**Desenvolvimento de Modelos *In Silico* de Propriedades de ADME
Para a Triagem de Novos Candidatos a Fármacos**

Tiago Luiz Moda

Dissertação apresentada ao Instituto de Física de São Carlos, da Universidade de São Paulo, para obtenção do título de Mestre em Ciências: Física Aplicada - Opção Física Biomolecular

Orientador: Prof. Dr. Adriano Defini Andricopulo

São Carlos, Fevereiro de 2007

Moda, Tiago Luiz,

Desenvolvimento de Modelos *In Silico* de Propriedades de ADME Para a Triagem de Novos Candidatos a Fármacos. Tiago Luiz Moda, São Carlos, 2007. 82 p.

Dissertação (Mestrado) – Área de Física Aplicada
Universidade de São Paulo, São Carlos, 2007.

“It’s hard to give up on a drug, if you know that thousands of people will die without it.”

Amgen’s Paul Reider, helped the Merck Company develop Crixivan, the first protease inhibitor to treat AIDS patients.

Science **2005**, 309, 721

AGRADECIMENTOS

- ✓ Ao meu orientador, Prof. Adriano D. Andricopulo, pela oportunidade, orientação e confiança, sobretudo por me mostrar o comprometimento e responsabilidade com que os trabalhos em Química Medicinal devem ser realizados.
- ✓ Ao Prof. Carlos A. Montanari, pelas discussões sobre Quimiometria.
- ✓ Aos amigos com os quais tive valiosas discussões sobre Química Medicinal e Computacional, sobretudo, ao Andrei Leitão, Káthia Honório, Livia Salum e Ricardo Nicoluci.
- ✓ Ao Matheus Postigo, pela ajuda com as figuras desta dissertação.
- ✓ À turma da sala 8.
- ✓ Ao Instituto de Física de São Carlos da Universidade de São Paulo, aos professores e funcionários pela colaboração e excelente suporte para realização deste trabalho.
- ✓ Aos professores e amigos do Centro de Biotecnologia Molecular Estrutural – CEPID-FAPESP, que contribuíram positivamente para a realização deste trabalho.
- ✓ À CAPES, pela bolsa de mestrado.

SUMÁRIO

Lista de Abreviaturas	viii
Lista de Figuras	x
Lista de Tabelas	xii
Resumo	xiv
Abstract	xv

CAPÍTULO I

I. INTRODUÇÃO	2
1. Química Medicinal e a Descoberta de Novos Fármacos	2
2. Fármacos e a Indústria Farmacêutica	3
3. Desenvolvimento de NCEs	5
4. Farmacodinâmica e Farmacocinética	7
5. Absorção, Distribuição, Metabolismo e Excreção (ADME)	9
5.1 Absorção (A)	9
5.2 Metabolismo (M) e Excreção (E)	11
5.3 Biodisponibilidade	14
5.4 Distribuição (D)	15
5.5 Ligação às Proteínas Plasmáticas	16
6. Modelagem <i>In Silico</i> de Propriedades de ADME	17

7. Holograma QSAR	19
8. VolSurf	24

CAPÍTULO II

II. OBJETIVOS	28
----------------------	----

CAPÍTULO III

III. PARTE EXPERIMENTAL	30
--------------------------------	----

1. Modelagem de QSAR	31
2. Conjuntos de Padrões Dados	31
3. Holograma QSAR	32
4. Método VolSurf	33

CAPÍTULO IV

IV. RESULTADOS E DISCUSSÃO	35
-----------------------------------	----

1. Conjunto Padrão de Dados de Biodisponibilidade	36
2. Holograma QSAR: Modelo de Biodisponibilidade	40
3. Método VolSurf: Modelo de Biodisponibilidade	50

4. Conjunto Padrão de Dados de Ligação às Proteínas Plasmáticas	57
5. Holograma QSAR: Modelo de Ligação Às Proteínas Plasmáticas	59
6. Método VolSurf: Modelo de Ligação às Proteínas Plasmáticas	64

CAPÍTULO V

V. CONCLUSÕES	70
----------------------	----

CAPÍTULO VI

VI. REFERÊNCIAS BIBLIOGRÁFICAS	73
---------------------------------------	----

LISTA DE ABREVIATURAS

2 D	Bidimensional
3D	Tridimensional
A	Átomos
ADME	Absorção, Distribuição, Metabolismo e Excreção
B	Ligação
C	Conectividade
CBME	Centro de Biotecnologia Molecular Estrutural
CEPID	Centros de Pesquisa, Inovação e Difusão
Ch	Quiralidade
CRC	Algoritmo Cíclico da Verificação de Redundância
CYP	Citocromo P450 Oxidase
DA	Doador e Aceptor
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
FFD	Análise de Planejamento Fatorial
H	Hidrogênio
HBA	Aceptores de Ligação Hidrogênio
HBD	Doadores de Ligação Hidrogênio
HQSAR	Holograma QSAR
IFSC	Instituto de Física de São Carlos
IND	Investigação de um Novo Fármaco
Log <i>P</i>	Coeficiente de Partição de uma Substância em Sistema Octanol/Água
LQMC	Laboratório de Química Medicinal e Computacional
MIF	Campos Moleculares de Interação
MM	Massa Molar
MSS	Planilha Molecular

NCE	Nova Entidade Química
NDA	Requisição de um Novo Fármaco
P&D	Pesquisa e Desenvolvimento
PCA	Análise de Componentes Principais
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
SAR	Relações Entre a Estrutura e Atividade
SPR	Relações Entre a Estrutura e Propriedade
VolSurf	Volume & Superfície
WDI	<i>World Drug Index</i>

LISTA DE FIGURAS

Figura 1	Etapas do processo de descoberta e desenvolvimento de fármacos	3
Figura 2	Principais razões de insucesso de NCEs nas fases clínicas de desenvolvimento	6
Figura 3	Estratégias em P&D de fármacos: (a) foco na otimização de potência; (b) foco na otimização integrada de potência e propriedades de ADME	8
Figura 4	Estrutura da citocromo CYP3A4 com o grupo heme representado no centro	13
Figura 5	Representação da biodisponibilidade oral em humanos	14
Figura 6	Esquema ilustrativo do processo de obtenção de hologramas moleculares	20
Figura 7	Etapas envolvidas no método HQSAR	23
Figura 8	Cálculo dos MIF gerados pela sonda água (OH ₂) no método VolSurf	24
Figura 9	Representação do espaço químico-biológico do conjunto de dados	40
Figura 10	Representação gráfica da distribuição dos valores de biodisponibilidade	43
Figura 11	Gráfico de HQSAR dos valores experimentais e preditos de biodisponibilidade para os 302 compostos do conjunto de dados	49
Figura 12	Relação entre os valores experimentais e preditos de biodisponibilidade para os 302 compostos do conjunto de dados	54
Figura 13	Coeficientes PLS dos descritores VolSurf	55
Figura 14	Representação da contribuição dos descritores na biodisponibilidade dos fármacos isonizida e atorvastatina	56
Figura 15	Relação entre os valores experimentais e preditos de PPB para os 304 compostos do conjunto de dados	63

Figura 16	Gráfico dos valores experimentais e preditos de PPB para os 304 compostos do conjunto de dados do conjunto de dados	68
Figura 17	Coeficientes PLS dos descritores VolSurf	69

LISTA DE TABELAS

Tabela 1	Parâmetros de distinção de fragmentos no método HQSAR	22
Tabela 2	Definição das principais sondas químicas usadas no método Volsurf	25
Tabela 3	Definição de descritores físico-químicos calculados pelo método Volsurf	26
Tabela 4	Conjunto padrão de dados de biodisponibilidade oral em humanos	36
Tabela 5	Faixas de distribuição de baixa, moderada e alta biodisponibilidade dos compostos do conjunto de dados	42
Tabela 6	Resultados de HQSAR para as várias combinações de distinção de fragmentos empregando o tamanho padrão do fragmento (4-7)	45
Tabela 7	Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando a melhor distinção observada (A/B/C/Ch)	46
Tabela 8	Valores experimentais e preditos de biodisponibilidade, associados aos resíduos para os compostos do conjunto teste	48
Tabela 9	Análise VolSurf para as várias combinações de sondas empregando o espaçamento padrão da caixa (0,5 Å)	51
Tabela 10	Influência do espaçamento da caixa nos parâmetros estatísticos do modelo VolSurf	52
Tabela 11	Comparação dos modelos 3 e 3/FFD nos parâmetros estatísticos do modelo VolSurf (PLS-LOO)	52
Tabela 12	Valores experimentais e preditos de biodisponibilidade, associados aos resíduos para os compostos do conjunto teste	53
Tabela 13	Conjunto Padrão de Dados de Ligação às Proteínas Plasmáticas	57
Tabela 14	Efeito da distinção de fragmentos no modelo de HQSAR para PPB	60
Tabela 15	Influência do tamanho dos fragmentos nos parâmetros estatísticos do modelo 4 (distinção de fragmentos A/B/C/Ch)	61

Tabela 16	Valores experimentais e preditos de PPB, associados aos resíduos para os compostos do conjunto teste	62
Tabela 17	Análise VolSurf para as várias combinações de sondas empregando o espaçamento padrão da caixa (0,5 Å)	64
Tabela 18	Influência do espaçamento da caixa nos parâmetros estatísticos do modelo VolSurf para PPB	65
Tabela 19	Comparação dos modelos 2 e 2/FFD nos parâmetros estatísticos do modelo VolSurf (PLS-LOO)	66
Tabela 20	Valores experimentais e preditos de PPB, associados aos resíduos para os compostos do conjunto teste	67

RESUMO

As ferramentas de modelagem molecular e de estudos das relações quantitativas entre a estrutura e atividade (QSAR) ou estrutura e propriedade (QSPR) estão integradas ao processo de planejamento de fármacos, sendo de extremo valor na busca por novas moléculas bioativas com propriedades farmacocinéticas e farmacodinâmicas otimizadas. O trabalho em Química Medicinal realizado nesta dissertação de mestrado teve como objetivo estudar as relações quantitativas entre a estrutura e as propriedades farmacocinéticas biodisponibilidade oral e ligação às proteínas plasmáticas.

Para a realização deste trabalho, conjuntos padrões de dados foram organizados para as propriedades biodisponibilidade e ligação às proteínas plasmáticas 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 os métodos holograma QSAR e VolSurf.

Os modelos finais de HQSAR e VolSurf gerados neste trabalho possuem elevada consistência interna e externa, apresentando bom poder de correlação e predição das propriedades alvo. 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.

ABSTRACT

Molecular modeling tools and quantitative structure-activity relationships (QSAR) or structure-property (QSPR) are integrated into the drug design process in the search for new bioactive molecules with good pharmacokinetic and pharmacodynamic properties. The Medicinal Chemistry work carried out in this Master's dissertation concerned studies of the quantitative relationships between chemical structure and the pharmacokinetic properties oral bioavailability and plasma protein binding.

In the present work, standard data sets for bioavailability and plasma protein binding were organized encompassing the structural information and corresponding pharmacokinetic data. The created data sets established the scientific basis for the development of predictive models using the hologram QSAR and VolSurf methods.

The final HQSAR and VolSurf models possess high internal and external consistency with good correlative and predictive power. 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.

CAPÍTULO I
INTRODUÇÃO

I. INTRODUÇÃO

1. Química Medicinal e a Descoberta de Novos Fármacos

A indústria farmacêutica vem passando por um momento crucial na sua história, marcado por grandes transformações guiadas pelas inovações científicas e tecnológicas. Os modelos tradicionais, baseados principalmente em química e biologia básica, estão progressivamente dando lugar a novos paradigmas fundamentados na integração de uma combinação de especialidades atuando em caráter multidisciplinar.¹ Essa nova geração de modelos e conceitos reflete o papel de destaque da Química Medicinal moderna nos processos inovativos da gênese planejada de fármacos.

A grande evolução da Química Medicinal tem permitido a descoberta de notáveis inovações terapêuticas, proporcionando sensíveis benefícios em dois componentes principais: saúde física e mental. As novas tecnologias a serviço do processo de descoberta e desenvolvimento de fármacos levaram, de forma inegável, a melhorias significativas na qualidade de vida das diversas populações no mundo.¹⁻³

Os métodos em quimio- e bioinformática estão integrados ao planejamento de fármacos.⁴ O seu emprego nos estudos em Química Medicinal vão desde a identificação, seleção e otimização de moléculas candidatas a novas entidades químicas (NCEs, do inglês, *New Chemical Entities*) com elevado potencial terapêutico. A integração desses métodos ao trabalho químico e biológico experimental é um requerimento essencial para a geração de novas moléculas bioativas qualificadas a cerca de uma série complexa de propriedades farmacodinâmicas e farmacocinéticas.^{5,6}

2. Fármacos e a Indústria Farmacêutica

O complexo processo de descoberta e desenvolvimento de uma NCE, ilustrado na Figura 1, é longo e envolve altos investimentos.⁷ Desde a concepção até a introdução de um único fármaco no mercado farmacêutico, são investidos de 12 a 15 anos em P&D (Pesquisa e Desenvolvimento, termo do inglês, *R&D, Research and Development*), com custos totais da ordem de US\$ 500-880 milhões, podendo alcançar em alguns casos cifras superiores a US\$ 1 bilhão.⁸⁻¹¹

Figura 1. Etapas do processo de descoberta e desenvolvimento de fármacos.⁹

Quando aprovadas na fase pré-clínica (fase de descoberta ou pesquisa básica), as NCEs são submetidas à Administração de Alimentos e Medicamentos dos Estados Unidos (FDA, do inglês, *Food and Drug Administration*), para uma proposta de

investigação de um novo fármaco (IND, do inglês, *Investigational New Drug*). Após a aprovação de uma IND, são iniciados os testes clínicos em humanos, examinando entre outras, a farmacocinética, dosagem, segurança e eficácia do fármaco, gerando uma série de dados que servem como documentos de suporte para uma posterior requisição de aprovação de um novo fármaco (NDA, do inglês, *New Drug Application*).

