human PPB [15, 16], and used the data to create predictive 2D QSAR models, employing the hologram QSAR (HQSAR) method [17]. Compounds whose PPB is strongly dose dependent were not included in the data set. Data set compounds containing one asymmetric (chiral) center, for which the corresponding PPB value was reported for the racemate, were treated as separate drugs (individual enantiomers) and modeled accordingly. The highly diverse compounds in the data set belong to a variety of therapeutic classes, including antiviral, antibacterial, anti-inflammatory, antipsychotic, antifungal, antihypertensive, antidepressant, immunosuppressive, analgesic, antineoplastic and antiemetic, among several others. PPB expressed as percentage was used for modeling purposes.

The data set was divided into training and test sets, containing 250 (001-250, Table 3) and 62 compounds (251-312, Table 3), respectively. The compounds of both training and test sets were carefully selected in order to ensure appropriate property coverage over the entire range of values (1-99.9%). A statistical cluster analysis carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) employing the complete linkage clustering method (Euclidean distances) with no standardization showed that the composition of both

*Address correspondence to this author at the Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970, São Carlos-SP, Brazil; Tel: + 55 16 3373-8095; Fax: +55 16 3373-9881; E-mail: aandrico@if.sc.usp.br

training and test sets is representative of the whole data set. The HQSAR modeling analyses, calculations and visualizations were performed using the SYBYL 7.3 package (Tripos Inc., St. Louis, USA) as previously described [18-22]. Briefly, HQSAR requires only 2D structures and the property value of interest as inputs. In this method, each molecule in the data set is broken down into several unique structural fragments, which are then arranged to form a molecular hologram, an extended form of fingerprint that encodes all possible molecular fragments (i.e., linear, branched, and overlapping) and maintains a count of the number of occurrences of each fragment. Incorporation of information about each fragment, and each of its constituent sub-fragments, implicitly encodes 3D structural information such as hybridization and chirality. With the transformation of the chemical representation of a molecule into its corresponding molecular hologram, this method requires no explicit 3D information for the ligands (e.g., determination of 3D structure, putative binding conformations, and molecular alignment) [19]. HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction [20, 21]. Several combinations of fragment distinction were considered during the QSAR modeling runs, using a standard default series of hologram lengths. The influence of different fragment sizes was also investigated. The patterns of fragment counts from the training set molecules were then related to the experimental plasma protein binding data using the full cross-validated r^2 (q^2) partial least squares (PLS) leave-one-out (LOO) method [18, 20].

In particular, the molecular fragments of the compounds contained in the training set were obtained using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). Several combinations of these param-

eters were considered using the fragment size default (4-7) as showed in Table 1. HQSAR analyses were performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401 bins. The statistical results from the PLS analyses for the 250 training set compounds using the several fragment distinction combinations are listed in Table 1.

The statistical quality of the models was evaluated by their corresponding q^2 and r^2 values. The best statistical results were obtained using the fragment distinction A/B/C/Ch (model 4, $q^2 = 0.54$, $r^2 = 0.87$). The influence of different fragment sizes in the statistical parameters was further investigated for the best HQSAR model generated in Table 1 (model 4), and the results are summarized in Table 2. Fragment size parameters control the minimum and maximum length of fragments to be included in the hologram fingerprint. The results show that the variation of the fragment size did provide a considerable improvement when compared to the results obtained using the fragment size default (4-7). The best statistical result ($q^2 = 0.72$, $r^2 = 0.91$) among all models was obtained with the fragment size 2-5, as shown in Table 2.

As the molecular structure encoded within a 2D hologram is directly related to the property value of the training set molecules, the HQSAR model should be able to predict the PPB for new compounds from its fingerprint. The q^2 LOO procedure used may give a suitable representation of the internal consistency and predictive power of the models. However, the real predictive ability of the HQSAR model derived with the 250 training set molecules (fragment distinction: A/B/C/Ch; fragment size: 2-5, Table 2) was assessed by predicting PPB values of an external test set of 62 molecules (compounds 251-312, Table 3). Prior to prediction, the test set compounds were processed identically to the

Table 1. Results of HQSAR Analyses for Various Fragment Distinctions on the Key Statistical Parameters Using Fragment Size Default (4-7)

Model	Fragment Distinction	Statistical Parameters			
		q^2	r^2	N	HL
1	A/B	0.50	0.77	8	307
2	A/B/DA	0.52	0.84	8	353
3	A/B/H/DA	0.52	0.78	7	401
4	A/B/C/Ch	0.54	0.87	8	353
5	A/B/C/Ch/DA	0.51	0.79	6	401
6	A/B/C/H/DA	0.50	0.80	8	199
7	A/B/C/H/Ch	0.42	0.82	8	353
8	A/B/C/H/Ch/DA	0.36	0.82	8	353
9	A/B/C	0.52	0.87	8	401
10	A/B/C/H	0.38	0.81	8	307
11	A/B/H	0.51	0.75	8	353
12	A/B/C/DA	0.42	0.72	5	353
13	A/B/H/Ch/DA	0.50	0.83	8	401

q^2 , cross-validated correlation coefficient; r^2 , noncross-validated correlation coefficient; HL, hologram length; N , optimal number of components. Fragment Distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor atoms.

Table 2. HQSAR Analyses for the Influence of Various Fragment Sizes on the Key Statistical Parameters

Fragment Size	Statistical Parameters			
	q^2	r^2	N	HL
2-5	0.72	0.91	8	353
3-6	0.59	0.87	8	401
4-7	0.54	0.86	8	353
5-8	0.55	0.86	8	353
6-9	0.38	0.65	5	353
7-10	0.32	0.58	5	307

Table 3. Experimental, Predicted and Residual Plasma Protein Binding Data for Both Training and Test Set Compounds For the Final HQSAR Model

Training Set									
No.	Compound	Exp ^a	Pred ^b	Res ^c	No.	Compound	Exp ^a	Pred ^b	Res ^c
001	25-Desacetyl Rifapentine ¹⁵	97.70	92.53	-5.17	002	Levetiracetam ¹⁵	10.00	18.60	8.60
003	5-Fluorouracil ¹⁶	10.00	24.92	14.92	004	Levo-Alpha Acetyl-Methadol ¹⁶	80.00	70.09	-9.91
005	Acetaminophen ¹⁶	20.00	41.43	21.43	006	Levofloxacin ¹⁶	31.50	21.67	-9.83
007	Acetylsalicylic ¹⁶	49.00	63.73	14.73	008	Linezolid ¹⁵	31.00	35.59	4.59
009	Acrivastine ¹⁵	50.00	62.25	12.25	010	Loratadine ¹⁶	97.00	100.96	3.96
011	Acyclovir ¹⁵	21.00	6.20	-14.8	012	Lorazepam ¹⁶ (R)	91.00	92.25	1.25
013	Albuterol ¹⁶ (R)	7.00	28.35	21.35	014	Lorazepam ¹⁶ (S)	91.00	91.73	0.73
015	Albuterol ¹⁶ (S)	7.00	27.95	20.95	016	Losartan ¹⁵	98.70	89.57	-9.13
017	Alfentanil ¹⁶	92.00	81.29	-10.71	018	Lovastatin ¹⁵	95.00	85.50	-9.50
019	Almotriptan ¹⁵	35.00	33.90	-1.10	020	M Pentamethylmelamine ¹⁵	75.00	92.30	17.30
021	Alosetron ¹⁵	82.00	68.83	-13.17	022	Medroxyprogesterone ¹⁵	90.00	88.95	-1.05
023	Alprazolam ¹⁶	71.00	72.99	1.99	024	Meloxicam ¹⁵	99.40	86.08	-13.32
025	Altretamine ¹⁵	94.00	99.06	5.06	026	Memantine ¹⁵	45.00	40.43	-4.57
027	Amlodipine ¹⁵ (S)	93.00	82.88	-10.12	028	Mercaptopurine ¹⁶	19.00	17.53	-1.47
029	Amoxicillin ¹⁵	20.00	16.42	-3.58	030	Methadone ¹⁶ (R)	89.00	83.52	-5.48
031	Aprepitant ¹⁵	95.00	108.94	13.94	032	Methadone ¹⁶ (S)	89.00	83.83	-5.17
033	Aripiprazole ¹⁵	99.00	108.56	9.56	034	Methotrexate ¹⁶ (R)	46.00	64.52	18.52
035	Atazanavir ¹⁵	86.00	72.54	-13.46	036	Methylprednisolone ¹⁶	78.00	74.85	-3.15
037	Atorvastatin ¹⁵	98.00	99.96	1.96	038	Metoprolol ¹⁶ (R)	12.00	27.59	15.59
039	Atovaquone ¹⁵	99.00	108.90	9.90	040	Metoprolol ¹⁶ (S)	12.00	26.88	14.88
041	Balsalazide ¹⁵	99.00	104.33	5.33	042	Metronidazole ¹⁶	11.00	15.83	4.83
043	Benazepril ¹⁵	96.70	80.21	-16.49	044	Mexiletine ¹⁶ (R)	63.00	48.26	-14.74
045	Benazeprilat ¹⁵	95.30	79.14	-16.16	046	Mexiletine ¹⁶ (S)	63.00	47.63	-15.37
047	Bexarotene ¹⁵	99.00	87.14	-11.86	048	Miglustat ¹⁵	1.00	-9.48	-10.48
049	Bicalutamide ¹⁵	96.00	99.37	3.37	050	Minocycline ¹⁶	76.00	70.62	-5.38
051	Biotin ¹⁵	20.00	35.60	15.60	052	Mirtazapine ¹⁶	85.00	77.42	-7.58
053	Bosentan ¹⁵	98.00	95.53	-2.47	054	Morphine ¹⁶	35.00	53.42	18.42
055	Bromocriptine ¹⁶	93.00	108.56	15.56	056	Mycophenol ¹⁵	97.00	92.83	-4.17

(Table 3). Contd.....