A sobrevivência, crescimento e desenvolvimento contínuo da indústria farmacêutica requerem investimentos significativos em programas de P&D.¹² Neste sentido, a capacidade da indústria em alocar recursos em P&D está diretamente ligada a sua capacidade em gerar receitas (capital bruto) através de um arsenal atrativo de fármacos.² Embora a indústria farmacêutica seja um dos setores mais rentáveis do mercado, o número de novos fármacos vem diminuindo consideravelmente na última década.^{13,14} É estimado que, de cada cinco mil compostos testados na fase pré-clínica, apenas cinco se tornam NCEs e avançam nas fases clínicas de desenvolvimento, sendo que, apenas uma obtém sucesso e alcança o mercado farmacêutico.^{15,16} Por essa razão, existe uma forte demanda por novos fármacos de variadas classes terapêuticas, capazes de gerar grandes receitas para as suas companhias.

O fator crítico para o sucesso da indústria farmacêutica é a descoberta de fármacos da categoria *blockbuster* (termo do inglês usado para designar os fármacos com vendas anuais superiores a US\$ 1 bilhão), que são os maiores responsáveis pelo crescimento da indústria a nível mundial.⁶

Atualmente, a categoria *blockbuster* inclui diversos fármacos de enorme sucesso como o agente hipolipemiante Lípitor® (Atorvastatina cálcica), da Pfizer; o antitrombótico Plavix® (Clopidogrel), da Sanofi Aventis; o antiulceroso Nexium®

(Esomeprazol), da AstraZeneca; o antiasmático Seretide® (Fluticasona + Salmeterol), da GlaxoSmithKline; o hipolipemiante Zocor® (Sinvastatina), da Merck; o anti-hipertensivo Novarsc® (Anlodipidina), da Pfizer; o antipsicótico Zyprexa® (Olanzapina), da Lilly; o antidepressivo Effexor® (Venlafaxine), da Wyeth; entre outros.

3. Desenvolvimento de NCEs

O processo de descoberta e desenvolvimento de novos fármacos ilustrado na Figura 1 é dividido em duas grandes fases: pré-clínica (ou descoberta) e clínica (ou desenvolvimento).^{8,14}

A fase pré-clínica envolve várias etapas de pesquisa básica: i. abordagem do problema médico; ii. seleção de um alvo; iii. desenvolvimento de estratégias; iv. estudos *in silico*; v. síntese e caracterização de moléculas; vi. ensaios *in vitro*; vii. ensaios *in vivo* (modelos em animais); viii. avaliação toxicológica básica; iv. síntese em escala industrial de princípios ativos; x. proteção da propriedade intelectual de NCEs e análise do mercado farmacêutico; xi. documentação e formulação da IND junto ao FDA.

A fase clínica (FC) envolve as etapas de desenvolvimento de NCEs. i. FC I: testes de dosagem dos princípios ativos em humanos saudáveis e avaliação preliminar dos níveis de toxidez; ii. FC II: testes farmacológicos em um número reduzido de pacientes para avaliação da dosagem, eficácia e segurança da composição; iii. FC III: testes em um número significativo de pacientes para avaliação das propriedades

farmacocinéticas, além de eficácia e segurança; e, por fim, iv. FC IV: monitoramento pós-lançamento do novo fármaco.⁹

O processo de desenvolvimento de fármacos é de alto risco. É estimado que 79% das NCEs em desenvolvimento clínico são eliminadas do processo.⁸ Dentre estas, 39% são descartadas por não apresentarem as propriedades farmacocinéticas requeridas para um fármaco, conforme a Figura 2.¹⁷

Figura 2. Principais razões de insucesso de NCEs nas fases clínicas de desenvolvimento.¹⁷

Existe grande interesse por parte da indústria farmacêutica em novas tecnologias úteis no planejamento de fármacos. Estas, idealmente, devem ser capazes de contribuir no processo de otimização de propriedades dos candidatos a NCEs, tendo como finalidade a redução do tempo e dos altos custos envolvidos no processo.^{18,19}

Modelos *in silico* (expressão usada com o significado de “executado em computador”) estão sendo empregados nos estágios iniciais de P&D na seleção e otimização de moléculas com maior potencial de desenvolvimento. Estes modelos devem ser integrados aos ensaios *in vitro* e *in vivo* com o propósito de diminuir a quantidade de moléculas avaliadas, maximizando as chances de sucesso na identificação de NCEs qualificadas.²⁰⁻²⁴

4. Farmacodinâmica e Farmacocinética

Por muitas décadas, o requisito principal para que uma molécula alcançasse o processo de desenvolvimento era a fase farmacodinâmica, ou seja, os candidatos deveriam possuir propriedades otimizadas como potência e seletividade.²⁵ Desta maneira, o foco predominante era nos estudos das interações intermoleculares responsáveis pelo processo de reconhecimento molecular.²⁶ Contudo, para que os efeitos farmacológicos observados resultem na ação terapêutica dos fármacos é necessário que os princípios ativos possam estar disponíveis por determinado período de tempo na periferia dos sítios alvos de ação. Os estudos acerca dos diversos caminhos que o princípio ativo deve percorrer no complexo compartimento biológico, desde a sua administração por via oral até atingir a circulação sistêmica para exercer o

efeito terapêutico desejado, fazem parte da chamada fase farmacocinética.²⁷ É amplamente reconhecida pela indústria farmacêutica a importância dos estudos de propriedades de ADME, que se referem aos processos farmacocinéticos de absorção (A), distribuição (D), metabolismo (M) e excreção (E), nos estágios iniciais do processo de descoberta de fármacos.

É fundamental, portanto, que as estratégias de planejamento de NCEs integrem a otimização de múltiplas propriedades farmacodinâmicas e farmacocinéticas, conforme ilustrado na Figura 3.^{28,29}

Figura 3. Estratégias em P&D de fármacos: (a) foco na otimização de potência; (b) foco na otimização integrada de potência e propriedades de ADME.⁸¹

Muitos esforços tem sido investidos em um campo científico considerado ainda emergente, o da modelagem *in silico* de propriedades de ADME.^{25,27,30} As propriedades farmacocinéticas podem ser estudadas através do uso de métodos como os estudos das relações quantitativas entre a estrutura e atividade (QSAR), e os estudos

das relações quantitativas entre a estrutura e propriedade (QSPR), entre muitos outros.²⁹

5. Absorção, Distribuição, Metabolismo e Excreção (ADME)

Por questões de mercado, conveniência e segurança, a maioria dos fármacos é administrada por via oral. O tratamento com medicamentos por esta via implica a introdução de um fármaco no organismo (processo de administração), de modo a que possa chegar ao sangue (absorção) e dirigir-se para o sítio específico de ação (distribuição). O fármaco pode ser também transformado em outras substâncias químicas (metabolismo), mas deve, por fim, abandonar o organismo (excreção), principalmente pela urina.

5.1 Absorção (A)

Após a administração, o fármaco deve se dissolver e ser solubilizado no trato gastrointestinal para poder ser absorvido através da membrana intestinal.^{17,31} A absorção ocorre por dois mecanismos principais. O primeiro, chamado de mecanismo transcelular, é a simples difusão passiva através das membranas celulares. O segundo, chamado de mecanismo paracelular, é a passagem entre as células intestinais. A absorção também pode ocorrer por transporte ativo, que requer energia e é mediado por carregadores.^{30,32}

Entre as propriedades moleculares que influenciam a absorção se destacam a 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 a uma alta lipofilia, entretanto, moléculas muito hidrofílicas (baixa lipofilia) também apresentam problemas de permeabilidade, resultando em uma absorção inadequada.^{28,33} A lipofilia, associada ao coeficiente de partição (Log *P*, partição de uma substância entre as fases aquosa e orgânica no sistema água/octanol), representa a passagem do meio extracelular para o meio intracelular através da bicamada lipídica.²⁴ O balanço ideal entre solubilidade e lipofilia é, portanto, um ponto fundamental no planejamento de NCEs.

Diversos estudos foram realizados para demonstrar o perfil dos fármacos disponíveis no mercado em relação à absorção e permeabilidade, a porção “A” de ADME. Um destes, desenvolvido por Lipinski e colaboradores, conhecido como “Regra dos cinco” (do inglês, *Rule of Five*), merece especial destaque.³⁴ Este estudo foi baseado em um conjunto de dados de 2245 NCEs coletadas do *World Drug Index* (WDI) e estabeleceu de maneira pioneira que candidatos a fármacos apresentariam baixa absorção e permeabilidade se infringissem duas das seguintes propriedades:

- Massa molar (MM) > 500
- Número de doadores de ligação hidrogênio (HBD) > 5
- Número de aceptores de ligação hidrogênio (HBA) > 10
- Coeficiente de partição octanol/água (Log *P*) > 5

A partir destes resultados, diversos outros estudos foram derivados para caracterizar outras propriedades moleculares importantes em grupos de fármacos.^{35,36} Uma importante inclusão foi o reconhecimento de que a área de superfície polar (PSA, do inglês, *polar surface area*) era um dos fatores que poderiam ser considerados na análise da permeabilidade celular (no limite, $PSA \leq 140 \text{ \AA}^2$).³⁷

5.2 Metabolismo (M) e Excreção (E)

O organismo humano dispõe de mecanismos próprios para eliminar xenobióticos, incluindo, os fármacos. O metabolismo é o processo químico através do qual o organismo altera um fármaco. O fígado é o principal, mas não o único lugar onde os fármacos são metabolizados. A excreção se refere aos processos que o organismo utiliza para eliminar um fármaco.³⁸

Após ser absorvido no intestino, o fármaco entra na via hepática que leva ao fígado onde pode sofrer o metabolismo de primeira passagem, facilitando o processo de excreção através de biotransformações específicas na estrutura química do xenobiótico, gerando os produtos do metabolismo, os metabólitos. A via principal de excreção é a urina.³⁰

O metabolismo de fármacos ocorre em duas fases distintas bem caracterizadas. A primeira fase (fase 1) engloba reações catalisadas pela família de isoformas da enzima citocromo P450 oxidase (identificada como CYP ou CYP450). Ocorrem predominantemente, reações de oxidação, redução e hidrólise, levando a metabólitos mais hidrofílicos. Neste processo, os metabólitos gerados podem ser inativos ou exibir

graus similares ou diferentes de potência ou toxicidade em comparação com o fármaco original. Os pró-fármacos são um caso especial, onde o princípio é administrado em uma forma inativa que sofre biotransformação liberando o princípio ativo, que exercerá a ação farmacológica através da modulação seletiva do alvo apropriado.^{39,40}

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 aminoácidos, formando conjugados mais hidrossolúveis, geralmente inativos, que são eliminados preferencialmente pela urina, ou então excretados pela bile e eliminados nas fezes.^{26,30}

A CYP é usada de forma genérica para identificar uma superfamília relacionada de enzimas oxidativas (EC 1.14). Sete das cinquenta e sete isoformas humanas são responsáveis pelo metabolismo de mais de 90% dos fármacos atualmente em uso clínico. Estas são: CYP1A2, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1 e CYP3A4.⁴¹

A principal dessas isoformas é a CYP3A4 (citocromo P450 3A4, EC 1.14.14.1), responsável pelo metabolismo de aproximadamente 50% dos fármacos.³⁰ A CYP3A4, mostrada na Figura 4, é a forma mais abundante no fígado. Embora com papel menos importante no metabolismo de fármacos, as diferentes isoformas do CYP também estão presentes em outros órgãos do corpo, como pulmão, rins e na parede intestinal.^{17,27}

Figura 4. Estrutura da citocromo CYP3A4 com o grupo heme representado no centro (fonte, PDB: 1W0E).

Uma grande variedade de ligantes é capaz de se ligar a isoforma CYP3A4. Alguns exemplos de fármacos importantes que modulam a atividade da CYP3A4 em diferentes categorias são: i. substratos: atorvastatina, ciclosporina, sildenafil, fluoxetina e diazepam; ii. indutores: efavirenz, nevirapina; iii. inibidores: saquinavir, cimetidina, ciprofloxacina, ketoconazol; entre vários outros.²⁷

5.3 Biodisponibilidade

A indústria farmacêutica tem como foco principal o desenvolvimento de moléculas bioativas que possam ser administradas por via oral. Uma das propriedades farmacocinéticas mais importantes no processo de descoberta é a biodisponibilidade oral (do inglês, *Oral Bioavailability*), que expressa o valor percentual da dose de um fármaco administrado por via oral que atinge a circulação sistêmica. Por definição, quando um fármaco é administrado por via intravenosa, a sua biodisponibilidade é de 100%.^{38,40}

A intensidade da biodisponibilidade de fármacos varia substancialmente se considerarmos a diversidade química e a classe terapêutica. A biodisponibilidade é afetada por diversos processos complexos que tem início desde a administração de um fármaco, tais como, desintegração e dissolução da formulação, solubilidade do princípio ativo, absorção intestinal, metabolismo mediado pela CYP, ligação às proteínas plasmáticas e excreção.¹⁷ A Figura 5 ilustra os caminhos para que um fármaco atinja a via sistêmica e apresente biodisponibilidade oral em humanos.

Figura 5. Representação da biodisponibilidade oral em humanos.¹⁷

Ao longo do processo de planejamento de NCEs, é comum que muitas moléculas que possuem comprovada potência e seletividade em ensaios *in vitro* não apresentem o mesmo perfil em modelos *in vivo*. A falta de ação farmacológica muitas vezes está associada a propriedades farmacocinéticas inapropriadas, que por sua vez, não permitem que o princípio ativo atinja o seu alvo para exercer o efeito terapêutico. Esse é um caso que evidencia a importância do estudo da biodisponibilidade em candidatos a novos fármacos.^{27,35}

Nesta dissertação de mestrado, a biodisponibilidade foi selecionada como propriedade farmacocinética alvo para o desenvolvimento de modelos preditivos de ADME.