057	Buprenorphine ¹⁵	96.00	83.18	-12.82	058	Mycophenolate ¹⁵	98.00	98.76	0.76
059	Bupropion ¹⁶ (R)	80.00	76.17	-3.83	060	Nab-Verapamil ¹⁵ (R)	94.00	93.44	-0.56
061	Bupropion ¹⁶ (S)	80.00	72.87	-7.13	062	Nab-Verapamil ¹⁵ (S)	90.00	93.44	3.44
063	Captopril ¹⁵	27.50	27.02	-0.48	064	Nalbuphine ¹⁶	50.00	47.12	-2.88
065	Carbamazepine ¹⁵	76.00	68.98	-7.02	066	Nalmefene ¹⁶	34.40	42.64	8.24
067	Carbidopa ¹⁵	36.00	54.91	18.91	068	Naloxone ¹⁵	45.00	38.82	-6.18
069	Carvedilol ¹⁵ (R)	98.00	95.11	-2.89	070	Nelfinavir ¹⁵	98.00	96.69	-1.31
071	Carvedilol ¹⁵ (S)	98.00	94.40	-3.60	072	Nevirapine ¹⁶	60.00	67.73	7.73
073	Cefdinir ¹⁵	65.00	59.83	-5.17	074	Nicardipine ¹⁶ (R)	98.70	94.88	-3.82
075	Cefixime ¹⁶	67.00	57.63	-9.37	076	Nicardipine ¹⁶ (S)	98.70	94.88	-3.82
077	Cefotetan ¹⁶	85.00	89.14	4.14	078	Nimodipine ¹⁵ (S)	95.00	86.20	-8.80
079	Cefprozil ¹⁵	36.00	49.80	13.8	080	Nitrofurantion ¹⁶	62.00	46.20	-15.80
081	Ceftibuten ¹⁵	65.00	59.42	-5.58	082	Norfloxacin ¹⁵	12.50	30.48	17.98
083	Cefuroxime ¹⁶	33.00	48.46	15.46	084	Ofloxacin ¹⁶ (R)	25.00	22.30	-2.70
085	Celecoxib ¹⁵	97.00	96.30	-0.70	086	Ofloxacin ¹⁶ (S)	25.00	21.67	-3.33
087	Cetirizine ¹⁶ (R)	55.00	65.00	10.0	088	Olanzapine ¹⁵	93.00	86.79	-6.21
089	Cetirizine ¹⁶ (S)	55.00	65.31	10.31	090	Olmesartan ¹⁵	99.00	100.91	1.91
091	Chlorpheniramine ¹⁶ (R)	70.00	74.92	4.92	092	Olsalazine ¹⁵	99.00	108.74	9.74
093	Chlorpheniramine ¹⁶ (S)	70.00	73.55	3.55	094	Omeprazole ¹⁶ (R)	95.00	102.80	7.80
095	Chlorpromazine ¹⁶	96.50	88.92	-7.58	096	Omeprazole ¹⁶ (S)	95.00	102.80	7.80
097	Chlorthalidone ¹⁶ (S)	72.00	90.39	18.39	098	Ondansetron ¹⁶ (R)	73.00	80.38	7.38
099	Cilostazo ¹⁵	96.50	76.91	-19.59	100	Ondansetron ¹⁶ (S)	73.00	79.25	6.25
101	Cimetidine ¹⁶	19.00	23.78	4.78	102	Oseltamivir ¹⁵	42.00	31.85	-10.15
103	Ciprofloxacin ¹⁵	30.00	33.40	3.40	104	Oxicarbazepine 10-Hidroxy ¹⁶ (R)	40.00	52.64	12.64
105	Cisapride ¹⁶ (R)	98.00	81.99	-16.01	106	Oxicarbazepine 10-Hidroxy ¹⁶ (S)	40.00	53.36	13.36
107	Citalopram ¹⁶ (R)	80.00	70.41	-9.59	108	Pancuronium ¹⁶	7.00	17.75	10.75
109	Clarithromycin ¹⁶	46.00	38.94	-7.06	110	Pantoprazole ¹⁵ (R)	98.00	103.03	5.03
111	Clonazepam ¹⁵	85.00	93.93	8.93	112	Pantoprazole ¹⁵ (S)	98.00	103.03	5.03
113	Clorazepate ¹⁶ (R)	97.50	83.63	-13.87	114	Pemoline ¹⁵ (R)	50.00	44.08	-5.92
115	Clorazepate ¹⁶ (S)	97.50	84.36	-13.14	116	Pemoline ¹⁵ (S)	50.00	45.68	-4.32
117	Clozapine ¹⁵	97.00	100.26	3.26	118	Penciclovir ¹⁵	20.00	22.03	2.03
119	Cyclophosphamide ¹⁶	13.00	21.59	8.59	120	Perindopril ¹⁵	15.00	25.69	10.69
121	Delavirdine ¹⁵	98.00	96.14	-1.86	122	Phenylephine ¹⁶ (R)	51.00	32.51	-18.49
123	Desloratadine ¹⁵	84.50	94.45	9.95	124	Phenylephine ¹⁶ (S)	51.00	32.12	-18.88
125	Desmethyl Venlafaxine ¹⁶ (R)	30.00	27.17	-2.83	126	Pilocarpine ¹⁵	1.00	15.16	14.16
127	Desmethyl Venlafaxine ¹⁶ (S)	30.00	29.02	-0.98	128	Pimozide ¹⁶	99.00	96.34	-2.66
129	Diazepam ¹⁶	98.70	79.32	-19.38	130	Pioglitazone ¹⁵ (R)	99.00	92.56	-6.44
131	Dicloxacillin ¹⁶	95.80	89.43	-6.37	132	Pioglitazone ¹⁵ (S)	99.00	94.48	-4.52
133	Didanosine ¹⁶	5.00	12.02	7.02	134	Pravastatin ¹⁵	50.00	62.10	12.10
135	Diflunisal ¹⁵	99.00	85.21	-13.79	136	Prazosin ¹⁶	95.00	95.87	0.87
137	Digoxin ¹⁵	25.00	35.85	10.85	138	Prednisolone ¹⁵	80.00	69.16	-10.84

(Table 3). Contd.....