5.4 Distribuição (D)

Após serem absorvidos, os fármacos alcançam a circulação sistêmica e são distribuídos de maneira não uniforme nos diferentes tecidos.⁴² A extensão da ligação às proteínas plasmáticas é um fator diretamente relacionado a concentração dos fármacos na circulação e, por consequência, afeta a magnitude da ação farmacológica (efeito terapêutico) e também a depuração dos fármacos por filtração e por processos ativos.

O termo distribuição se refere à transferência reversível do fármaco de um local a outro dentro do organismo. Os padrões de distribuição de um fármaco refletem alguns fatores fisiológicos, bem como suas propriedades físico-químicas e farmacocinéticas.³³

5.5 Ligação às Proteínas Plasmáticas

A ligação às proteínas plasmáticas (PPB, do inglês, *Plasma Protein Binding*) é outra propriedade farmacocinética muito importante no planejamento de fármacos. Possui papel central no transporte de fármacos no organismo humano, afetando de forma significativa a distribuição e excreção, e conseqüentemente, os efeitos farmacológicos e toxicológicos.

Ao alcançarem a circulação sangüínea, 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 ácida.⁴³

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 que seja alcançado um equilíbrio, 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 sangüínea, retardando a chegada de fármacos aos sítios de eliminação e disponibilizando-os no local de ação quando necessário.^{44,45} Quando a ligação à proteína 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, reciprocamente, quando a ligação é fraca, pode ocorrer um aumento na intensidade de

ação do fármaco. Em grande parte dos casos, fármacos que apresentam PPB > 90%, possuem baixa toxidez por não ultrapassarem os índices da faixa terapêutica.⁴⁰

Um outro aspecto importante refere-se à competição de fármacos pelos sítios de PPB, sendo deslocado o que possui menor afinidade relativa, ficando, portanto, com a sua fração livre aumentada no plasma. Também pode ocorrer que, ao ligar-se à proteína plasmática, o fármaco modifique a estrutura terciária desta, alterando assim, a afinidade da proteína por outros xenobióticos. Estes mecanismos moleculares resultam no aumento da fração livre do fármaco deslocado. A mudança de 1% da PPB de um fármaco que se liga 99% às proteínas plasmáticas, terá duplicada a percentagem da fração livre, podendo por isso ocorrer um aumento de sua atividade farmacológica. Em geral, as diferenças individuais na PPB podem contribuir significativamente para a variabilidade na resposta clínica dos fármacos.

A PPB foi a segunda propriedade farmacocinética escolhida nessa dissertação de mestrado para o desenvolvimento de modelos preditivos de ADME.

6. Modelagem *In Silico* de Propriedades de ADME

Devido à emergência de tecnologias em larga escala na indústria farmacêutica moderna, dados para uma variedade de propriedades farmacocinéticas começaram a ser acumulados.^{5,21,25,46,47} Tendo em vista a grande importância dessas propriedades no panorama atual do processo de descoberta e desenvolvimento de fármacos, diversos conjuntos de dados, comerciais ou não, foram (e continuam a ser) criados e empregados na modelagem *in silico* de propriedades de ADME.^{48,49} O uso de modelos

in silico tem evoluído com avanços notáveis nas áreas de correlação, predição, simplificação, automatização, entre outros aspetos relevantes.

Claramente, os modelos de ADME *in silico*, em comparação com os ensaios experimentais tradicionais, possuem uma aplicabilidade maior para atender a enorme demanda gerada na triagem em larga escala de novas moléculas.^{21,25,50-53} Além disso, os ensaios *in vitro* e *in vivo* apresentam desvantagens que limitam o seu uso em larga escala: são complexos e dispendiosos em termos de materiais, infra-estrutura e pessoal qualificado.⁵⁴ Por essas razões, há grande interesse da indústria na geração de modelos de ADME *in silico* que possam rapidamente auxiliar na seleção de moléculas promissoras e guiar a eliminação de compostos com perfil farmacocinético inadequado. Por outro lado, a integração de modelos de ADME (*in silico*, *in vitro* e *in vivo*) parece ser um caminho essencial a ser seguido em todos os estágios do processo de descoberta de fármacos.⁵⁵⁻⁵⁷

Existe uma grande variedade de modelos de ADME *in silico* disponíveis, com diferentes níveis de complexidade e capacidade de processamento. Em particular, métodos de QSAR têm sido empregados na geração de modelos para diversas propriedades farmacocinéticas, tais como metabolismo mediado pela CYP, absorção intestinal, biodisponibilidade oral, passagem através da barreira hemato-encefálica, entre outras.⁵⁸⁻⁶⁶ A rica história de quase 50 anos do QSAR continua revelando o potencial notável desta técnica para um amplo conjunto de aplicações. QSAR é uma tecnologia que gera descritores baseados em estruturas moleculares e usa algoritmos computacionais para relacionar os descritores principais com o valor da propriedade de interesse.⁶⁷ As técnicas de QSAR buscam identificar e quantificar as relações

predominantes no complexo campo de modelagem, representadas pelas propriedades da estrutura química e atividade biológica, da estrutura química e propriedade, ou ainda, num paradigma mais moderno, pelas propriedades da estrutura química e propriedades de ADME (QSAR-ADME).^{57,68-70} No campo de ADME, o uso do QSAR é especialmente atrativo devido ao grande acúmulo de dados na literatura de algumas propriedades farmacocinéticas. O maior desafio do QSAR no campo de modelagem de propriedades de ADME é a criação de modelos capazes de prever a propriedade farmacocinética alvo de NCEs ainda não sintetizadas.⁷¹

Na presente dissertação de mestrado foi empregado o método holograma QSAR (HQSAR), de forma inédita, na criação de modelos preditivos de QSAR-ADME para biodisponibilidade e PPB.⁷²⁻⁷³ O método VolSurf foi também empregado na modelagem destas propriedades farmacocinéticas.⁷⁴⁻⁷⁵

7. Holograma QSAR

O método holograma QSAR (HQSAR) emprega hologramas moleculares originários da fragmentação molecular 2D. Os hologramas definem a quantidade de tipos diferentes de fragmentos que dão origem aos descritores. Os descritores codificam todos os fragmentos moleculares 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 (PLS, do inglês, *Partial Least Squares*) 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.^{72,73,76,77}

Durante o processo de modelagem de HQSAR, vários parâmetros que afetam diretamente a geração, e por consequência, a qualidade dos modelos, podem ser controlados. Uma representação do processo envolvido na geração de hologramas moleculares no método HQSAR é apresentada na Figura 6.

Figura 6. 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) imposto aos

fragmentos a serem gerados. Como na maioria dos casos o tamanho do holograma é menor que o número total de fragmentos gerados, ocorre à colisão de fragmentos, ou seja, fragmentos distintos são agrupados no mesmo holograma.^{72,73,76,77} 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 apresenta os diferentes tipos de distinção de fragmentos que podem ser considerados em HQSAR.

Os fragmentos gerados são relacionados com a propriedade alvo através do método PLS. Dessa forma, o método HQSAR identifica nas moléculas bioativas padrões de fragmentos (subestruturas) que são relevantes para a propriedade alvo. Além disso, a incorporação progressiva de informações sobre cada fragmento e cada um de seus constituintes, implicitamente codifica informações sobre a estrutura tridimensional.⁷²⁻⁷³

Tabela 1. 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
Ligação (B)	Permite que os fragmentos sejam distinguidos com base nos tipos de ligações: simples, duplas, triplas e aromáticas
Conectividade (C)	Permite que os fragmentos sejam diferenciados baseando-se no estado de hibridização dos átomos contidos no fragmento
Hidrogênio (H)	A distinção dos fragmentos é feita com base na presença e número de átomos de hidrogênio
Quiralidade (Ch)	Permite que os fragmentos sejam distinguidos com base na presença de centros estereogênicos
Doador e Aceptor (DA)	Esta opção permite que os fragmentos sejam distinguidos com base na presença de átomos aceptores ou doadores de ligações hidrogênio

Um esquema geral das etapas envolvidas na construção de um modelo HQSAR é ilustrado na Figura 7.

Figura 7. Etapas envolvidas no método HQSAR.

Um papel importante de um modelo de HQSAR, além de prever a propriedade alvo de moléculas não testadas, é fornecer indicações sobre quais fragmentos moleculares são importantes para a propriedade em questão. Estas informações podem orientar a síntese de novas moléculas com propriedades otimizadas. No método HQSAR, este objetivo é atingido através da construção de um mapa de contribuição 2D, que utiliza um código de cores para indicar a contribuição dos diferentes fragmentos moleculares para a propriedade alvo das moléculas. O método HQSAR, que atualmente faz parte de um leque de tecnologias disponíveis em Química Medicinal, é capaz de gerar modelos com alta capacidade de interpolação e extrapolação em diversos conjuntos de dados de propriedades alvo.

8. VolSurf

O método VolSurf é empregado na modelagem de várias propriedades associadas aos eventos farmacocinéticos de ADME.⁷⁴ Este método gera descritores 2D baseados nos campos de interação 3D providos do método GRID. Os campos moleculares de interação (MIF, do inglês, *Molecular Interaction Fields*) exploram o espaço das propriedades físico-químicas das moléculas a partir das energias de interação entre estas e uma sonda química virtual (Figura 8).^{78,79} Os modelos VolSurf são desenvolvidos empregando o método PLS, a análise de componentes principais (PCA, do inglês, *Principal Components Analysis*) e técnicas de validação cruzada.⁸⁰

Figura 8. Cálculo dos MIF gerados pela sonda água (OH2) no método VolSurf. (a) determinação dos pontos nos quais a sonda explora as interações moleculares; (b) representação de um campo favorável na região em azul.⁸¹

O mapa tridimensional pode ser visto como uma matriz a qual contém a medida das energias atrativas e repulsivas entre a molécula e a sonda. Cada vértice dos cubos que seccionam a caixa tridimensional (espaçamento da caixa) representa um ponto a partir do qual as energias são exploradas pela sonda química virtual sobre a molécula em questão (Figura 8). O espaçamento da caixa pode ter a dimensão variada de 0,1 a 1 Å no sentido de melhor explorar as energias de interação.

As interações calculadas através da combinação de sondas virtuais representam as interações entre a molécula e os vários compartimentos biológicos por onde ela pode circular.⁸¹ As sondas descritas na Tabela 2 podem ser usadas em conjunto para a composição dos MIF que tem por objetivo representar as interações associadas aos processos inerentes aos sistemas de ADME.

Tabela 2. Definição das principais sondas químicas usadas no método Volsurf.

Sondas Químicas	
OH2	Água: simula o processo de solvatação-desolvatação
DRY	Hidrofóbica: calcula a energia de interação hidrofóbica e simula interações entre a molécula e à membrana biológica
N1	Amida: Representa um nitrogênio de amida (ex: NH de ligações peptídicas em proteínas) e não doa ou aceita ligação de hidrogênio.
O	Oxigênio Carbonílico: simula interações entre a molécula e a membrana biológica.
O::	Oxigênio Carboxílico: simula duas ligações de hidrogênio na direção de dois pares isolados contidos na sonda.

Uma grande vantagem do método VolSurf é a sua simplicidade. Os mapas e os descritores podem ser facilmente interpretados quimicamente e são especificamente designados para a otimização *in silico* das propriedades farmacocinéticas.⁸² Alguns descritores 2D gerados a partir dos MIF são apresentados na Tabela 3.

Tabela 3. Definição de descritores físico-químicos calculados pelo método Volsurf.

Descritores 2D	Definição
VOH2 (volume molecular)	Definido quando a sonda água (OH2) interage com a molécula. Representa o volume (em Å ³) acessível pela água, calculado a 0,20 kcal/mol
SOH2 (superfície molecular)	Definida quando água (OH2) interage com a molécula. Representa a superfície acessível (em Å ²) a sonda água interagindo a 0,20 kcal/mol
W1-8OH2 (região hidrofílica)	Pode ser definida como o envelope molecular acessível pelas moléculas de água
Iw1-8OH2 (momento hidrofílico “integy”)	Mede o desequilíbrio entre o centro de massa da molécula e a posição das regiões hidrofílicas em volta. Calculado pela sonda água (OH2) a -0,2 -0,5 -1,0 -2,0 -3,0 -4,0 -5,0 -6,0 kcal/mol
Cw1-8OH2 (fator de capacidade)	Representa a razão entre as regiões hidrofílicas e a superfície molecular calculada pela sonda água (OH2) a -0,2 -0,5 -1,0 -2,0 -3,0 -4,0 -5,0 -6,0 kcal/mol
D1-8DRY (região hidrofóbicas)	São definidas quando a sonda DRY interage com a molécula. Regiões hidrofílicas indicam interações com a sonda hidrofílica a oito diferentes níveis de energia (-0,2 -0,4 -0,6 -0,8 -1,0 -1,2 -1,4 -1,6 kcal/mol)
ID1-8DRY (momento hidrofóbico “integy”)	Mede o desequilíbrio entre o centro de massa da molécula e as posições das regiões hidrofóbicas em volta medidas a -0,2 -0,5 -1,0 -2,0 -3,0 -4,0 -5,0 -6,0 kcal/mol
HB1-8 (ligação de hidrogênio)	Representa detalhes sobre as ligações de hidrogênio das moléculas estudadas. Esse descritor é obtido através das diferenças entre as regiões negativas hidrofílicas geradas pela sonda água (W1-8OH2) com qualquer outra sonda polar como: N1, O ou O::
MW (massa molar)	Determina a massa molar
Log P	Determina o coeficiente de partição no sistema octanol/água

Os descritores moleculares 2D mostrados na Tabela 3 são relacionados com a propriedade farmacocinética através das análises PLS ou PCA. Dessa forma, o método VolSurf é capaz de identificar padrões físico-químicos relevantes que estão associados a propriedade farmacocinética estudada.

O método VolSurf foi empregado nesta dissertação de mestrado na modelagem de conjuntos de dados das propriedades farmacocinéticas biodisponibilidade e PPB.