139	Diltiazem ¹⁶	78.00	90.65	12.65	140	Prednisone ¹⁶	75.00	71.91	-3.09
141	Diphenhydramine ¹⁶	78.00	72.36	-5.64	142	Proguanil ¹⁵	75.00	57.34	-17.66
143	Dofetilide ¹⁵	65.00	61.13	-3.87	144	Propafenone ¹⁵ (R)	95.00	89.10	-5.90
145	Donepezil ¹⁶ (R)	96.00	108.56	12.56	146	Propafenone ¹⁵ (S)	95.00	88.39	-6.61
147	Donepezil ¹⁶ (S)	96.00	107.80	11.80	148	Propranolol ¹⁶ (S)	90.00	75.07	-14.93
149	Doxepin ¹⁶	82.00	76.83	-5.17	150	Pyrazinamide ¹⁵	10.00	28.68	18.68
151	Doxorubicin ¹⁶	76.00	83.75	7.75	152	Pyrimethamine ¹⁵	87.00	78.08	-8.92
153	Efavirenz ¹⁵	99.50	95.90	-3.60	154	Quinidine 3-Hydroxy ¹⁶	60.00	43.22	-16.78
155	Enalaprilat ¹⁶	55.00	44.44	-10.56	156	Quinidine ¹⁶	87.00	64.69	-22.31
157	Entacapone ¹⁵	98.00	78.02	-19.98	158	Rabeprazole ¹⁵ (R)	96.30	92.09	-4.21
159	Escitalopram Oxalate ¹⁵ (S)	56.00	70.41	14.41	160	Rabeprazole ¹⁵ (S)	96.30	92.09	-4.21
161	Esmolol ¹⁶ (R)	55.00	43.02	-11.98	162	Ranitidine ¹⁵	15.00	13.04	-1.96
163	Esmolol ¹⁶ (S)	55.00	42.31	-12.69	164	Rifabutin ¹⁶	71.00	75.27	4.27
165	Esomeprazole Magnesium ¹⁵ (S)	97.00	102.80	5.80	166	Rifapentine ¹⁵	93.20	93.24	0.04
167	Ethionamide ¹⁵	30.00	37.45	7.45	168	Risperidone 9-Hydroxy ¹⁶ (S)	77.00	85.06	8.06
169	Etoposide ¹⁶	96.00	105.60	9.60	170	Risperidone ¹⁵	90.00	92.93	2.93
171	Exemestane ¹⁵	90.00	103.31	13.31	172	Ritonavir ¹⁵	98.50	105.44	6.94
173	Famotidine ¹⁵	17.50	27.35	9.85	174	Rivastigmine ¹⁵	40.00	31.07	-8.93
175	Felodipine ¹⁶ (R)	99.60	103.50	3.90	176	Rofecoxib ¹⁵	87.00	89.46	2.46
177	Felodipine ¹⁶ (S)	99.60	103.50	3.90	178	Rosiglitazone ¹⁵	99.80	95.37	-4.43
179	Fenofibrate ¹⁵	99.00	99.08	0.08	180	Saquinavir ¹⁶	98.00	97.93	-0.07
181	Fexofenadine ¹⁶ (R)	65.00	69.84	4.84	182	Sertraline ¹⁵	98.00	102.66	4.66
183	Fexofenadine ¹⁶ (S)	65.00	70.35	5.35	184	Sinvastatin ¹⁵	95.00	98.13	3.13
185	Finasteride ¹⁵	90.00	81.30	-8.70	186	Sirolimus ¹⁵	92.00	104.63	12.63
187	Flecainide ¹⁶ (R)	61.00	54.00	-7.00	188	Sulfamethoxazole ¹⁶	53.00	67.92	14.92
189	Fluconazole ¹⁵	11.50	22.09	10.59	190	Sulfasalazine ¹⁶	99.30	110.51	11.21
191	Fluoxetine ¹⁶ (R)	94.00	92.72	-1.28	192	Sulindac ¹⁶ (R)	94.00	82.67	-11.33
193	Fluoxetine ¹⁶ (S)	94.00	91.95	-2.05	194	Sumatriptan ¹⁵	17.50	39.67	22.17
195	Flutamide 2-Hydroxy ¹⁶	93.00	92.10	-0.90	196	Tacrolimus ¹⁵	99.00	85.66	-13.34
197	Flutamide ¹⁶	95.00	100.03	5.03	198	Tamoxifen ¹⁶	98.00	98.54	0.54
199	Fluvastatin ¹⁵	98.00	96.26	-1.74	200	Tamsulosin ¹⁵	96.50	75.91	-20.59
201	Fluvoxamine ¹⁶	77.00	88.10	11.10	202	Terbinalite ¹⁵	99.00	93.68	-5.32
203	Furosemide ¹⁵	95.00	81.87	-13.13	204	Tetracycline ¹⁶	65.00	54.15	-10.85
205	Ganciclovir ¹⁵	1.50	0.90	-0.60	206	Tiagabine ¹⁵	96.00	88.14	-7.86
207	Gatifloxacin ¹⁶ (R)	20.00	26.69	6.69	208	Timolol ¹⁵	10.00	2.32	-7.68
209	Gatifloxacin ¹⁶ (S)	20.00	28.01	8.01	210	Topotecan Carboxylic Acid ¹⁶	6.60	18.78	12.18
211	Gemfibrozil ¹⁶	97.00	78.44	-18.56	212	Topotecan ¹⁶	21.30	19.56	-1.74
213	Gemifloxacin ¹⁵ (R)	64.00	44.30	-19.70	214	Toremifene ¹⁵	99.50	101.85	2.35
215	Glipizide ¹⁵	98.50	84.20	-14.30	216	Torseamide ¹⁵	99.00	95.24	-3.76
217	Glyburide ¹⁶	99.80	111.97	12.17	218	Trazodone ¹⁶	93.00	87.70	-5.30
219	Granisetron ¹⁵	65.00	62.14	-2.86	220	Tretinoin ¹⁵	95.00	97.95	2.95

(Table 3). Contd.....

221	Haloperidol ¹⁶	92.00	75.91	-16.09	222	Triamterene ¹⁶	61.00	55.04	-5.96
223	Hydrochlorotiazide ¹⁵	67.90	64.01	-3.89	224	Triazolam ¹⁶	90.10	89.84	-0.26
225	Idarubicin ¹⁶	97.00	109.79	12.79	226	Valacyclovir ¹⁵	15.70	14.22	-1.48
227	Idarubicinol ¹⁶	94.00	98.22	4.22	228	Valdecocixib ¹⁵	98.00	96.62	-1.38
229	Imipramine ¹⁶	90.10	76.71	-13.39	230	Venlafaxine ¹⁶ (R)	27.00	26.61	-0.39
231	Indapamide ¹⁵ (R)	75.00	91.84	16.84	232	Venlafaxine ¹⁶ (S)	27.00	28.45	1.45
233	Indinavir ¹⁵	60.00	73.45	13.45	234	Verapamil ¹⁵ (S)	88.00	93.67	5.67
235	Indomethacin ¹⁵	99.00	103.57	4.57	236	Verapamil ¹⁵ (R)	90.00	93.67	3.67
237	Irbesartan ¹⁵	90.00	102.02	12.02	238	Verdenafil ¹⁵	95.00	79.55	-15.45
239	Isosorbide 5-Mononitrate ¹⁶	1.00	5.48	4.48	240	Warfarin ¹⁶ (R)	99.00	107.94	8.94
241	Isotretinoin ¹⁵	99.00	101.58	2.58	242	Warfarin ¹⁶ (S)	99.00	108.33	9.33
243	Isradipine ¹⁶ (R)	95.00	93.83	-1.17	244	Zidovudine ¹⁵	38.00	39.70	1.70
245	Isradipine ¹⁶ (S)	95.00	93.83	-1.17	246	Zileuton ¹⁶ (R)	93.40	80.57	-12.83
247	Lansoprazole ¹⁶ (R)	97.00	98.58	1.58	248	Zileuton ¹⁶ (S)	93.40	81.35	-12.05
249	Lansoprazole ¹⁶ (S)	97.00	98.58	1.58	250	Ziprasidone ¹⁵	90.00	94.88	4.88
Test Set									
No.	Compound	Exp ^a	Pred ^b	Res ^c	No.	Compound	Exp ^a	Pred ^b	Res ^c
251	Abacavir ¹⁶	50.00	55.66	5.66	252	Methotrexate ¹⁵ (S)	46.00	63.10	17.10
253	Acitretin ¹⁶	99.90	89.72	-10.18	254	Moxifloxacin ¹⁶	50.00	34.29	-15.71
255	Albendazole ¹⁶	70.00	59.96	-10.04	256	Naproxen ¹⁶	99.00	86.44	-12.56
257	Aminosalicylic Acid ¹⁶	55.00	59.80	4.80	258	Nicotine ¹⁵	4.90	22.56	17.66
259	Amiodarone ¹⁶	96.00	106.72	10.72	260	Nifedipine ¹⁶	95.00	83.35	-11.65
261	Amitriptyline ¹⁵	94.80	76.32	-18.48	262	Nimodipine ¹⁶ (R)	95.00	86.20	-8.80
263	Amlodipine ¹⁶ (R)	93.00	82.88	-10.12	264	Nizatidine ¹⁶	35.00	23.63	-11.37
265	Azathioprine ¹⁶	30.00	20.95	-9.05	266	Nortriptyline ¹⁵	92.00	75.35	-16.65
267	Bumetanide ¹⁵	99.00	90.32	-8.68	268	Progesterone ¹⁶	97.50	97.28	-0.22
269	Cevimeline ¹⁶	20.00	21.46	1.46	270	Propranolol ¹⁶ (R)	90.00	75.78	-14.22
271	Chlorthalidone ¹⁶ (R)	72.00	90.39	18.39	272	Quetiapine ¹⁶	83.00	67.05	-15.95
273	Cisapride ¹⁵ (S)	98.00	81.36	-16.64	274	Repaglinide ¹⁶	98.00	92.07	-5.93
275	Citalopram ¹⁶ (S)	80.00	70.41	-9.59	276	Ribavirin ¹⁵	1.00	-1.28	-2.28
277	Clavulanate ¹⁶	25.00	21.13	-3.87	278	Rifampin ¹⁶	80.00	97.88	17.88
279	Diclofenac ¹⁶	99.00	104.35	5.35	280	Risedronate ¹⁶	24.00	33.57	9.57
281	Dolasetron Hydrolysed ¹⁶	73.00	63.43	-9.57	282	Risperidone 9-Hydroxy ¹⁵ (R)	77.00	86.66	9.66
283	Drospirenone ¹⁶	97.00	100.72	3.72	284	Ritonavir ¹⁶	98.50	83.80	-14.70
285	Duloxetine ¹⁶ (S)	90.00	91.89	1.89	286	Ropinirole ¹⁶	40.00	44.75	4.75
287	Eprosartam ¹⁶	98.00	105.07	7.07	288	Sulfisoxazole ¹⁶	85.00	74.50	-10.50
289	Estazolam ¹⁶	93.00	74.48	-18.52	290	Sulindac ¹⁵ (S)	94.00	82.67	-11.33
291	Ethinyl Estradiol ¹⁶	97.00	97.33	0.33	292	Tadalafil ¹⁶	94.00	95.82	1.82
293	Ezetimibe ¹⁶	90.00	84.10	-5.90	294	Thalidomide ¹⁶ (R)	55.00	76.88	21.88
295	Flecainide ¹⁵ (S)	61.00	55.02	-5.98	296	Thalidomide ¹⁶ (S)	66.00	78.41	12.41
297	Galantamine ¹⁶	18.00	27.91	9.91	298	Theophylline ¹⁶	40.00	34.56	-5.44
299	Gemifloxacin ¹⁶ (S)	64.00	44.30	-19.70	300	Tinidazole ¹⁶	12.00	12.86	0.86

(Table 3). Contd.....