CAPÍTULO II
OBJETIVOS

II. OBJETIVOS

Os objetivos fundamentais desta dissertação de mestrado são:

- ✓ Organizar amplos conjuntos padrões de dados das propriedades farmacocinéticas biodisponibilidade oral e ligação às proteínas plasmáticas.
- ✓ Desenvolver modelos de QSAR-ADME, empregando a técnica holograma QSAR, para os conjuntos de dados das propriedades farmacocinéticas alvo.
- ✓ Desenvolver modelos de ADME para os conjuntos de dados das propriedades farmacocinéticas, empregando o método VolSurf.
- ✓ Avaliar a consistência interna e externa dos modelos através do seu poder de correlação nos conjuntos treinamento e predição nos conjuntos teste.
- ✓ Determinar a potencialidade dos modelos e a sua aplicação em Química Medicinal na identificação, seleção e otimização de moléculas bioativas candidatas a fármacos.
- ✓ Estabelecer as limitações dos modelos, discutindo possíveis mecanismos de otimização.

CAPÍTULO III
PARTE EXPERIMENTAL

III. PARTE EXPERIMENTAL

1. Modelagem de QSAR

Os trabalhos de modelagem de QSAR, análises estatísticas e geração dos modelos HQSAR e VolSurf foram realizados em estações computacionais no Laboratório de Química Medicinal e Computacional (LQMC) do Centro de Biotecnologia Molecular Estrutural (CBME), no Instituto de Física de São Carlos - USP. Estes trabalhos foram realizados na plataforma computacional SYBYL 7.2 (Tripos, St. Louis, EUA), operando no sistema operacional Linux.

2. Conjuntos Padrões de Dados

Foram organizados dois conjuntos padrões de dados de propriedades farmacocinéticas. Os dados estrutura-propriedade farmacocinética foram coletados da literatura.⁸³⁻⁸⁵ O conjunto de dados de biodisponibilidade é constituído de 302 moléculas, enquanto que 304 moléculas constituem o conjunto de dados de ligação às proteínas plasmáticas. Estes conjuntos de dados, contendo os compostos e os valores associados das propriedades farmacocinéticas alvo, são apresentados nas Tabelas 4 e 13, no Capítulo IV desta dissertação. As estruturas 3D empregadas nos métodos HQSAR e VolSurf foram geradas no programa SYBYL 7.2, sendo posteriormente submetidas aos procedimentos de minimização de energia empregando o campo de força Tripos.

3. Holograma QSAR

O módulo de HQSAR da plataforma SYBYL 7.2 foi empregado na produção dos modelos para os conjuntos padrões de dados de biodisponibilidade e ligação às proteínas plasmáticas.

As estruturas das moléculas dos conjuntos de dados e os valores correspondentes da propriedade farmacocinética alvo (biodisponibilidade ou ligação às proteínas plasmáticas) foram transferidos para uma planilha molecular (MSS, do inglês, *molecular spreadsheet*), onde ficou armazenada toda a informação para o processo de modelagem de QSAR. 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*) contendo todos os fragmentos moleculares, os quais são designados por um algoritmo cíclico da verificação de redundância (CRC, do inglês, *Cyclic Redundancy Check*).^{72,73} O holograma contém informações sobre o número e o tipo de cada fragmento de cada molécula. O comprimento L 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 x L, onde K é o número de observações do conjunto treinamento e L é o comprimento do holograma molecular. Para a proposta de QSAR foi criada uma matriz da variável dependente (propriedade farmacocinética alvo).

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 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 IV desta dissertação.

4. Método VolSurf

A primeira etapa do método VolSurf foi à geração dos MIF a partir das estruturas 3D do conjunto treinamento. O módulo VolSurf da plataforma SYBYL 7.2 foi usado para gerar os descritores 2D a partir dos MIF, e relacioná-los às propriedades farmacocinéticas alvo.

Os descritores 2D foram gerados com base nas interações moleculares de combinações de cinco sondas diferentes: OH2, DRY, N1, O e O::.. As sondas OH2 e DRY, que representam interações fundamentais com os compartimentos biológicos, estiveram presentes em todas as análises de combinações entre três, quatro e cinco sondas. A quantidade de descritores 2D gerados pelo VolSurf é dependente do número de sondas usadas na análise, sendo que 78 descritores foram gerados quando as sondas OH2 e DRY estavam presentes, 94 quando uma terceira sonda foi adicionada, 110 descritores para quatro sondas e 126 para cinco sondas. Os descritores gerados foram ordenados na forma de uma matriz $K \times M$, onde K é o número de observações do conjunto treinamento e M o número de descritores 2D gerados a partir dos MIF. A

matriz foi desta maneira analisada estatisticamente. Várias análises PLS foram conduzidas com diversas combinações de sondas e espaçamento da caixa tridimensional com o objetivo de produzir modelos VolSurf com elevada consistência interna. Após a criação do modelo PLS para o conjunto treinamento, procedeu-se uma avaliação por análise de planejamento fatorial (FFD, do inglês, *Fractional Factorial Design*)⁸⁰ para checar a ocorrência de redundância entre os descritores através da combinação entre eles. Por fim, técnicas de validação cruzada determinaram o número de componentes principais 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 IV desta dissertação.

CAPÍTULO IV
RESULTADOS E DISCUSSÃO

IV. RESULTADOS E DISCUSSÃO

1. Conjunto Padrão de Dados de Biodisponibilidade

O conjunto padrão de dados empregado no desenvolvimento dos modelos de HQSAR e VolSurf possui 302 moléculas (a maioria fármacos) associadas aos respectivos valores de biodisponibilidade. A lista com as moléculas e os valores da propriedade farmacocinética (% biodisponibilidade) é apresentada na Tabela 4.

Tabela 4. Conjunto padrão de dados de biodisponibilidade oral em humanos.

Conjunto de Dados					
Composto	%	Composto	%	Composto	%
Abacavir ⁸⁵	86,00	Bufuralol ⁸⁴ (S)	46,00	Clorambucil ⁸⁵	87,00
Acebutolol ⁸⁵ (S)	37,00	Bumetanida ⁸⁵	81,00	Cloranfenicol Palmitate ⁸⁴	80,00
Acetaminofeno ⁸⁵	88,00	Bupropiona ⁸⁴ (R)	70,00	Cloranfenicol ⁸⁴	69,00
Ácido Acetilsalicílico ⁸⁴	68,00	Bupropiona ⁸⁴ (S)	70,00	Cloroquina ⁸⁵ (S)	80,00
Albuterol ⁸⁵ (S)	71,00	Busulfan ⁸³	80,00	Clorpromazina ⁸⁵	32,00
Almotriptano ⁸³	70,00	Cafeína ⁸⁴	100,00	Clortetraciclina ⁸⁴	27,50
Alosetron ⁸³	55,00	Calcitriol ⁸⁵	61,00	Clortalidona ⁸⁵ (R)	64,00
Amiodarona ⁸³	50,00	Candesartan ⁸³ (R)	15,00	Clortalidona ⁸⁵ (S)	64,00
Amrinona ⁸⁵	93,00	Candesartan ⁸³ (S)	15,00	Claritromicina ⁸³	50,00
Anastrozol ⁸⁵	84,00	Carbamazepina ⁸⁴	70,00	Clavulanato ⁸⁵	75,00
Aprepitant ⁸³	62,50	Cefaclor ⁸⁴	90,00	Clofibrato ⁸⁵	95,00
Atenolol ⁸⁵ (S)	54,00	Cefazolim ⁸⁵	90,00	Clonazepam ⁸³	90,00
Atomoxetina ⁸³	63,00	Cefalexina ⁸⁵	90,00	Clonidina ⁸⁵	60,00
Atorvastatina ⁸³	14,00	Ciclofosfamida ⁸⁵	74,00	Clorazepata ⁸⁵ (R)	91,00
Atovaquona ⁸³	23,00	Cicloprolol ⁸⁴ (R)	100,00	Clorazepata ⁸⁵ (S)	91,00
Bepridil ⁸⁴ (R)	60,00	Cicloprolol ⁸⁴ (S)	100,00	Cloroquina ⁸⁵ (R)	80,00
Bepridil ⁸⁴ (S)	60,00	Cimetidina ⁸⁵	60,00	Clorpromazina ⁸⁴	32,00
Bosentan ⁸³	50,00	Cimetropium ⁸⁴	2,00	Cloxacilina ⁸⁴	37,00
Bromocriptina ⁸⁵	4,50	Ciprofloxacina ⁸³	70,00	Clozapina ⁸⁵	55,00
Budesonida ⁸⁴	11,00	Citalopram ⁸⁵ (R)	80,00	Codeína ⁸⁴	55,00
Bufuralol ⁸⁴ (R)	46,00	Citalopram ⁸⁵ (S)	80,00	Dapsona ⁸⁴	93,00

Delavirdina ⁸³	96,00	Flucloxacilina ⁸⁴	49,00	Lormetazepam ⁸⁴ (R)	75,00
Diazepam ⁸⁵	100,00	Flucitosina ⁸³	83,50	Lormetazepam ⁸⁴ (S)	75,00
Dicloxacilina ⁸⁴	67,50	Fluocortolona ⁸⁴	83,50	Losartan ⁸³	33,00
Diflunisal ⁸⁴	100,00	Flupentixol ⁸⁵	47,50	Lovastatina ⁸⁵	5,00
Digitoxina ⁸⁴	88,50	Flufenazina ⁸⁵	2,70	Melfalan ⁸⁵ (R)	71,00
Digoxina ⁸³	95,00	Flurbiprofeno ⁸⁵ (R)	92,00	Melfalan ⁸⁵ (S)	71,00
Dihidroergosina ⁸⁴	10,00	Flurbiprofeno ⁸⁵ (S)	92,00	Mepindolol ⁸⁴ (R)	82,00
Diltiazem ⁸³	40,00	Fluticasona ⁸³	1,00	Mepindolol ⁸⁴ (S)	82,00
Difenhidramina ⁸⁵	72,00	Fluvastatina ⁸³	24,00	Metergolina ⁸⁴	23,00
Disopiramida ⁸⁵ (R)	83,00	Gabapentina ⁸³	60,00	Metformina ⁸³	55,00
Disopiramida ⁸⁵ (S)	83,00	Galantamina ⁸³	90,00	Metaciclina ⁸⁵	58,00
Dixirazina ⁸⁴ (R)	10,00	Gatifloxacina ⁸⁵ (R)	96,00	Metadona ⁸⁵ (R)	92,00
Dixirazina ⁸⁴ (S)	10,00	Gatifloxacina ⁸⁵ (S)	96,00	Metadona ⁸⁵ (S)	92,00
Domperidona ⁸⁴	14,00	Gefitinib ⁸³	60,00	Metimazol ⁸⁴	93,00
Doxapram ⁸⁴ (R)	61,00	Gemifloxacina ⁸³ (S)	71,00	Metotrexato ⁸⁵ (R)	70,00
Doxapram ⁸⁴ (S)	61,00	Gemifloxacina ⁸³ (R)	71,00	Metotrexato ⁸⁵ (S)	70,00
Doxazosina ⁸⁴	65,00	Glimepirida ⁸⁵	100,00	Metilfenobarbital ⁸⁴ (R)	73,00
Doxorubicina ⁸⁴	5,00	Glipizida ⁸⁵	95,00	Metilfenobarbital ⁸⁴ (S)	73,00
Doxiciclina ⁸⁵	93,00	Gliburida ⁸⁵	95,00	Metilprednisolona ⁸⁵	82,00
Dronabinol ⁸³	15,00	Granisetron ⁸⁵	60,00	Metisergida ⁸⁴	13,00
Drospirenona ⁸³	76,00	Haloperidol ⁸⁵	60,00	Metoclopramida ⁸⁵	76,00
Dutasterida ⁸³	60,00	Hidroclorotiazida ⁸⁵	71,00	Metopimazina ⁸⁴	19,00
Eletriptano ⁸³	62,50	Hidromorfona ⁸³	24,00	Metoprolol ⁸⁵ (S)	50,00
Entricitabina ⁸³	93,00	Ibuprofeno ⁸⁴ (R)	80,00	Mexiletina ⁸⁵ (R)	87,00
Enalapril ⁸⁵	41,00	Ibuprofeno ⁸⁴ (S)	80,00	Mexiletina ⁸⁵ (S)	87,00
Endralazina ⁸⁴	75,00	Idarubicina ⁸⁵	28,00	Mianserin ⁸⁴ (S)	20,00
Eproxindina ⁸⁴	70,00	Imipramina ⁸⁵	42,00	Midalcipran ⁸⁴	84,00
Eritromicina ⁸⁵	35,00	Indobufeno ⁸⁵ (S)	85,00	Midazolam ⁸⁴	40,50
Escitalopram ⁸³ (S)	80,00	Indobufeno ⁸⁵ (S)	85,00	Midodrina ⁸³ (R)	93,00
Esomeprazol ⁸³	90,00	Irbesartan ⁸³	70,00	Midodrina ⁸³ (S)	93,00
Estradiol ⁸⁴	3,00	Isoniazid	100,00	Milrinona ⁸⁴	92,00
Etambutol ⁸⁵	77,00	Isradipina ⁸⁵ (R)	19,50	Minociclina ⁸⁵	97,50
Etinil Estradiol ⁸³	40,00	Isradipina ⁸⁵ (S)	19,50	Moexiprila ⁸³	13,00
Etosuximida ⁸⁵ (S)	100,00	Itraconazol ⁸³	55,00	Montelukast ⁸³	73,00
Felodipina ⁸⁵ (R)	15,00	Ketocorolac ⁸³ (R)	100,00	Morfina ⁸³	40,00
Fenciclidina ⁸⁴	72,00	Ketocorolac ⁸³ (S)	100,00	Moxifloxacina ⁸³	90,00
Fanciclovir ⁸³	77,00	Ketoprofeno ⁸⁴ (R)	85,00	Micofenolata ⁸³	72,00
Fenflumizol ⁸⁴	50,00	Ketoprofeno ⁸⁴ (S)	85,00	Nadolol ⁸⁴ (R)	34,00
Fenfluramina ⁸⁴ (R)	89,00	LAAM ⁸⁵	47,00	Nadolol ⁸⁴ (S)	34,00
Fenfluramina ⁸⁴ (S)	89,00	Lamivudina ⁸³	86,00	Nafcillina ⁸⁵	30,00
Feniletilmalonamida ⁸⁴	91,00	Lansoprazol ⁸⁵ (R)	80,00	Nalbuphina ⁸⁵	11,00
Fenitoina ⁸⁴	90,00	Lansoprazol ⁸⁵ (S)	80,00	Naloxona ⁸⁵	2,00
Fenobarbital ⁸⁴	96,00	Letrozol ⁸⁵	99,90	Naltrexona ⁸⁴	20,00
Fenoprofen ⁸⁴ (R)	80,00	Levofloxacina ⁸⁵	99,00	Naratriptano ⁸⁵	67,50
Fenoprofen ⁸⁴ (S)	80,00	Levonorgestrel ⁸⁵	87,00	Nateglinida ⁸³	73,00
Fenoximon ⁸⁴	53,00	Linezolid ⁸³	100,00	Neostigmina ⁸⁵	15,00
Fisostigmina ⁸⁴	6,00	Lisinopril ⁸⁴	25,00	Nevirapina ⁸³	93,00
Flecainida ⁸⁴ (R)	95,00	Lorazepam ⁸⁵ (R)	93,00	Nifedipina ⁸⁵	50,00
Flecainida ⁸⁴ (S)	95,00	Lorazepam ⁸⁵ (S)	93,00	Nitrazepam ⁸⁴	78,00

Nitrendipina ⁸⁴ (R)	16,00	Prednisolona ⁸⁵	82,00	Sulprida ⁸⁵ (S)	27,00
Nitrendipina ⁸⁴ (S)	16,00	Prednisona ⁸⁵	80,00	Suprofeno ⁸⁵ (R)	92,00
Nitrofurantiona ⁸⁵	87,00	Primaquina ⁸⁴ (R)	96,00	Tacrolimus ⁸³	17,00
Nizatidina ⁸³	70,00	Primaquina ⁸⁴ (S)	96,00	Telenzepina ⁸⁴	54,00
Nomifensina ⁸⁴ (R)	27,00	Primidona ⁸⁴	92,00	Telithromicina ⁸³	57,00
Nomifensina ⁸⁴ (S)	27,00	Procainamida ⁸⁵	83,00	Terbinalita ⁸³	40,00
Norethindrona ⁸⁴	64,00	Prociclidina ⁸⁴ (R)	75,00	Testosterone Undecanoato ⁸⁵	7,00
Norfenfluramina ⁸⁴ (R)	85,00	Prociclidina ⁸⁴ (S)	75,00	Tetraciclina ⁸⁵	77,00
Norfenfluramina ⁸⁴ (S)	85,00	Prometazina ⁸⁴ (S)	25,00	Teofilina ⁸³	98,80
Norzimelidina ⁸⁴	66,00	Propiomazina ⁸⁴ (R)	33,00	Tiagabina ⁸³	90,00
Ofloxacina ⁸⁵ (R)	97,50	Propiltiouracil ⁸⁴	78,00	Tiapamil ⁸⁴	22,00
Ofloxacina ⁸⁵ (S)	97,50	Proscilaridina ⁸⁴	7,00	Tiotropium ⁸³	2,50
Olmesartan ⁸³	26,00	Protriptilina ⁸⁴	85,00	Tocainida ⁸⁵ (R)	89,00
Ondansetron ⁸⁵ (R)	62,00	Proxifilina ⁸⁴ (R)	100,00	Tocainida ⁸⁵ (S)	89,00
Ondansetron ⁸⁵ (S)	62,00	Proxifilina ⁸⁴ (S)	100,00	Toliprolol ⁸⁴ (R)	90,00
Oseltamivir ⁸³	75,00	Pirazinamida	96,80	Tramadol ⁸³	75,00
Oxitetraciclina ⁸⁴	58,00	Quinidina ⁸⁵	75,00	Trandolapril ⁸³	10,00
Pantoprazol ⁸³ (R)	77,00	Raloxifeno ⁸³	2,00	Triantereno ⁸⁵	51,00
Pantoprazol ⁸³ (S)	77,00	Repaglinida ⁸³	56,00	Triazolam ⁸⁵	44,00
Penicillina G ⁸⁴	22,50	Ribavirina ⁸³	64,00	Trimetoprima ⁸⁴	63,00
Penicillina V ⁸⁵	60,00	Rifampina ⁸³	88,80	Trospium ⁸³	9,60
Pentazocina ⁸⁴	18,00	Risperidona ⁸³	70,00	Valdecoxib ⁸³	83,00
Pimozida ⁸⁵	50,00	Rizatriptan ⁸³	45,00	Valpróico-Ácido ⁸³	90,00
Pinacidil ⁸⁴ (R)	57,00	Rofecoxib ⁸³	93,00	Venlafaxina ⁸⁵ (R)	45,00
Pinacidil ⁸⁴ (S)	57,00	Ropinirol ⁸³	55,00	Verapamil ⁸³ (S)	13,50
Pipotiazina ⁸⁴	26,00	Rosuvastatina ⁸³	20,00	Verdenafil ⁸³ (R)	15,00
Pirazolac ⁸⁴	93,50	Saquinavir ⁸³	4,00	Ziprasidona ⁸³	60,00
Piroxicam ⁸⁴	100,00	Sildenafil ⁸³	40,00	Zolmitriptano ⁸³	40,00
Pramipexol ⁸³	90,00	Sinvastatina ⁸⁵	5,00	Zolpiclona ⁸⁵ (S)	80,00
Pravastatina ⁸³	17,00	Sirolimus ⁸³	14,00	Zolpidem ⁸⁴	72,00
Prazosina ⁸⁵	68,00	Spironolactona ⁸⁴	25,00		

Durante o extensivo processo de coleta de dados, alguns aspectos mereceram atenção especial com o objetivo de padronizar o conjunto de dados. Por exemplo, não foram incluídos no conjunto de dados compostos que apresentavam biodisponibilidade dose dependente, ou seja, quando o valor da propriedade variava em função da dose administrada.

Um outro aspecto importante considerado foi o limite aceito 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 reportado da propriedade

não ultrapassasse 20 unidades percentuais. Para ilustrar essa situação, pode ser citado o fármaco cardiotônico do grupo dos digitálicos Lanoxicab[®] (Digoxina), que possuía valores reportados entre 90 e 100% (no limite de 10 unidades percentuais). Esse composto, portanto, foi incorporado ao conjunto de dados.

No caso de compostos que apresentavam mais de um valor percentual de biodisponibilidade disponível, foi usada a média dos valores. Desta maneira, foi atribuído um valor de 95% para a digoxina (Tabela 4). Vale salientar que todos os valores coletados referem-se a biodisponibilidade absoluta, que é a fração (%) do fármaco administrado por via oral que atinge a circulação sistêmica em comparação com a dose administrada por via intravenosa.

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 de biodisponibilidade para duas formas. Essa estratégia já foi empregada em conjuntos de dados de ADME descritos na literatura.^{79,82} Todas as moléculas foram consideradas em sua forma neutra.

O conjunto padrão de dados possui alta diversidade química, com notável cobertura do espaço químico-biológico. Entre os fármacos do conjunto de dados são encontrados diversos inibidores enzimáticos, agonistas e antagonistas de receptores, abrangendo diversas classes terapêuticas. Como exemplos, podem ser citados agentes analgésicos, antibacterianos, antidepressivos, hipoglicemiantes, antieméticos, anticonvulsivantes, antifúngicos, anti-hipertensivos, antiinflamatórios, antimaláricos, antiparkinsonianos, antipsicóticos, antitumorais, antivirais, hipolipemiantes,

contraceptivos, corticosteróides, hipnóticos, imunossupressores, psicotrópicos, entre outros.

As estruturas de alguns fármacos são mostradas na Figura 9 para ilustrar o espaço químico-biológico do conjunto de dados.

Figura 9. Representação do espaço químico-biológico do conjunto de dados.

2. Holograma QSAR: Modelo de Biodisponibilidade

Biodisponibilidade é uma propriedade muito importante na descoberta e desenvolvimento de fármacos. No campo da modelagem *in silico* de propriedades farmacocinéticas, existe grande interesse na geração de modelos capazes de prever esta propriedade em diversos estágios do planejamento de NCEs.

A proposta deste estudo é a geração de modelos preditivos a partir de um conjunto padrão de dados, empregando um método de QSAR bastante usado em nosso grupo. A Tabela 4 lista os compostos do conjunto de dados associados ao respectivo valor da propriedade farmacocinética. As moléculas do conjunto de dados cobrem praticamente toda a faixa de distribuição de valores possíveis de biodisponibilidade, entre 1,00 e 100%.

Para o trabalho de modelagem de HQSAR, os 302 compostos do conjunto de dados foram divididos em conjuntos treinamento e teste. O conjunto treinamento para o desenvolvimento dos modelos é formado por 250 compostos, enquanto que o conjunto teste para validação externa dos modelos é formado por 52 compostos. A razão do número de compostos do conjunto treinamento para o teste é de 4,8:1, cerca de 80:20 (%).

Para uma análise mais conveniente e simplificada, o conjunto de dados foi dividido em três faixas de distribuição de valores de biodisponibilidade, como mostrado na Tabela 5.

Como pode ser observado, 20% dos 302 compostos do conjunto de dados encontram-se na faixa de distribuição de baixa biodisponibilidade (com valores de até 30%). Uma fração de 29,80% dos compostos cai na faixa de distribuição de biodisponibilidade moderada (valores entre 30 e 70%). Os 50,20% compostos restantes encontram-se na faixa de alta biodisponibilidade (valores maiores que 70%). Essa observação comprova a tendência natural de que os fármacos devem possuir alta biodisponibilidade, sendo, portanto, uma propriedade essencial a ser considerada nos estágios iniciais do planejamento de NCEs.

Tabela 5. Faixas de distribuição de baixa, moderada e alta biodisponibilidade dos compostos do conjunto de dados.

Biodisponibilidade (%)	Frequência	Compostos (%)	Faixas de Distribuição	
1-10	18	5,96	20,00 %	Baixa
11-20	27	8,94		
21-30	15	5,00		
31-40	14	4,62	29,80 %	Moderada
41-50	16	5,30		
51-60	28	9,27		
61-70	32	10,60		
71-80	50	16,56	50,20 %	Alta
81-90	48	15,87		
91-100	54	17,88		
Total	302	100,00		

A distribuição dos valores de biodisponibilidade para o conjunto de dados e para os conjuntos treinamento e teste é mostrada no histograma da Figura 10.

O conjunto de dados possui, além de elevada diversidade estrutural, uma distribuição apropriada do valor da propriedade farmacocinética. É necessário salientar a boa distribuição existente nos conjuntos treinamento e teste em relação ao conjunto padrão de dados. O conjunto teste, em particular, apresenta uma distribuição representativa em comparação ao conjunto treinamento. Essa distribuição verificada ao longo da faixa de distribuição da propriedade farmacocinética representa um aspecto fundamental na avaliação da consistência interna e externa de modelos de QSAR.

Figura 10. Representação gráfica da distribuição dos valores de biodisponibilidade.

As propriedades mais comuns em estudos de QSAR são potência (e.g., IC_{50} , EC_{50}) e afinidade (e.g., K_i , K_d), cujas faixas de distribuição podem variar em alguns conjuntos de dados por fatores superiores a 20.000 vezes.⁸⁶ Nestes casos, para que ocorra uma linearização dos dados, os valores das propriedades (dados em concentração Molar) são convertidos nos valores correspondentes na escala logarítmica (e.g., $pIC_{50} = -\log IC_{50}$). É importante notar que essa enorme faixa de distribuição do valor da propriedade não se aplica a maior parte dos conjuntos de dados de propriedades de ADME, que possui valores distribuídos em uma faixa definida geralmente de 0,00 a 100%. Essa característica confere a estes conjuntos a clara vantagem no que diz respeito ao conhecimento dos limites da distribuição do

valor da propriedade para todos os compostos considerados no processo de modelagem, sejam do conjunto treinamento ou do conjunto teste. Este é um aspecto relevante, pois um dos maiores desafios do QSAR moderno é prever o valor da propriedade alvo (no caso, potência e afinidade) além dos limites de distribuição do conjunto de dados. O poder de extrapolar valores em limites de confiança aceitáveis é o fator mais limitante do processo.

Após a padronização do conjunto de dados, o passo seguinte foi iniciar 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 implementado na plataforma SYBYL 7.2.

A obtenção dos fragmentos moleculares foi realizada utilizando as seguintes distinções de fragmentos: átomos (A), ligações (B), conectividade (C), átomos de hidrogênio (H), quiralidade (Ch) e doadores e aceptores (DA). Foram empregadas 13 combinações destes parâmetros para o tamanho padrão do fragmento (4-7): A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/H, A/B/C/Ch, A/C/Ch/DA, A/DA, A/C/DA, A/H/DA, A/C/Ch/DA, A/C/H/DA, A/Ch/D-A. 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). Os resultados estatísticos obtidos a partir das análises PLS para os 250 compostos do conjunto treinamento, empregando as várias combinações de distinção de fragmentos, são mostrados na Tabela 6.