301	Glimepiride ¹⁵	99.50	97.20	-2.30	302	Tolcapone ¹⁶	99.90	108.47	8.57
303	Indapamide ¹⁶ (S)	75.00	91.35	16.35	304	Tolterodine ¹⁶	96.30	71.85	-24.45
305	Lamivudine ¹⁶	36.00	25.70	-10.30	306	Topiramate ¹⁶	15.00	7.92	-7.08
307	Leflunomide ¹⁶	99.30	88.14	-11.16	308	Valsartan ¹⁶	95.00	84.86	-10.14
309	Levonorgestrel ¹⁶	97.00	92.22	-4.78	310	Zaleplon ¹⁶	60.00	53.89	-6.11
311	Levothyroxine ¹⁶	99.00	79.27	-19.73	312	Zanamivir ¹⁵	10.00	19.73	9.73

^aLiterature. ^bPredicted. ^cResidual: the difference between predicted and experimental values. Training set compounds (001-250) were used for QSAR model generation. Test set compounds (251-312) were used for external model validation.

training set compounds as previously described. The external validation process can be considered the most valuable validation method as these compounds were completely excluded during the training of the model. The results are listed in Table 3, and the graphic results simultaneously displayed in (Fig. 1).

The good agreement between experimental and predicted values for the test set compounds establishes the reliability of the constructed HQSAR model ($r^2_{\text{pred}} = 0.86$). From the moderate residual values, it can be seen that the HQSAR model obtained is quite reliable and should be useful to predict PPB values of novel structurally diverse compounds.

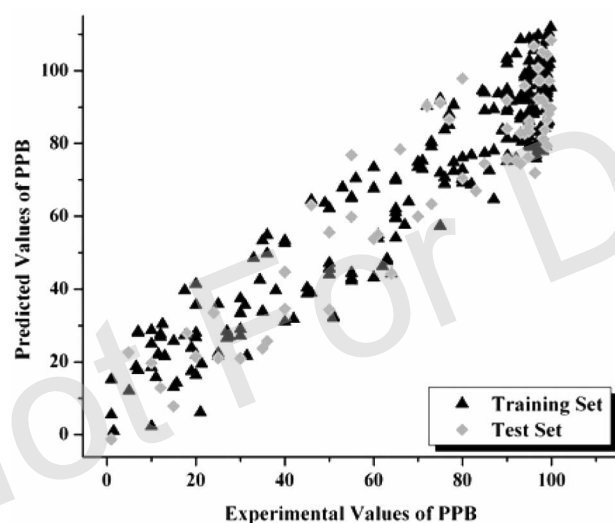


Fig. (1). Plot of Experimental vs. Predicted Plasma Protein Binding for Both Training and Test Sets.

Computational tools are integrated throughout the drug discovery process to fully meet the ADME throughput demands for the screening of large data sets of compounds, creating tools that are faster, simpler and more cost-effective than traditional experimental methods. PPB is an important pharmacokinetic property that has been attracting attention recently for the generation of predictive *in silico* models [23-26]. In this work, we developed a simple useful model employing special molecular holograms for predicting the approximate PPB of chemically diverse compounds based on the molecular structures alone. Although such a model cannot completely replace *in vitro* and *in vivo* testing, it is a useful tool for rapidly screening compounds for their probable behavior, allowing early elimination of unfavorable candidates. In general, *in silico* ADME approaches face considerable challenges in drug design. PPB is a difficult property to

measure because of the complexity of the underlying biological processes. These include the variability of the experimental data, determination of the standard error, standardization and validation of the massive amount of data that is generated, among others. However, it is important to note that the final HQSAR model described herein shows both good internal and external consistency, with substantial predictive power, and should be useful for the design of new drug candidates with improved pharmacokinetic properties.

ACKNOWLEDGEMENT

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation), CNPq (The National Council for Scientific and Technological Development) and CAPES (The Brazilian Coordination for the Improvement of Higher Education Personnel), Brazil.

REFERENCES

- [1] O'Brien, S. E.; de Groot, M. J. *J. Med. Chem.*, **2005**, *48*, 1287.
- [2] Mager, D. E. *Adv. Drug Deliv. Rev.*, **2006**, *58*, 1326.
- [3] van de Waterbeemd, H.; Gifford, E. *Nat. Rev. Drug Discov.*, **2003**, *2*, 192.
- [4] Lombardo, F.; Gifford, E.; Shalaeva, M. *Mini-Rev. Med. Chem.*, **2003**, *3*, 861.
- [5] Saunders, K. C. *Drug Discov. Today*, **2004**, *1*, 373.
- [6] Ekins, S.; Wrighton, S. A. *J. Pharmacol. Toxicol. Methods*, **2001**, *45*, 65.
- [7] Cai, H.; Stoner, C.; Reddy, A.; Freiwald, S.; Smith, D.; Winters, R.; Stankovic, C.; Surendran, N. *Int. J. Pharm.*, **2006**, *308*, 133.
- [8] Aureli, L.; Cruciani, G.; Mariconi, A. *J. Med. Chem.*, **2005**, *48*, 2469.
- [9] Kratochwil, N. A.; Gerber, P. R. *Biochem. Pharmacol.*, **2002**, *64*, 1355.
- [10] Leitão, A.; Andricopulo, A. D.; Oliva, G.; Pupo, M. T.; de Marchi, A. A.; Vieira, P. C.; da Silva, M. F. G. F.; Ferreira, V. F.; de Souza, M. C. B. V.; Sa, M. M.; Moraes, V. R. S.; Montanari, C. A. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 2199.
- [11] Yoshida, F.; Topliss, J. G. *J. Med. Chem.*, **2000**, *43*, 2575.
- [12] Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.*, **2001**, *101*, 619.
- [13] Hansch, C.; Leo, A.; Meikapati, S. B.; Kurup, A. *Bioorg. Med. Chem.*, **2004**, *12*, 3391.
- [14] Jónsdóttir, S. Ó.; Jørgensen, F. S.; Brunak, S. *Bioinformatics*, **2005**, *21*, 2145.
- [15] Physicians' Desk Reference (PDR), 59th ed.; Thomson, **2005**.
- [16] Hardman, J. G.; Limbird, L. E.; Gilman, A. G. *Pharmacokinetic Data. Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 10th ed.; McGraw-Hill: United States of America, **2001**; pp 1924.
- [17] HQSAR™ Manual Release in SYBYL 7.3, Tripos Inc., St. Louis, MO, **2007**.
- [18] Honorio, K. M.; Garratt, R. C.; Polikarpov, I.; Andricopulo, A. D. *Lett. Drug Des. Discov.*, **2006**, *3*, 261.

- [19] Castilho, M. S.; Postigo, M. P.; de Paula, C. B. V.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D. *Bioorg. Med. Chem.*, **2006**, *14*, 516.
- [20] Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 3119.
- [21] Salum, L. B.; Polikarpov, I.; Andricopulo, A. D. *J. Mol. Graph. Model.*, **2007**, *25*, doi:10.1016/j.jmgm.2007.02.001
- [22] Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D. *Lett. Drug Des. Discov.*, **2007**, *4*, 106.
- [23] Saiakhov, R. D.; Stefan, L. R.; Klopman, G. *Perspec. Drug Disc. Des.*, **2000**, *19*, 133.
- [24] Lobell, M.; Sivarajah, V. *Mol. Divers.*, **2003**, *7*, 69.
- [25] Wang, J.; Krudy, G.; Xie, X. Q.; Wu, C.; Holland, G. J. *Chem. Inf. Model.*, **2006**, *46*, 2674.
- [26] Gleeson, M. P. *J. Med. Chem.*, **2007**, *50*, 101.

Not For Distribution

Anexo

ANEXO A

Tabela A-1 – Conjunto padrão de dados de solubilidade aquosa¹²⁰

Composto (CAS)	logS (mol/L)	Composto (CAS)	logS (mol/L)	Composto (CAS)	logS (mol/L)	Composto (CAS)	logS (mol/L)
100-01-6	-2,37	105-53-3	-0,82	108-47-4	0,38	111-44-4	-1,12
100-02-7	-0,74	105-54-4	-1,28	108-48-5	0,45	1114-71-2	-3,35
10004-44-1	-0,07	105-55-5	-1,46	108-67-8	-3,40	111-65-9	-5,24
100-17-4	-2,41	105-57-7	-0,43	108-68-9	-1,40	111-66-0	-4,44
1002-84-2	-4,31	105-66-8	-1,92	108-70-3	-4,48	111-68-2	-1,85
100-41-4	-2,77	105-67-9	-1,19	108-85-0	-2,30	111-70-6	-1,81
100-42-5	-2,82	106-30-9	-2,71	108-86-1	-2,55	111-71-7	-1,70
100-44-7	-2,39	106-32-1	-3,39	108-87-2	-3,85	111-76-2	-0,42
100-46-9	-1,53	106-34-3	-1,73	108-88-3	-2,21	111-82-0	-4,69
100-47-0	-1,00	106-37-6	-4,07	108-90-7	-2,38	111-83-1	-5,06
100-51-6	-0,40	106-38-7	-3,19	108-93-0	-0,44	111-86-4	-2,75
100-52-7	-1,19	106-39-8	-3,63	108-94-1	-0,60	111-87-5	-2,39
100-55-0	0,96	106-41-2	-1,09	108-95-2	0,00	111-92-2	-1,44
100-61-8	-1,28	106-42-3	-2,77	108-98-5	-2,12	112-05-0	-2,75
100-63-0	0,07	106-43-4	-3,08	109-00-2	-0,46	112-30-1	-3,63
100-65-2	-0,74	106-44-5	-0,70	109-02-4	1,00	112-37-8	-3,55
100-66-3	-1,85	106-46-7	-3,27	109-07-9	0,74	112-38-9	-3,40
100-68-5	-2,39	106-47-8	-1,66	109-60-4	-0,72	112-53-8	-4,67
100-75-4	-0,17	106-48-9	-0,70	109-64-8	-2,08	1125-78-6	-1,99
101-21-3	-3,38	106-49-0	-1,21	109-65-9	-2,37	1125-84-4	-0,87
101-27-9	-4,37	106-50-3	-0,38	109-66-0	-3,18	112-72-1	-6,05
101-31-5	-1,91	106-51-4	-0,99	109-67-1	-2,68	1127-76-0	-4,17
101-42-8	-1,60	106-55-8	0,49	109-69-3	-2,03	1129-41-5	-1,80
101-77-9	-2,30	106-57-0	-0,83	109-70-6	-1,85	1134-23-2	-3,40
101-81-5	-4,17	106-63-8	-1,21	109-79-5	-2,18	114-26-1	-2,05
101-84-8	-3,96	106-70-7	-2,00	109-87-5	0,48	115-43-5	-2,18
102-07-8	-3,15	106-93-4	-1,68	109-92-2	-0,85	1156-19-0	-3,68
102170-44-5	-1,11	106-94-5	-1,73	109-94-4	0,15	115-90-2	-2,30
1022-46-4	-4,61	106-95-6	-1,50	109-97-7	-0,17	116-44-9	-3,70
10238-21-8	-5,09	106-99-0	-1,87	109-99-9	0,56	117-34-0	-3,22
1024-57-3	-6,29	107-02-8	0,57	110-00-9	-0,82	1177-87-3	-4,90
102-69-2	-2,28	107-04-0	-1,32	110-02-1	-1,45	117-81-7	-6,96
102-76-1	-0,60	107-05-1	-1,36	110-12-3	-1,33	117-89-5	-4,52
102-82-9	-3,12	107-06-2	-1,06	110-15-6	-0,20	118-08-1	-4,11
10311-84-9	-6,34	107-08-4	-2,29	110-38-3	-4,10	118-10-5	-3,09
103-29-7	-4,62	107-13-1	0,15	110-42-9	-4,63	118-52-5	-2,60
103-33-3	-4,45	107-31-3	0,58	110-43-0	-1,42	1185-33-7	-1,04
103-36-6	-3,00	1077-16-3	-5,21	110-44-1	-1,77	118-55-8	-3,15
103-65-1	-3,37	107-82-4	-2,89	110-53-2	-3,07	118-90-1	-2,06
103-69-5	-1,70	107-83-5	-3,74	110-54-3	-3,84	118-91-2	-1,89
103-70-8	-0,68	107-87-9	-0,19	110-62-3	-0,85	118-92-3	-1,52
103-73-1	-2,33	107-92-6	-0,19	110-74-7	-0,49	119-36-8	-2,34
103-82-2	-0,89	108-03-2	-0,80	1107-99-9	-4,58	1194-65-6	-4,24
103-84-4	-1,33	108-08-7	-4,26	110-82-7	-3,10	119-61-9	-3,12
103-85-5	-1,77	108-10-1	-1,42	110-83-8	-2,59	119-65-3	-1,45
103-88-8	-3,08	108-11-2	-0,80	110-85-0	1,07	119-68-6	-2,88
103-90-2	-1,03	108-20-3	-1,10	110-86-1	0,76	1198-37-4	-1,94
104-04-1	-2,69	108-21-4	-0,55	110-89-4	1,07	119-91-5	-5,40
104-46-1	-3,13	108-36-1	-3,54	110-91-8	1,06	120-07-0	-0,73
104-51-8	-4,06	108-37-2	-3,21	110-94-1	1,00	120-12-7	-6,35
104-75-6	-1,71	108-38-3	-2,82	111-11-5	-3,39	120-36-5	-2,45
104-76-7	-2,11	108-39-4	-0,68	111-13-7	-2,05	120-47-8	-2,35
104-98-3	-1,96	108-42-9	-1,37	111-25-1	-3,81	120-57-0	-1,63

Continua

Continuação

105-05-5	-3,75	108-43-0	-0,70	111-26-2	-1,10	120-72-9	-1,52
105-30-6	-1,11	108-44-1	-0,85	111-27-3	-1,24	120-80-9	0,62
105-37-3	-0,66	108-46-3	0,81	111-43-3	-1,62	120-82-1	-3,59
120-83-2	-1,55	129-00-0	-6,19	149-32-6	0,70	20168-99-4	-5,54
120-89-8	-0,40	129-20-4	-3,73	149-91-7	-1,16	2029-64-3	-3,30
121-34-6	-2,05	13029-08-8	-5,27	150-13-0	-0,40	2032-59-9	-2,36
121-44-8	-0,14	130-95-0	-2,76	150-68-5	-2,89	20354-26-1	-2,82
121-61-9	-1,61	13110-37-7	-3,26	150-69-6	-2,17	2040-96-2	-4,74
121-69-7	-1,92	131-11-3	-1,66	151-67-7	-1,71	204443-27	-0,91
121-75-5	-3,36	131-17-9	-3,13	15258-73-8	-2,10	204443-30	-1,55
121-79-9	-1,78	132-64-9	-4,60	15271-41-7	-2,08	2049-95-8	-4,15
121-92-6	-1,68	132-65-0	-4,38	15299-99-7	-3,57	2050-67-1	-5,80
121-98-2	-2,41	1330-78-5	-6,70	15310-01-7	-4,21	2050-68-2	-6,56
122-11-2	-2,96	133-67-5	-2,68	1540-35-8	-0,88	2051-24-3	-11,62
122-14-5	-4,04	133-90-4	-2,47	15414-82-1	-0,06	2051-60-7	-4,54
122-15-6	-0,85	133-91-5	-3,31	156-59-2	-1,30	2051-61-8	-4,88
122-34-9	-4,55	134-32-7	-1,92	15687-27-1	-3,99	2051-62-9	-5,20
122-39-4	-3,51	13476-55-6	-3,95	1582-09-8	-5,68	205-82-3	-8,00
122-59-8	-1,10	135-19-3	-2,28	15862-07-4	-6,27	205-99-2	-8,23
122-66-7	-2,92	135-98-8	-3,89	15879-93-3	-1,84	206-44-0	-6,00
122-80-5	-0,98	136-60-7	-3,48	15968-05-5	-7,39	207-08-9	-8,49
122-85-0	-1,58	136-77-6	-2,59	1596-84-5	-0,20	208-96-8	-3,96
122-88-3	-2,29	13698-49-2	-4,95	15972-60-8	-3,26	2104-96-3	-6,09
122-97-4	-1,38	137-26-8	-3,90	16118-49-3	-1,83	2114-15-0	-0,30
122-99-6	-0,71	137-32-6	-0,47	1629-60-3	-0,83	2114-20-7	-1,92
123-08-0	-0,96	137-58-6	-1,76	1634-04-4	-0,24	2131-41-1	-4,92
123-11-5	-1,49	138-52-3	-0,85	16529-66-1	-0,96	2136-99-4	-9,15
123-19-3	-1,30	139-40-2	-4,43	16606-02-3	-6,25	2152-44-5	-4,71
123-25-1	-0,96	13993-65-2	-3,94	1668-19-5	-3,40	2163-69-1	-2,36
123-29-5	-3,80	140-67-0	-2,92	1678-91-7	-4,25	2164-17-2	-3,43
123-30-8	-0,80	14073-97-3	-2,49	16878-76-5	-2,36	21725-46-2	-3,15
123-31-9	-0,17	140-88-5	-0,74	16878-77-6	-1,76	217-59-4	-6,74
123-38-6	0,58	14090-87-0	-1,56	1689-83-4	-3,61	218-01-9	-8,06
123-51-3	-0,52	141-76-4	-0,43	1689-84-5	-3,33	2188-67-2	-3,27
123-54-6	0,22	141-78-6	-0,04	1698-60-8	-2,87	2207-01-4	-4,27
123-56-8	0,30	141-84-4	-1,77	1730-37-6	-5,22	22071-15-4	-3,70
123-66-0	-2,31	141-90-2	-2,26	1732-13-4	-5,96	22204-53-1	-4,16
123-72-8	-0,01	14214-32-5	-4,16	1746-81-2	-2,57	22259-30-9	-2,34
123-75-1	1,15	142-28-9	-1,62	17560-51-9	-3,78	2235-90-7	-2,57
123-86-4	-1,24	142-29-0	-2,10	18181-70-9	-6,62	2244-16-8	-2,06
123-92-2	-1,92	142-62-1	-1,06	18181-80-1	-4,93	22862-76-6	-1,61
123-96-6	-2,09	142-68-7	-0,03	18259-05-7	-7,92	229-87-8	-2,78
123-99-9	-1,89	142-82-5	-4,53	1836-75-5	-5,46	22994-85-0	-2,81
124-04-9	-0,82	142-84-7	-0,46	18559-94-9	-1,22	2303-16-4	-4,08
124-07-2	-2,30	142-92-7	-2,46	1861-40-1	-5,53	2303-17-5	-4,88
124-11-8	-5,05	14293-44-8	-3,79	1897-45-6	-5,64	230-46-6	-2,68
124-13-0	-2,36	142-96-1	-1,85	19044-88-3	-5,16	23103-98-2	-1,95
124-19-6	-3,17	143-07-7	-4,62	1912-24-9	-3,55	2315-68-6	-2,67
124-48-1	-1,90	143-08-8	-3,01	1912-26-1	-4,06	23184-66-9	-4,19
124-83-4	-1,42	14309-40-1	-4,60	191-24-2	-9,03	2363-58-8	-5,41
124-94-7	-3,69	14309-41-2	-5,40	1918-02-1	-2,75	2367-82-0	-2,31
125-33-7	-2,64	14437-17-3	-3,77	1918-00-9	-1,70	2370-12-9	-1,52
126-07-8	-4,61	144-62-7	0,38	1918-16-7	-2,48	2385-74-2	-0,29
127-07-1	1,12	144-80-9	-1,23	192-97-2	-7,80	238-84-6	-6,68
127-18-4	-2,54	144-82-1	-2,41	1929-82-4	-3,76	23947-60-6	-3,02
127-19-5	1,11	144-83-2	-2,70	19550-30-2	-0,39	239-64-5	-7,42
127-31-1	-3,43	145-13-1	-4,65	19666-30-9	-5,69	243-17-4	-8,04
127-33-3	-2,52	145-73-3	-0,27	1982-47-4	-4,89	2432-20-4	-0,12
127-69-5	-2,91	146-22-5	-3,80	1982-49-6	-4,11	2432-21-5	-0,47
12771-68-5	-2,60	146-54-3	-5,30	1982-55-4	-4,98	2432-99-7	-2,70