Tabela 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	0,43	0,60	18,41	5	53
2	A/B/C	0,52	0,89	9,65	8	353
3	A/B/C/H	0,50	0,81	12,94	8	257
4	A/B/C/H/Ch	0,48	0,79	13,46	7	257
5	A/B/H	0,46	0,75	14,55	7	401
6	A/B/C/Ch	0,53	0,90	9,14	8	353
7	A/B/C/Ch/DA	0,47	0,87	10,65	7	353
8	A/B/DA	0,41	0,73	15,20	8	353
9	A/B/C/DA	0,52	0,81	12,80	6	307
10	A/B/H/DA	0,38	0,70	16,16	7	401
11	A/B/C/H/DA	0,44	0,76	14,39	6	353
12	A/B/Ch/DA	0,42	0,69	16,27	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. 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 do modelo **2** ($q^2 = 0,52$ e $r^2 = 0,89$ com $SEE = 9,65$), A/B/C/Ch do modelo **6** ($q^2 = 0,53$ e $r^2 = 0,90$ com $SEE = 9,14$) e A/B/C/DA do modelo **9** ($q^2 = 0,52$, e $r^2 = 0,81$ com $SEE = 12,80$).

O modelo **6** foi selecionado 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 forma, foram realizadas análises individuais empregando os seguintes tamanhos dos fragmentos: 2-5, 3-6, 4-7,

5-8, 6-9, 7-10, 8-11 e 9-12, representados nestes intervalos pelo número mínimo e máximo de átomos. A Tabela 7 apresenta os resultados obtidos neste processo.

Tabela 7. Influência do tamanho dos fragmentos nos parâmetros estatísticos utilizando a melhor distinção observada (A/B/C/Ch).

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	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 , 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 quando comparado com o modelo **6**, obtido utilizando o tamanho padrão do fragmento (4-7). Como pode ser observado na Tabela 7, o melhor resultado estatístico ($q^2 = 0,70$ e $r^2 = 0,93$ com $SEE = 7,6$) dentre todos os modelos testados foi obtido com o tamanho do fragmento 7-10. Estes resultados confirmam a importância da avaliação da influência do tamanho dos fragmentos na geração do holograma molecular. Além disso, os resultados ilustrados nas Tabelas 6 e 7 mostram que não há uma relação entre o poder preditivo dos modelos de HQSAR e o comprimento do holograma, o que pode ser visto sem surpresa, pois o padrão de ocupação no

holograma molecular mudará de forma não trivial em relação às mudanças na composição do holograma.

O modelo final de HQSAR gerado com o conjunto treinamento possui a validação cruzada com “deixe um de fora” (LOO, do inglês, *Leave-One-Out*), como indicativo maior de sua robustez. Para melhor avaliar a consistência interna desse modelo, foi empregado o método de validação cruzada com “deixe muitos de fora” (LMO, do inglês, *Leave-Many-Out*). Neste processo, grupos de 17 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 do modelo. 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 um modelo de HQSAR apresente utilidade no planejamento de NCEs candidatas a fármacos, é necessário que o mesmo passe por um processo de validação externa para uma avaliação da sua real capacidade preditiva. Como a estrutura molecular codificada nos hologramas moleculares está diretamente relacionada com a propriedade alvo das moléculas do conjunto de dados, os modelos robustos de HQSAR devem ser capazes de prever a biodisponibilidade de novos candidatos baseados nessa informação. A validação externa do melhor modelo de HQSAR obtido para o conjunto treinamento de 250 moléculas, utilizando a distinção de fragmento A/B/C/Ch e o tamanho do fragmento 7-10, foi feita com o emprego de um conjunto teste de 52 moléculas. As moléculas contidas no conjunto teste foram completamente excluídas

durante o treinamento do modelo. Os resultados do processo de validação externa são apresentados na Tabela 8.

Tabela 8. Valores experimentais e preditos de biodisponibilidade, associados aos resíduos para os compostos do conjunto teste.

Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c	Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c
Conjunto Teste									
01	Amiodarona ⁸⁵	50,00	55,77	-5,77	27	Isradipina ⁸⁵ (S)	19,50	10,74	8,76
02	Bepridil ⁸⁴ (S)	60,00	49,36	10,64	28	Ketoprofeno ⁸⁴ (R)	85,00	96,79	-11,79
03	Cafeína ⁸⁴	100,00	94,56	5,44	29	Levofloxacina ⁸⁵	99,00	96,89	2,11
04	Clortalidona ⁸⁵ (S)	64,00	50,81	13,19	30	Lorazepam ⁸⁵ (R)	93,00	83,31	9,69
05	Cimetidina ⁸⁵	60,00	77,71	-17,71	31	Lorazepam ⁸⁵ (S)	93,00	78,53	14,47
06	Claritromicina ⁸³	50,00	44,82	5,18	32	Lormetazepam ⁸⁴ (S)	75,00	81,87	-6,87
07	Clavulanata ⁸⁵	75,00	63,22	11,78	33	Metotrexato ⁸⁵ (S)	70,00	59,45	10,55
08	Clofibrato ⁸⁵	95,00	80,76	14,24	34	Mexiletina ⁸⁵ (S)	87,00	76,76	10,24
09	Clonazepam ⁸³	90,00	79,95	10,05	35	Midazolam ⁸⁴	40,50	54,81	-14,31
10	Cloxacilina ⁸⁴	37,00	35,60	1,40	36	Nateglinida ⁸³	73,00	60,01	12,99
11	Codeína ⁸⁴	55,00	45,49	9,51	37	Nimodipina ⁸³	13,00	26,34	-13,34
12	Dixirazina ⁸⁴ (R)	10,00	12,33	-2,33	38	Nitrazepam ⁸⁴	78,00	87,02	-9,02
13	Dronabinol ⁸³	15,00	5,68	9,32	39	Nitrendipina ⁸⁴ (R)	16,00	30,80	-14,80
14	Escitalopram ⁸³ (S)	80,00	68,42	11,58	40	Nomifensina ⁸⁴ (R)	27,00	43,55	-16,55
15	Fenfluramina ⁸⁴ (S)	89,00	81,39	7,61	41	Noretindrona ⁸⁴	64,00	77,49	-13,49
16	Fenoprofeno ⁸⁴ (S)	80,00	88,32	-8,32	42	Pirazolac ⁸⁴	93,50	101,81	-8,31
17	Fenoximona ⁸⁴	53,00	63,32	-10,32	43	Pramipexola ⁸³	90,00	81,72	8,28
18	Flecainida ⁸⁵ (R)	70,00	82,15	-12,15	44	Prazosin ⁸⁵	68,00	74,90	-6,90
19	Fluphenazina ⁸⁵	2,70	19,42	-16,72	45	Procainamida ⁸⁵	83,00	70,91	12,09
20	Flurbiprofeno ⁸⁵ (S)	92,00	78,71	13,29	46	Risperidona ⁸³	70,00	80,11	-10,11
21	Fluvastatina ⁸³	24,00	30,79	-6,79	47	Ropinirola ⁸³	55,00	68,84	-13,84
22	Gabapentina ⁸³	60,00	73,38	-13,38	48	Sinvastatina ⁸⁵	5,00	4,90	0,10
23	Gatifloxacina ⁸⁵ (S)	96,00	104,69	-8,69	49	Spirolactona ⁸⁴	25,00	18,91	6,09
24	Gemifloxacina ⁸³	71,00	66,47	4,53	50	Tetraciclina ⁸⁵	77,00	94,25	-17,25
25	Glimepirida ⁸⁵	100,00	107,33	-7,33	51	Toliprolol ⁸⁴ (R)	90,00	77,18	12,82
26	Ibuprofeno ⁸⁵ (R)	80,00	70,43	9,57	52	Venlafaxina ⁸⁵ (R)	45,00	47,45	-2,45

^a Literatura. ^b Predito. ^c Resíduo: diferença entre o valor da literatura e o predito.

A diferença entre os valores experimentais e preditos (resíduos da Tabela 8) indica a boa capacidade preditiva do modelo gerado. Os valores preditos se aproximam

dos experimentais, desviando-se em não mais que 15 unidades percentuais. As únicas exceções são os compostos **5**, **19**, **40** e **50**, para os quais os valores apresentam uma diferença mais significativa (entre 16,55 e 17,71 unidades percentuais). A Figura 11 apresenta o gráfico dos valores experimentais contra os valores preditos de biodisponibilidade para os conjuntos treinamento e teste.

Figura 11. Gráfico de HQSAR dos valores experimentais e preditos de biodisponibilidade para os 302 compostos do conjunto de dados.

A análise gráfica confirma a boa concordância dos valores preditos e experimentais. O modelo final gerado apresenta, portanto, considerável consistência interna e externa. O modelo possui capacidade de classificar novos compostos nas faixas de baixa, moderada e alta biodisponibilidade.

3. Método VolSurf: Modelo de Biodisponibilidade

O conjunto de dados empregado foi exatamente o mesmo apresentado na Tabela 4 deste Capítulo. Neste estudo, foi usado o módulo VolSurf implementado na plataforma SYBYL 7.2.

A primeira etapa foi à geração dos MIF a partir das estruturas 3D das moléculas do conjunto treinamento, com o programa GRID. Esse procedimento requer a seleção de sondas químicas virtuais e a definição do espaçamento da caixa tridimensional estruturada em torno das moléculas do conjunto de dados. A obtenção dos MIF foi realizada usando as sondas que representam interações: hidrofílica (OH2), hidrofóbica (DRY), com nitrogênio de amida (N1), interações entre a molécula e a membrana biológica (O), ligações de hidrogênio (O::). As seguintes combinações de sondas foram consideradas para a caixa tridimensional com secção padrão de 0,5 Å: OH2/DRY/N1, OH2/DRY/O, OH2/DRY/O::, OH2/DRY/N1/O, OH2/DRY/N1/O::, OH2/DRY/O/O::, OH2/DRY/N1/O/O::.

As informações providas dos MIF foram em seguida convertidas nos descritores 2D relacionados às propriedades farmacocinéticas.^{74,81} A análise VolSurf foi realizada considerando todos os descritores 2D gerados (Tabela 9). Os resultados estatísticos obtidos a partir das análises PLS para os 250 compostos do conjunto treinamento, empregando diferentes combinações de sondas e espaçamento da caixa (0,5 Å), são mostrados na Tabela 9.

Tabela 9. Análise VolSurf para as várias combinações de sondas empregando o espaçamento padrão da caixa (0,5 Å).

Modelo	Sondas	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	2D
1	OH2/DRY/N1	0,62	0,66	15,91	3	94
2	OH2/DRY/O	0,66	0,69	14,91	3	94
3	OH2/DRY/O::	0,68	0,71	14,72	3	94
4	OH2/DRY/N1/O	0,66	0,74	13,95	4	110
5	OH2/DRY/N1/O::	0,67	0,70	14,86	3	110
6	OH2/DRY/O/O::	0,68	0,73	13,95	4	110
7	OH2/DRY/N1/O/O::	0,68	0,73	14,00	4	126

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; 2D, número de descritores.

Os melhores resultados estatísticos foram obtidos usando as combinações OH2/DRY/O:: do modelo **3** ($q^2 = 0,68$ e $r^2 = 0,71$ com $SEE = 14,72$), OH2/DRY/O/O:: do modelo **6** ($q^2 = 0,68$ e $r^2 = 0,73$ com $SEE = 13,95$) e OH2/DRY/N1/O/O:: do modelo **7** ($q^2 = 0,68$ e $r^2 = 0,73$ com $SEE = 14,00$). Como pode ser observado, os modelos **3**, **6** e **7** não apresentaram diferenças significativas nos parâmetros estatísticos. O modelo **3** foi selecionado para a investigação da influência de diferentes espaçamentos da caixa tridimensional. O parâmetro referente ao espaçamento da caixa controla a distância entre os vértices de cada cubo que seccionam a caixa e define os pontos de interação entre a sonda e a molécula alvo. Os resultados estatísticos, para os diferentes espaçamentos utilizados 0,2; 0,5; 0,7 e 1 Å, são sumarizados na Tabela 10.

A variação do espaçamento da caixa não levou a qualquer melhora nos parâmetros estatísticos quando comparado ao modelo com espaçamento padrão de 0,5 Å.

Tabela 10. Influência do espaçamento da caixa nos parâmetros estatísticos do modelo VolSurf.

Espaçamento da Caixa	Parâmetros Estatísticos			
	q^2	r^2	SEE	N
0,2	0,44	0,55	18,17	2
0,5	0,68	0,71	14,72	3
0,7	0,66	0,70	14,83	3
1	0,68	0,68	15,24	3

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.

Para evitar o problema da redundância entre os descritores 2D, o modelo **3** foi investigado empregando-se a análise de FFD (taxa de combinação = 5; número de componentes (N) = 3). Os resultados da comparação entre o modelo **3** e o modelo **3/FFD** são mostrado na Tabela 11.

Tabela 11. Comparação dos modelos **3** e **3/FFD** nos parâmetros estatísticos do modelo VolSurf (PLS-LOO).

Modelo	Número de Descritores	Parâmetros Estatísticos			
		q^2	r^2	SEE	N
3	94	0,68	0,71	14,72	3
3/FFD	25	0,72	0,75	13,76	3

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.

A análise FFD (modelo **3/FFD**) proporcionou a redução do número de descritores de 94 para 25, melhorando os parâmetros estatísticos. O método de validação cruzada LMO foi aplicado ao modelo com 25 descritores. A retirada de 15 moléculas não levou a uma variação considerável nos resultados estatísticos,

confirmando a consistência interna do modelo. Os modelos **6** e **7** também foram sujeitos a esta mesma sequência de análises, contudo, o modelo **3/FFD** foi o que apresentou os melhores resultados, justificando a sua escolha.

O modelo **3/FFD** foi empregado no processo de validação externa de um conjunto teste contendo 52 moléculas. Os resultados são apresentados na Tabela 12.

Tabela 12. Valores experimentais e preditos de biodisponibilidade, associados aos resíduos para os compostos do conjunto teste.

Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c		Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c
Conjunto Teste										
01	Acebutolol ⁸⁵ (S)	37,00	35,70	1,30		27	Ketocorolac ⁸³ (R)	100,00	90,90	9,10
02	Ácido Acetilsalicílico ⁸⁴	68,00	74,37	-6,37		28	Lamivudina ⁸³	86,00	78,78	7,22
03	Amiodarona ⁸³	50,00	30,55	19,45		29	Lisinopril ⁸⁴	25,00	16,42	8,58
04	Amrinona ⁸⁵	93,00	106,46	-13,46		30	Metaciclina ⁸⁵	58,00	42,11	15,89
05	Atenolol ⁸⁵ (S)	54,00	42,61	11,39		31	Metotrexato ⁸⁴	70,00	56,03	13,97
06	Budesonida ⁸⁵	11,00	28,90	-17,90		32	Metilfenobarbital ⁸⁵	73,00	83,68	-10,68
07	Bufuralol ⁸⁵ (S)	46,00	55,15	-9,15		33	Metopimazina ⁸⁵	19,00	35,50	-16,50
08	Busulfan ⁸⁵	80,00	67,19	12,81		34	Metoprolol ⁸⁵ (S)	50,00	55,03	-5,03
09	Candesartan ⁸³	15,00	41,91	-26,91		35	Nafcilina ⁸⁵	30,00	51,72	-21,72
10	Carbamazepina ⁸⁴	70,00	80,08	-10,08		36	Neostigmina ⁸⁵	15,00	4,22	10,78
11	Cimetropium	2,00	-7,02	9,02		37	Penicilina ⁸⁵ V	60,00	55,96	4,04
12	Clavulanato ⁸⁴	75,00	83,99	-8,99		38	Prometazina ⁸⁵ (S)	25,00	46,26	-21,26
13	Clorazepato ⁸⁴	91,00	89,12	1,88		39	Rofecoxib ⁸³	93,00	74,12	18,88
14	Clorpromazina ⁸⁴	32,00	40,30	-8,30		40	Ropinirola ⁸³	55,00	68,23	-13,23
15	Dicloxacilina ⁸⁴	67,50	55,87	11,63		41	Sildenafil ⁸³	40,00	55,93	-15,93
16	Doxapram ⁸⁵ (S)	61,00	42,22	18,78		42	Sulprida ⁸⁵ (S)	27,00	55,22	-28,22
17	Doxorubicina ⁸⁴	5,00	24,59	-19,59		43	Suprofen ⁸⁵ (R)	92,00	73,56	18,44
18	Etambutol ⁸⁴	77,00	60,60	16,40		44	Terbinalita ⁸³	40,00	42,35	-2,35
19	Etosuximida ⁸⁵ (S)	100,00	115,2	-15,20		45	Testosterona ⁸⁵	7,00	16,37	-9,37
20	Fenitoína ⁸⁴	90,00	112,09	-22,09		46	Teopilina ⁸³	98,80	110,29	-11,49
21	Flupentixol ⁸⁵	47,50	34,88	12,62		47	Trimetoprima ⁸⁴	63,00	75,10	-12,10
22	Galantamina ⁸³	90,00	71,35	18,65		48	Valdecocib ⁸³	83,00	74,47	8,53
23	Gefitinib ⁸³	60,00	46,29	13,71		49	Valpróico-Ácido ⁸³	90,00	101,99	-11,99
24	Ibuprofeno ⁸⁴ (S)	80,00	80,93	-0,93		50	Ziprasidona ⁸³	60,00	73,29	-13,29
25	Imipramina ⁸⁴	42,00	35,45	6,55		51	Zolpiclona ⁸⁵ (S)	80,00	55,98	24,02
26	Indobufeno ⁸⁵ (S)	85,00	67,33	17,67		52	Zolpidem ⁸⁴	72,00	58,93	13,07

^a Literatura. ^b Predito. ^c Resíduo: diferença entre o valor da literatura e o predito.

Os valores preditos desviam dos experimentais em não mais do que 20 unidades percentuais. As únicas exceções são os compostos **9, 20, 35, 38, 42 e 51**, para os quais os valores preditos apresentam um erro mais significativo, entre 21,26 e 28,22 unidades percentuais. O gráfico das relações entre os valores experimentais e preditos de biodisponibilidade dos compostos do conjunto de dados é apresentado na Figura 12.

Figura 12. Relação entre os valores experimentais e preditos de biodisponibilidade para os 302 compostos do conjunto de dados.

A concordância entre os valores experimentais e preditos indica a utilidade do modelo preditivo no planejamento de NCEs. Podendo este classificar moléculas em faixas de distribuição pré-definidas de biodisponibilidade (Tabela 12).

No método VolSurf, a análise PLS é capaz de discriminar a importância dos descritores em relação à propriedade alvo, conforme mostra a Figura 13. No caso estudado, a biodisponibilidade é favorecida para altos valores de **Iw1,2,8_{OH2}**, **Cw4_{OH2}**, **HB1,2_{O::}**. Por outro lado, pode ser observado que os descritores **V_{OH2}**, **S_{OH2}**, **ID7,8_{DRY}**, **W1,2,3_{O::}** são desfavoráveis a biodisponibilidade.

Figura 13. Coeficientes PLS dos descritores VolSurf.

De forma ilustrativa, as contribuições dos descritores para dois fármacos do conjunto de dados são apresentadas na Figura 14.

Figura 14. Representação da contribuição dos descritores na biodisponibilidade dos fármacos isonizida e atorvastatina.

Para o fármaco antibacteriano de alta biodisponibilidade isoniazida (Rifater®), observa-se um excelente acordo entre estes descritores e aqueles apresentados na Figura 13. Em contrapartida, os descritores para o fármaco hipolipemiante de baixa biodisponibilidade atorvastatina (Lípitor®), apresentam uma relação inversa em comparação aos descritores da Figura 13.

4. Conjunto Padrão de Dados de Ligação às Proteínas Plasmáticas

O conjunto padrão de dados empregado no desenvolvimento dos modelos de HQSAR e VolSurf possui 304 moléculas associadas aos respectivos valores da ligação às proteínas plasmáticas (PPB). A lista com as moléculas e os valores da propriedade farmacocinética (% PPB) é apresentada na Tabela 13.

Tabela 13. Conjunto Padrão de Dados de Ligação às Proteínas Plasmáticas.

Conjunto de Dados					
Composto	PPB (%)	Composto	PPB (%)	Composto	PPB (%)
5-Fluorouracila ⁸⁴	10,00	Benazepril ⁸³	96,70	Cetirizina ⁸³ (S)	93,00
Abacavir ⁸³	50,00	Benazeprilat ⁸³	95,30	Cevimelina ⁸³	20,00
Acetaminofeno ⁸⁴	20,00	Bexarotena ⁸³	99,00	Cidofovir ⁸⁴	6,00
Ácido Acetilsalicílico ⁸⁴	49,00	Bicalutamida ⁸³	96,00	Cilostazol ⁸³	96,50
Ácido Aminosalicílico ⁸³	55,00	Biotin ⁸³	20,00	Cinacalcet ⁸³	95,00
Acitretina ⁸³	99,90	Bosentan ⁸³	98,00	Ciprofloxacina ⁸³	30,00
Acrivastina ⁸³	50,00	Bromocriptina ⁸⁴	93,00	Cisapride ⁸⁴ (R)	98,00
Aciclovir ⁸³	21,00	Bumetanida ⁸⁴	99,00	Cisapride ⁸⁴ (S)	98,00
Albendazol ⁸³	70,00	Buprenorfina ⁸³	96,00	Citalopram ⁸³ (R)	80,00
Albuterol ⁸⁴ (R)	7,00	Bupropiona ⁸⁴ (R)	80,00	Citalopram ⁸³ (S)	80,00
Albuterol ⁸⁴ (S)	7,00	Bupropiona ⁸⁴ (S)	80,00	Clavulanate ⁸³	25,00
Alfentanil ⁸⁴	92,00	Busulfan ⁸³	39,50	Clavulanate ⁸³	25,00
Alosetron ⁸³	82,00	Calcitrol ⁸³	99,00	Clonazepam ⁸³	85,00
Amantadina ⁸³	67,00	Candesartan ⁸³	99,00	Clorazepate ⁸⁴ (R)	97,50
Amiodarona ⁸³	96,00	Capecitabina ⁸³	60,00	Clorazepate ⁸⁴ (S)	97,50
Amitriptilina ⁸⁴	94,80	Carbamazepina ⁸³	76,00	Clorambucil ⁸⁴	99,00
Anfotericina ⁸⁴ B	90,00	Carbidopa ⁸³	36,00	Cloropheniramina ⁸⁴ (R)	70,00
Anlodipina ⁸³ (R)	93,00	Carvedilol ⁸³ (S)	98,00	Cloropheniramina ⁸⁴ (S)	70,00
Anlodipina ⁸³ (S)	93,00	Cefdinir ⁸³	65,00	Cloroquina ⁸³ (R)	55,00
Aprepitant ⁸³	95,00	Cefixima ⁸⁴	67,00	Cloroquina ⁸³ (S)	55,00
Aripiprazol ⁸³	99,00	Cefotetan ⁸⁴	85,00	Clorpromazina ⁸⁴	96,50
Atazanavir ⁸³	86,00	Cefpodoxima Proxetil ⁸³	25,00	Clortalidona ⁸³ (R)	72,00
Atomoxetina ⁸³	98,00	Cefprozil ⁸³	36,00	Clortalidona ⁸³ (S)	72,00
Atorvastatina ⁸³	98,00	Ceftibuten ⁸³	65,00	Clozapine ⁸³	97,00
Atovaquona ⁸³	99,00	Celecoxib ⁸³	97,00	Delavirdina ⁸³	98,00
Azatioprina ⁸³	30,00	Cetirizina ⁸³ (R)	93,00	Desloratadina ⁸³	84,50

Desmetil-Velanfaxina ⁸³ (R)	30,00	Galantamina ⁸³	18,00	Metronidazol ⁸⁴	11,00
Diazepam ⁸⁴	98,70	Ganciclovir ⁸³	1,50	Mexiletina ⁸⁴ (R)	63,00
Diclofenaco ⁸³	99,00	Gatifloxacina ⁸⁴ (R)	20,00	Mexiletina ⁸⁴ (S)	63,00
Dicloxacilina ⁸⁴	95,80	Gemifloxacina ⁸³ (R)	64,00	Miglustat ⁸³	1,00
Difenhidramina ⁸⁴	78,00	Gemifloxacina ⁸³ (S)	64,00	Minociclina ⁸⁴	76,00
Digoxina ⁸³	25,00	Glimepirida ⁸⁴	99,50	Moxifloxacina ⁸³	50,00
Diltiazem ⁸³	78,00	Gliburida ⁸⁴	99,80	Nalbufina ⁸⁴	50,00
Dofetilida ⁸³	65,00	Granisetron ⁸³	65,00	Naloxona ⁸³	45,00
Dolasetron ⁸³	73,00	Haloperidol ⁸⁴	92,00	Naproxen ⁸³	99,00
Donepezil ⁸³ (R)	96,00	Hidroclorotiazida ⁸³	67,90	Nelfinavir ⁸³	98,00
Donepezil ⁸³ (S)	96,00	Hidromorфона ⁸³	13,50	Nevirapina ⁸³	60,00
Doxepina ⁸⁴	82,00	Imipramina ⁸⁴	90,10	Niacina ⁸³	20,00
Doxiciclina ⁸⁴	88,00	Indapamida ⁸³ (R)	75,00	Nicardipina ⁸⁴ (R)	98,70
Doxorubicina ⁸⁴	76,00	Indapamida ⁸³ (S)	75,00	Nicardipina ⁸⁴ (S)	98,70
Dronabinol ⁸³	97,00	Indarubicina ⁸⁴	97,00	Nicotina ⁸⁴	4,90
Drospirenona ⁸³	97,00	Indarubicinol ⁸⁴	94,00	Nifedipina ⁸³	95,00
Drospirenona ⁸³	97,00	Indometacina ⁸³	99,00	Nimodipina ⁸³ (R)	95,00
Duloxetina ⁸³ (R)	90,00	Irbesartan ⁸³	90,00	Nitazoxanida ⁸³	99,00
Duloxetina ⁸³ (S)	90,00	Isoniazida ⁸³	1,00	Nitrofurantiona ⁸⁴	62,00
Efavirenz ⁸³	99,50	Isoniazida ⁸³	1,00	Nizatidina ⁸³	35,00
Enalaprilat ⁸⁴	55,00	Isotretinoína ⁸³	99,00	Nortriptilina ⁸⁴	92,00
Eprosartam ⁸³	98,00	Itraconazol ⁸³	99,80	Olmesartan ⁸³	99,00
Escitalopram ⁸³ (S)	56,00	Ketorolac ⁸³ (S)	99,00	Omeprazol ⁸³ (R)	95,00
Esmolol ⁸⁴ (R)	55,00	LAAM ⁸⁴	80,00	Omeprazol ⁸³ (S)	95,00
Esmolol ⁸⁴ (S)	55,00	Lamivudina ⁸³	36,00	Oseltamivir ⁸³	42,00
Esomeprazol ⁸³ (S)	97,00	Lamotrigina ⁸³	55,00	Oxicarbazepina ⁸³ (R)	40,00
Estazolan ⁸³	93,00	Lansoprasol ⁸³ (R)	97,00	Oxicarbazepin ⁸³ (S)	40,00
Etinyl Estradiol ⁸³	97,00	Lansoprazol ⁸³ (S)	97,00	Pancuronium ⁸⁴	7,00
Etosuximida ⁸⁴ (R)	1,00	Leflunomida ⁸³	99,30	Pantoprazol ⁸³ (R)	98,00
Etosuximida ⁸⁴ (S)	1,00	Levetiracetama ⁸³	10,00	Pantoprazol ⁸³ (S)	98,00
Exemestana ⁸³	90,00	Levodopa ⁸³	20,00	Pemolina ⁸³ (R)	50,00
Ezetimiba ⁸³	90,00	Levonorgestrel ⁸³	97,00	Pemolina ⁸³ (S)	50,00
Ezetimiba ⁸³	90,00	Lopinavir ⁸³	98,50	Pergolida ⁸³	90,00
Felodipina ⁸⁴ (R)	99,60	Loratadine ⁸⁴	75,00	Perindopril ⁸³	60,00
Felodipina ⁸⁴ (S)	99,60	Loratadina ⁸⁴	97,00	Perindoprilat ⁸³	15,00
Fenilefina ⁸⁴ (S)	51,00	Lorazepam ⁸⁴ (R)	91,00	Phenilephina ⁸⁴ (S)	51,00
Fenitoína ⁸⁴	89,00	Lorazepam ⁸⁴ (S)	91,00	Pioglitazona ⁸³ (R)	99,00
Fenofibrato ⁸³	99,00	Lovastatina ⁸³	95,00	Pioglitazona ⁸³ (S)	99,00
Fexofenadina ⁸³	65,00	M Tetrametilmelamina ⁸³	50,00	Pramipexola ⁸³	15,00
Finasterida ⁸³	90,00	Medroxiprogesterona ⁸³	90,00	Pravastatina ⁸³	50,00
Flecainida ⁸⁴ (R)	61,00	Meloxicam ⁸³	99,40	Prazosina ⁸⁴	95,00
Flecainida ⁸⁴ (S)	61,00	Melfalan ⁸⁴ (S)	90,00	Prednisolona ⁸³	80,00
Fluoxetina ⁸⁴ (R)	94,00	Metadona ⁸⁴ (R)	89,00	Prednisona ⁸⁴	75,00
Fluoxetina ⁸⁴ (S)	94,00	Metadona ⁸⁴ (S)	89,00	Procainamida ⁸⁴	16,00
Flurbiprofeno ⁸⁴ (R)	99,50	Metotrexato ⁸⁴ (S)	46,00	Progesterona ⁸³	97,50
Flurbiprofeno ⁸⁴ (S)	99,50	Metilprednisolona ⁸⁴	78,00	Proguanil ⁸³	75,00
Fluvastatina ⁸³	98,00	Metoclopramida ⁸⁴	40,00	Propafenona ⁸³ (R)	95,00
Fluvoxamina ⁸⁴	77,00	Metoprolol ⁸³ (R)	12,00	Propafenona ⁸³ (S)	95,00
Gabapentina ⁸³	3,00	Metoprolol ⁸³ (S)	12,00	Propranolol ⁸⁴ (R)	90,00