Continua

Continuação

127-79-7	-2,85	14686-13-6	-3,82	198-55-0	-8,80	2437-79-8	-6,51
128-13-2	-4,29	148-24-3	-2,42	2008-41-5	-3,68	2460-49-3	-2,53
128-62-1	-3,14	149-30-4	-3,18	2011-67-8	-3,80	2463-84-5	-4,31
24691-76-7	-2,56	330-54-1	-3,76	3970-62-5	-1,15	50-04-4	-4,00
24691-80-3	-3,30	330-55-2	-3,52	40186-72-9	-10,26	50-06-6	-2,34
2487-01-6	-4,47	33089-61-1	-5,47	40596-69-8	-5,19	50-11-3	-2,23
25057-89-0	-2,68	33146-45-1	-5,21	41464-39-5	-6,47	501-52-0	-1,41
2536-31-4	-4,18	33284-50-3	-5,25	41464-40-8	-6,57	50-23-7	-2,97
2539-17-5	-4,02	33284-53-6	-7,16	41464-41-9	-6,80	50-24-8	-3,21
2597-11-7	-5,46	333-41-5	-3,76	41483-43-6	-4,16	502-56-7	-2,59
26002-80-2	-5,24	3337-71-1	-1,66	4170-30-3	0,32	50-29-3	-8,08
260-94-6	-3,67	334-48-5	-3,44	41814-78-2	-2,07	50-32-8	-8,19
262-12-4	-5,31	3380-34-5	-4,46	4248-19-5	0,10	50-33-9	-3,81
2622-26-6	-3,98	339-43-5	-2,18	4248-20-8	-2,62	50-36-2	-2,23
26225-79-6	-3,42	33979-03-2	-8,71	426-13-1	-4,10	50-47-5	-3,66
2623-33-8	-1,91	34123-59-6	-3,54	4282-40-0	-4,81	50-48-6	-4,46
2642-98-0	-6,20	3424-82-6	-6,36	43121-43-3	-3,61	50-49-7	-4,19
2655-19-8	-4,24	3452-09-3	-4,24	433-97-6	-1,60	50-53-3	-5,01
26644-46-2	-4,19	34883-43-7	-5,28	434-03-7	-5,66	50-65-7	-4,70
27134-24-3	-1,93	3495-42-9	-5,43	4360-12-7	-2,82	50-70-4	1,09
271-44-3	-2,16	35065-27-1	-8,56	437-50-3	-2,93	50-71-5	-1,25
27314-13-2	-4,04	35065-28-2	-8,32	439-14-5	-3,76	507-45-9	-2,69
273-53-0	-1,16	351-83-7	-1,78	4394-00-7	-4,17	507-70-0	-2,32
27541-88-4	-5,03	352-11-4	-2,54	4396-01-4	-0,45	50-78-2	-1,61
28176-10-5	-2,28	352-34-1	-3,13	443-48-1	-1,26	508-96-3	-5,51
28299-41-4	-4,85	352-97-6	-1,51	4516-69-2	-4,48	509-86-4	-3,00
287-92-3	-2,64	35367-38-5	-6,02	452-35-7	-3,81	50-99-7	0,74
28805-78-9	-3,01	35693-92-6	-6,14	453-20-3	1,05	5104-49-4	-4,49
28911-01-5	-4,08	35693-99-3	-7,00	458-88-8	-1,50	512-04-9	-7,32
289-80-5	1,10	35694-08-7	-9,16	461-58-5	-0,31	51-21-8	-1,07
289-95-2	1,10	357-57-3	-2,09	461-72-3	-0,40	51218-45-2	-2,73
29091-05-2	-5,47	3625-25-0	-2,64	461-98-3	-1,28	5131-24-8	-3,35
291-64-5	-3,52	36322-90-4	-4,16	462-06-6	-1,80	513-36-0	-2,00
2921-88-2	-5,67	363-24-6	-2,47	464-07-3	-0,62	51-34-3	-0,48
292-64-8	-4,15	36330-85-5	-5,06	467-98-1	-1,87	514-10-3	-3,80
2939-80-2	-5,40	36378-49-1	-6,70	4695-62-9	-1,85	514-17-0	-3,59
2974-92-7	-6,39	364-62-5	-3,18	470-82-6	-1,64	51481-61-9	-1,35
298-00-0	-3,68	366-18-7	-1,42	471-03-4	-1,50	51-52-5	-2,15
298-02-2	-4,11	36653-82-4	-7,26	474-25-9	-3,64	51-55-8	-2,12
298-04-4	-4,23	372-18-9	-2,00	474-86-2	-5,28	515-64-0	-2,24
298-81-7	-3,66	3734-48-3	-5,64	475-25-2	-2,70	51-66-1	-1,30
29883-15-6	-0,77	374-01-6	0,30	475-31-0	-5,15	51-67-2	-1,12
299-42-3	-0,42	37517-28-5	-0,50	479-18-5	-0,17	517-09-9	-5,24
29973-13-5	-2,09	3766-60-7	-3,90	480-16-0	-3,08	51-79-6	0,85
299-84-3	-5,72	37680-65-2	-6,02	481-06-1	-3,09	519-05-1	-1,92
3002-23-1	-1,60	37680-73-2	-7,89	481-29-8	-4,16	51940-44-4	-2,98
3042-84-0	-3,12	3772-43-8	-3,84	481-37-8	-0,02	51953-17-4	0,59
30564-38-6	-1,11	3772-76-7	-2,54	482-05-3	-2,28	51-98-9	-4,79
309-00-2	-7,33	37764-25-3	-1,62	484-12-8	-4,31	52-01-7	-4,28
30979-48-7	-2,15	378-44-9	-3,77	4849-32-5	-2,93	52-21-1	-4,37
3121-79-7	-1,55	37924-13-3	-3,80	485-35-8	0,36	5221-53-4	-2,24
3124-46-7	-0,80	3813-05-6	-2,61	485-65-4	-2,63	52251-71-5	-6,89
31251-03-3	-8,40	382-45-6	-3,48	485-71-2	-3,07	52-31-3	-2,17
31367-46-1	-1,72	38380-02-8	-7,91	486-64-6	-2,07	5234-68-4	-3,14
31431-39-7	-3,88	38380-07-3	-8,01	488-73-3	-0,17	52-39-1	-3,85
3149-00-6	-3,05	38380-08-4	-7,82	4911-70-0	-0,89	52-43-7	-2,06
31508-00-6	-7,39	38411-22-2	-8,65	492-38-6	-2,06	5259-88-1	-2,43
315-30-0	-2,38	38444-81-4	-6,01	495-69-2	-1,68	526-08-9	-2,32
319-86-8	-4,51	38444-85-8	-6,26	496-11-7	-3,04	52645-53-1	-6,29
3248-05-3	-3,97	38444-86-9	-6,29	496-15-1	-1,04	52663-62-4	-7,05

Continua

Continuação

32598-10-0	-7,80	38444-93-8	-7,28	499-75-2	-2,08	52663-63-5	-7,42
32598-11-1	-7,25	389-08-2	-3,37	50-02-2	-3,64	52663-69-1	-7,92
327-54-8	-2,38	39196-18-4	-1,62	50-03-3	-4,30	52663-71-5	-8,30
52663-77-1	-10,41	553-26-4	-1,54	579-10-2	-0,95	597-35-3	0,04
526-73-8	-3,20	55-38-9	-4,57	579-44-2	-2,85	597-49-9	-0,85
52-67-5	-0,13	554-12-1	-0,14	579-66-8	-2,35	59756-60-4	-4,44
526-75-0	-1,43	554-57-4	-1,83	57-97-6	-7,02	597-96-6	-1,00
52-68-6	-0,22	554-84-7	-1,01	580-13-2	-4,40	598-50-5	1,13
52704-70-8	-8,60	555-37-3	-4,77	580-16-5	-2,16	598-53-8	-0,06
52712-04-6	-7,68	55-63-0	-2,22	58-08-2	-0,97	598-55-0	0,97
52712-05-7	-8,94	556-88-7	-1,37	581-40-8	-4,72	598-75-4	-0,20
52-86-8	-4,43	55702-45-9	-6,29	581-42-0	-4,89	5989-27-5	-4,00
529-86-2	-4,73	557-17-5	-0,39	58-15-1	-0,63	59-92-7	-1,60
53-06-5	-3,11	558-13-4	-3,14	58-18-4	-3,99	599-88-2	-2,82
530-78-9	-4,36	5582-82-1	-1,60	5819-08-9	-1,16	600-36-2	-1,22
53-16-7	-3,96	55880-77-8	-4,23	58-22-0	-4,07	60-12-8	-0,74
53-19-0	-6,51	5598-13-0	-4,82	5826-73-3	-1,20	60168-88-9	-4,38
533-17-5	-1,40	56030-54-7	-3,71	58-27-5	-3,03	60-18-4	-2,57
533-74-4	-2,13	56-04-2	-2,43	583-53-9	-3,50	60-29-7	-0,09
53-41-8	-4,40	56-23-5	-2,31	583-58-4	0,36	6032-29-7	-0,29
53-43-0	-4,12	56-29-1	-2,74	5836-10-2	-4,53	60-34-4	1,34
535-65-9	-3,74	563-12-2	-5,54	583-60-8	-0,94	60-35-5	1,58
535-80-8	-2,59	5635-50-7	-4,43	583-61-9	0,38	60-54-8	-3,12
53648-05-8	-3,04	56-38-2	-4,29	58-38-8	-4,40	60-57-1	-6,29
53-70-3	-8,66	564-25-0	-2,87	58-39-9	-4,16	607-67-0	-1,20
537-47-3	-2,33	56-47-3	-4,63	58-40-2	-4,30	60-80-0	0,39
53780-34-0	-3,24	56-49-5	-7,92	584-02-1	-0,24	608-21-9	-5,04
537-92-8	-2,09	56-53-1	-4,35	58-55-9	-1,37	60-87-7	-4,26
53808-87-0	-2,10	56-54-2	-3,37	58-63-9	-1,23	608-93-5	-5,65
538-32-9	-0,95	56558-16-8	-7,32	58-89-9	-4,60	60-89-9	-4,74
53-86-1	-4,62	565-60-6	-0,72	589-18-4	-1,20	60-99-1	-4,22
538-68-1	-4,64	565-67-3	-0,70	58-93-5	-2,62	611-14-3	-3,21
538-93-2	-4,12	56-72-4	-5,38	58-94-6	-3,05	613-12-7	-6,96
539-03-7	-2,84	56-75-7	-2,11	589-55-9	-1,40	613-13-8	-5,17
5392-40-5	-2,06	56-81-5	1,12	589-82-2	-1,47	614-28-8	-2,27
539-82-2	-1,75	56-85-9	-0,55	589-90-2	-4,47	614-29-9	-2,52
539-86-6	-0,83	569-31-3	-1,89	589-91-3	-0,88	614-61-9	-2,16
540-36-3	-1,97	57-10-3	-6,81	589-93-5	0,40	614-80-2	-2,24
540-54-5	-1,47	57-11-4	-5,68	590-12-5	-1,32	615-15-6	-1,96
54-11-5	0,79	57-13-6	0,96	5902-51-2	-2,48	615-54-3	-4,50
541-33-3	-2,40	571-58-4	-4,14	590-27-2	-3,22	61-57-4	-3,22
541-48-0	1,08	571-61-9	-4,74	590-36-3	-0,49	615-77-0	-0,80
541-73-1	-3,04	57-24-9	-3,32	591-07-1	-0,90	616-20-6	-2,63
542-55-2	-1,01	57-27-2	-3,28	591-17-3	-3,52	616-44-4	-2,39
542-69-8	-2,96	57-41-0	-3,99	591-22-0	0,38	616-45-5	1,07
542-76-7	-0,29	57-42-1	-1,89	591-24-2	-1,87	6164-98-3	-2,86
54-31-9	-3,66	57-43-2	-2,57	591-49-1	-3,27	61-68-7	-3,78
543-49-7	-1,55	57-44-3	-1,39	591-50-4	-2,78	618-51-9	-3,27
543-59-9	-2,73	57-48-7	0,64	591-78-6	-0,80	619-02-3	-2,18
544-10-5	-3,12	57-53-4	-1,67	591-93-5	-2,09	619-45-4	-1,59
544-25-2	-2,15	575-41-7	-4,29	592-35-8	-0,66	620-71-3	-1,92
544-63-8	-5,33	57-62-5	-2,88	592-41-6	-3,23	621-82-9	-2,48
547-75-1	-4,35	576-26-1	-1,31	592-42-7	-2,68	621-84-1	-0,35
548-00-5	-3,66	57-63-6	-4,30	59-31-4	-2,14	62-23-7	-2,80
54-85-3	0,01	57-67-0	-1,99	594-36-5	-2,51	622-45-7	-1,67
551-06-4	-4,60	57-68-1	-2,27	594-60-5	-0,41	622-50-4	-3,25
55-21-0	-0,96	57-74-9	-6,86	59-46-1	-1,40	622-96-8	-3,11
55215-17-3	-7,43	578-06-3	-4,22	594-83-2	-0,72	623-37-0	-0,80
55215-18-4	-8,42	57-83-0	-4,43	59-50-7	-1,57	623-42-7	-0,82
552-32-9	-1,91	57837-19-1	-1,60	595-33-5	-5,35	6236-05-1	-2,19

Continua

Continuação

552-58-9	-3,62	57-85-2	-5,37	595-41-5	-0,85	624-24-8	-1,36
552-63-6	-0,93	578-67-6	-2,54	59-66-5	-2,36	62-44-2	-2,37
55312-69-1	-7,21	57-91-0	-4,84	59-67-6	-0,84	624-95-3	-0,50
625-06-9	-0,92	6893-02-3	-5,22	75-30-9	-2,09	79-41-4	0,01
625-23-0	-1,08	68-94-0	-2,29	75-34-3	-1,29	79-46-9	-0,62
625-25-2	-1,72	68-96-2	-4,71	75-35-4	-1,64	79-57-2	-3,14
625-29-6	-2,63	691-37-2	-3,24	75-50-3	0,84	80-05-7	-2,82
62-53-3	-0,41	69-23-8	-4,15	75-56-9	-0,59	80-08-0	-2,82
625-54-7	-0,55	693-02-7	-2,36	7568-93-6	-0,48	80-35-3	-3,28
62-56-6	0,32	69-33-0	-1,95	7581-97-7	-2,70	80-60-4	0,31
626-48-2	-1,26	693-54-9	-3,31	75-83-2	-3,55	80-62-6	-0,80
626-64-2	1,02	694-80-4	-3,19	75-84-3	-0,40	80-74-0	-3,59
626-67-5	0,23	694-85-9	0,96	75-85-4	0,08	81-07-2	-1,64
626-89-1	-1,14	6966-78-5	-2,36	759-73-9	-0,96	81-25-4	-3,37
626-93-7	-0,89	69-72-7	-1,82	76-01-7	-2,60	81451-15-2	-2,71
627-08-7	-1,34	6988-21-2	-1,57	76-13-1	-3,04	81-81-2	-3,89
627-19-0	-1,64	700-12-9	-4,00	76-14-2	-3,12	82-02-0	-2,40
627-59-8	-1,38	7012-37-5	-6,21	76-22-2	-1,99	821-55-6	-2,58
628-28-4	-0,99	70-47-3	-0,74	76-24-4	-1,99	82-68-8	-5,82
628-32-0	-0,66	70-55-3	-1,74	76-25-5	-4,32	83-07-8	-0,62
628-41-1	-1,97	70-69-9	-2,63	76-44-8	-6,32	83-26-1	-4,11
628-63-7	-1,89	708-79-2	-1,56	76466-16-5	-2,22	83-32-9	-4,63
628-71-7	-3,01	709-98-8	-3,00	76-57-3	-1,52	83-34-1	-2,42
629-04-9	-4,43	71-00-1	-0,53	76674-21-0	-3,37	834-12-8	-3,04
629-05-0	-3,66	71-30-7	-1,14	76-74-4	-2,52	83-43-2	-2,99
629-06-1	-3,99	71-36-3	0,00	76-75-5	-3,36	83-44-3	-3,95
629-14-1	-0,77	71-41-0	-0,60	76-76-6	-2,21	83-49-8	-3,82
630-20-6	-2,18	71-43-2	-1,64	7681-93-8	-3,21	83-56-7	-2,92
6305-71-1	-1,60	71-55-6	-2,00	76-93-7	-2,21	83-79-4	-4,42
63-05-8	-3,69	72-14-0	-2,43	77-02-1	-1,71	83-88-5	-3,68
632-22-4	0,94	72-19-5	-0,09	77-06-5	-1,84	84-06-0	-4,70
63-25-2	-3,28	72-20-8	-6,18	77-09-8	-2,90	84-17-3	-4,95
634-66-2	-4,57	723-46-6	-2,62	77-21-4	-2,34	846-49-1	-3,60
634-90-2	-4,63	72-43-5	-6,89	77-27-0	-3,46	84-65-1	-5,19
636-26-0	-2,45	72-48-0	-2,78	77-28-1	-1,64	84-66-2	-2,35
63-68-3	-0,42	72-54-8	-7,20	77-47-4	-5,18	84-69-5	-4,66
63-74-1	-1,36	72-55-9	-6,90	77-65-6	-2,68	84-74-2	-4,40
63-75-2	0,81	72762-00-6	1,02	77-74-7	-0,38	85-01-8	-5,26
638-42-6	-1,47	7286-84-2	-3,26	779-02-2	-5,89	85-02-9	-3,36
64-19-7	1,22	7287-19-6	-4,10	77-92-9	0,51	853-23-6	-4,46
643-12-9	0,35	7294-05-5	-1,27	781-43-1	-6,57	853-34-9	-3,27
646-04-8	-2,54	729-99-7	-2,44	78-40-0	0,43	85-34-7	-3,08
64-77-7	-3,39	7307-04-2	-1,63	78-48-8	-5,14	85-41-6	-2,61
64-85-7	-3,75	73-22-3	-1,23	78-59-1	-1,06	85-44-9	-1,39
651-06-9	-2,58	73-24-5	-2,12	786-19-6	-5,74	85-68-7	-5,64
65-45-2	-1,76	73-48-3	-3,59	78-70-6	-1,99	85-79-0	-3,70
65-71-4	-1,52	73-49-4	-3,29	78-75-1	-2,15	86-30-6	-3,75
65-85-0	-1,55	738-70-5	-2,86	78-77-3	-2,43	86-57-7	-3,54
65-86-1	-1,93	74-11-3	-3,31	78-79-5	-2,03	86-73-7	-4,91
65882-61-3	-2,77	741-58-2	-4,20	78-83-1	0,04	86-74-8	-4,97
66-02-4	-2,86	74472-42-7	-7,66	78-86-4	-1,96	86-87-3	-2,65
66-22-8	-1,48	74-88-4	-1,00	78-87-5	-1,60	87-17-2	-3,59
66-25-1	-1,30	74-95-3	-1,17	789-02-6	-6,62	87-20-7	-3,16
664-95-9	-4,21	74-96-4	-1,09	78-92-2	0,43	872-50-4	1,00
66-81-9	-1,13	74-97-5	-0,89	78-93-3	0,52	872-55-9	-2,59
67-20-9	-3,38	75-03-6	-1,60	79-00-5	-1,48	87-33-2	-2,63
67-66-3	-1,17	75-08-1	-0,60	79-01-6	-1,96	87-47-8	-2,09
67-72-1	-3,67	75-09-2	-0,63	79-06-1	0,96	87-51-4	-2,07
67-78-7	-4,13	751-97-3	-1,42	79-07-2	-0,02	87-58-1	-3,46
68-22-4	-4,57	75-25-2	-1,91	79-11-8	0,93	87-61-6	-4,00

Continua

Continuação

68-35-9	-3,51	75-26-3	-1,59	79-20-9	0,52	87-68-3	-4,91
684-93-5	-0,85	75-27-4	-1,54	79-31-2	0,28	87-72-9	0,39
6876-23-9	-4,33	75-29-6	-1,41	79-34-5	-1,74	87-86-5	-4,28
88-04-0	-2,80	91-57-6	-3,77	94-82-6	-3,67	96-88-8	-1,55
88-06-2	-2,34	91-58-7	-4,14	94-96-2	-0,54	97-23-4	-3,95
88-09-5	-0,81	91-64-5	-1,89	95-06-7	-3,39	97-53-0	-1,56
88-13-1	-1,47	91-66-7	-3,03	95-14-7	-0,78	97-59-6	-1,60
88-14-2	-0,48	91-75-8	-2,60	95-16-9	-1,50	97-77-8	-4,86
88-19-7	-2,02	91-80-5	-2,64	95-47-6	-2,80	977-79-7	-5,27
886-50-0	-4,00	91-81-6	-2,64	95-48-7	-0,62	97-95-0	-1,17
88-67-5	-2,73	91-94-1	-4,92	95-49-8	-3,52	98-01-1	-0,10
88-72-2	-2,33	92-24-0	-8,60	95-50-1	-3,05	98-06-6	-3,66
88-74-4	-1,96	92-36-4	-3,68	95-51-2	-1,52	98-10-2	-1,56
88-75-5	-1,74	92-52-4	-4,31	95-53-4	-0,85	98-16-8	-1,47
88-85-7	-3,38	92-69-3	-3,48	95-54-5	-0,42	98-54-4	-2,41
88-96-0	-2,92	92-87-5	-2,70	95-55-6	-0,72	98-82-8	-3,27
88-99-3	-2,11	92-94-4	-7,11	95-57-8	-1,06	98-85-1	-0,92
89-78-1	-2,53	93-37-8	-1,94	95-63-6	-3,31	98-86-2	-1,28
89-83-8	-2,22	93-55-0	-1,83	95-65-8	-1,41	98-89-5	-1,45
90-00-6	-1,36	93-58-3	-1,85	957-51-7	-2,98	98-92-0	0,61
90-01-7	-0,29	93-60-7	-0,46	95-76-1	-3,24	98-95-3	-1,80
90-02-8	-0,86	936-40-3	0,06	95-87-4	-1,54	98-96-4	-0,91
90-05-1	-1,96	93-65-2	-2,55	95-95-4	-2,21	99-03-6	-1,28
90-11-9	-4,35	93-76-5	-2,96	96-09-3	-1,60	99-04-7	-2,14
90-12-0	-3,70	93-89-0	-2,32	96-12-8	-2,38	99-08-1	-2,44
90-13-1	-3,93	939-27-5	-4,29	96-14-0	-3,68	99-09-2	-2,19
90-15-3	-2,22	94-09-7	-2,32	96-18-4	-1,92	99-24-1	-1,24
90-39-1	-1,89	94-12-2	-2,33	96-23-1	-0,11	99-66-1	-1,86
90-43-7	-2,39	94-19-9	-1,94	96-33-3	-0,22	99-76-3	-1,78
91-01-0	-2,55	94-20-2	-3,03	96-37-7	-3,30	99-87-6	-3,77
91-15-6	-2,38	94-23-5	-2,71	96-45-7	-0,71	99-92-3	-1,61
91-18-9	0,02	94-26-8	-2,72	96-47-9	0,11	99-94-5	-2,60
91-20-3	-3,60	94-62-2	-3,46	96-48-0	1,07	99-96-7	-1,41
91-22-5	-1,30	94-74-6	-2,23	96-50-4	-0,36	99-99-0	-2,49
91-23-6	-1,96	94-75-7	-2,51	968-81-0	-2,06		