Pirazinamida ⁸³	10,00	Sibutramine ⁸⁴ (R)	97,00	Tolcapona ⁸³	99,90
Pirimetamina ⁸³	87,00	Sibutramine ⁸⁴ (S)	97,00	Tolterodina ⁸³	96,30
Quetiapina ⁸³	83,00	Sildenafil ⁸³	96,00	Topiramata ⁸³	15,00
Quinidina 3-Hidroxi ⁸⁴	60,00	Sinvastatin ⁸³	95,00	Toremifena ⁸³	99,50
Quinidina ⁸⁴	87,00	Sirolimus ⁸³	92,00	Torsemda ⁸³	99,00
Rabeprazol ⁸³ (R)	96,30	Stavudine ⁸⁴	1,00	Trandolapril ⁸³	80,00
Rabeprazol ⁸³ (S)	96,30	Sufentanil ⁸⁴	93,00	Trazodona ⁸⁴	93,00
Raloxifeno ⁸³	95,00	Sulfamethoxazole ⁸⁴	53,00	Tretinoin ⁸³	95,00
Repaglinida ⁸³	98,00	Sulfasalazine ⁸⁴	99,30	Triamterena ⁸⁴	61,00
Ribavirina ⁸⁴	1,00	Sulfisoxazole ⁸³	85,00	Triazolam ⁸⁴	90,10
Rifampina ⁸³	80,00	Sulindac ⁸⁴ (R)	94,00	Trimetoprima ⁸⁴	37,00
Rifampina ⁸³	80,00	Sulindac ⁸⁴ (S)	94,00	Ursodiol ⁸³	70,00
Rifapentina ⁸³	93,20	Sumatriptan ⁸³	17,50	Valaciclovir ⁸³	15,70
Riluzol ⁸³	96,00	Tacrolimus ⁸³	99,00	Valdecoxib ⁸³	98,00
Risedronato ⁸³	24,00	Tadalafil ⁸³	94,00	Valpróio-Ácido ⁸³	14,25
Risedronato ⁸³	24,00	Talidomida ⁸³ (R)	55,00	Valsartan ⁸³	95,00
Risperidona Hidroxi ⁸⁴ (R)	77,00	Talidomida ⁸³ (S)	66,00	Verapamil ⁸³ (R)	94,00
Risperidone Hidroxi ⁸⁴ (S)	77,00	Tamoxifen ⁸⁴	98,00	Verapamila ⁸³ (S)	88,00
Risperidona ⁸³	90,00	Telithromycin ⁸³	65,00	Verdenafila ⁸³	95,00
Ritonavir ⁸³	98,50	Terazosin ⁸³	92,00	Warfarina ⁸³ (R)	99,00
Rofecoxib ⁸³	87,00	Terbinalite ⁸³	99,00	Warfarina ⁸³ (S)	99,00
Ropinirola ⁸³	40,00	Tetracycline ⁸⁴	65,00	Zaleplon ⁸⁴	60,00
Rosiglitazona ⁸³	99,80	Teofilina ⁸³	40,00	Zanamivir ⁸⁴	10,00
Saquinavir ⁸³	98,00	Ticlopidina ⁸³	98,00	Zidovudina ⁸³	38,00
Selegilina ⁸⁴ (R)	94,00	Timolol ⁸³	10,00	Ziprasidona ⁸³	90,00
Selegilina ⁸⁴ (S)	94,00	Tinidazol ⁸³	12,00		
Sertralina ⁸³	98,00	Tolcapona ⁸³	99,90		

Os aspectos principais discutidos na organização do conjunto de dados de biodisponibilidade são aplicáveis a este conjunto de dados de PPB e, portanto, não serão discutidos novamente. O conjunto padrão de dados de PPB possui também alta diversidade química, com notável cobertura do espaço químico-biológico.

5. Holograma QSAR: Modelo de Ligação Às Proteínas Plasmáticas

A PPB é uma propriedade farmacocinética atrativa para a geração de modelos de QSAR-ADME úteis nos diversos estágios do processo de planejamento de NCEs. O conjunto padrão de dados possui 304 moléculas, sendo que 250 compõem o conjunto

treinamento e 54 o conjunto teste. A razão de compostos do conjunto treinamento para o conjunto teste é de 4,6:1, cerca de 82:18 (%).

Embora apresente uma maior concentração na região de alta PPB (91-99,9%), o conjunto de dados tem uma distribuição aceitável ao longo da faixa (limites de 0 a 100%) de valores de PPB. Os conjuntos treinamento e teste possuem distribuições razoavelmente uniformes, assegurando a qualidade das avaliações internas e externas do modelo (ver o gráfico da Figura 15 na pág. 63, para uma visualização da distribuição dos conjuntos).

O processo de modelagem de HQSAR foi conduzido de maneira similar ao descrito para biodisponibilidade. A distinção de fragmentos foi avaliada para a geração dos modelos de HQSAR usando o tamanho padrão do fragmento (4-7). Os modelos foram gerados com várias combinações de distinção de fragmentos, variando-se o tamanho dos hologramas entre 53 e 401. Os principais resultados são apresentados na Tabela 14.

Tabela 14. Efeito da distinção de fragmentos no modelo de HQSAR para PPB.

Modelo	Distinção do Fragmento	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	HL
1	A/B	0,50	0,77	14,98	8	307
2	A/B/DA	0,52	0,84	12,35	8	353
3	A/B/H/DA	0,52	0,78	14,59	7	401
4	A/B/C/Ch	0,54	0,86	11,65	8	353
5	A/B/C/Ch/DA	0,51	0,79	14,02	6	401
6	A/B/C/H/DA	0,50	0,80	13,93	8	199
7	A/B/C/H/Ch	0,42	0,82	13,05	8	353
8	A/B/C/H/Ch/DA	0,36	0,82	13,03	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. Distinção do fragmento: A, átomos; B, ligação; C, conexão; H, átomos de hidrogênio; Ch, quiralidade; DA, doador eceptor.

Como pode ser observado, o melhor resultado estatístico foi obtido para o modelo **4** com a combinação de distinção de fragmentos A/B/C/Ch ($q^2 = 0,54$, $r^2 = 0,86$ e SEE 11,65). Contudo, este modelo não apresenta a robustez desejada para ser submetido ao processo de validação externa. Desta maneira, o modelo foi submetido a um estudo sistemático variando o tamanho dos fragmentos de 2-5 até 7-10. A Tabela 15 mostra os resultados desta avaliação.

Tabela 15. Influência do tamanho dos fragmentos nos parâmetros estatísticos do modelo **4** (distinção de fragmentos A/B/C/Ch).

Tamanho dos Fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	N	HL
2-5	0,72	0,91	10,21	8	353
3-6	0,59	0,87	11,17	8	401
4-7	0,54	0,86	11,65	8	353
5-8	0,55	0,86	11,67	8	353
6-9	0,38	0,65	18,11	5	353
7-10	0,32	0,58	20,09	5	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.

Pode ser verificado que o melhor modelo foi gerado com tamanho dos fragmentos 2-5 para a distinção A/B/C/Ch. Este modelo possui parâmetros estatísticos ($q^2 = 0,72$ e $r^2 = 0,91$ com SEE = 10,21) superiores aos gerados no modelo **4**, com o tamanho padrão do fragmento (4-7). A submissão deste modelo ao procedimento PLS-LMO, confirmou a consistência interna do modelo.

O modelo final de HQSAR foi avaliado quanto à sua capacidade preditiva através de um processo de validação externa empregando um conjunto teste com 54 moléculas que não foram incluídas no conjunto treinamento para a geração do modelo.

Os resultados do processo de validação externa são apresentados na Tabela 16.

Tabela 16. Valores experimentais e preditos de PPB, associados aos resíduos para os compostos do conjunto teste.

Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c	Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c
Conjunto Teste									
01	Abacavir ⁸³	50,00	56,04	-6,04	28	Levonorgestrel ⁸³	97,00	97,26	-0,26
02	Ácido Aminosalicílico ⁸³	55,00	59,81	-4,81	29	Metotrexato ⁸⁴ (S)	46,00	60,94	-14,94
03	Acitretina ⁸³	99,90	91,09	8,81	30	Moxifloxacina ⁸³	50,00	40,74	9,26
04	Amiodarona ⁸³	96,00	106,83	-10,83	31	Naproxeno ⁸³	99,00	86,32	12,68
05	Amitriptilina ⁸⁴	94,80	76,13	18,671	32	Nicotina ⁸⁴	4,90	21,08	-16,18
06	Anlodipina ⁸³ (R)	93,00	81,31	11,69	33	Nifedipina ⁸³	95,00	83,62	11,38
07	Azatioprina ⁸³	30,00	20,69	9,31	34	Nizatidina ⁸³	35,00	21,04	13,96
08	Bumetanida ⁸⁴	99,00	94,33	4,67	35	Nortriptilina ⁸⁴	92,00	75,98	16,02
09	Cevimelina ⁸³	20,00	21,02	-1,02	36	Progesterona ⁸³	97,50	105,10	-7,60
10	Clortalidona ⁸⁴ (R)	72,00	89,92	-17,92	37	Quetiapina ⁸³	83,00	63,42	19,58
11	Cisaprida ⁸⁴ (S)	98,00	79,05	18,95	38	Repaglinida ⁸³	98,00	98,40	-0,40
12	Citalopram ⁸⁴ (S)	80,00	68,92	11,08	39	Ribavirina ⁸⁴	1,00	-3,26	4,26
13	Clavulanato ⁸³	25,00	17,42	7,58	40	Risedronato ⁸³	24,00	28,70	-4,70
14	Diclofenaco ⁸³	99,00	105,98	-6,98	41	Risperidona 9-Hidroxi ⁸⁴	77,00	84,35	-7,35
15	Dolasetron ⁸³	73,00	61,88	11,11	42	Ritonavir ⁸³	98,50	80,70	17,79
16	Drospirenona ⁸³	97,00	106,64	-9,64	43	Ropinirola ⁸³	40,00	43,17	-3,17
17	Estazolan ⁸³	93,00	76,64	16,35	44	Sulfisoxazol ⁸³	85,00	75,50	9,50
18	Etinil Estradiol ⁸³	97,00	104,76	-7,76	45	Sulindac ⁸⁴ (S)	94,00	85,85	8,15
19	Fenilefina ⁸⁴ (S)	51,00	30,20	20,79	46	Tadalapril ⁸³	94,00	91,82	2,18
20	Flecainida ⁸⁴ (S)	61,00	53,57	7,43	47	Talidomida ⁸³ (R)	55,00	73,24	-18,24
21	Galantamina ⁸³	18,00	28,58	-10,58	48	Talidomida ⁸³ (S)	66,00	74,71	-8,71
22	Gatifloxacina ⁸⁴ (R)	20,00	30,34	-10,34	49	Teofilina ⁸³	40,00	34,08	5,92
23	Gemifloxacina ⁸³ (S)	64,00	44,97	19,03	50	Tinidazol ⁸³	12,00	12,85	-0,85
24	Glimepirida ⁸⁴	99,50	99,36	0,14	51	Tolterodina ⁸³	96,30	71,35	24,95
25	Indapamida ⁸³ (S)	75,00	91,71	-16,71	52	Topiramato ⁸³	15,00	4,70	10,30
26	Lamivudina ⁸³	36,00	23,26	12,74	53	Zaleplon ⁸⁴	60,00	54,54	5,46
27	Leflunomida ⁸³	99,30	89,06	10,24	54	Zanamivir ⁸⁴	10,00	16,88	-6,88

^a Literatura. ^b Predito. ^c Resíduo: diferença entre o valor da literatura e o predito.

Os valores preditos se aproximam dos experimentais, desviando-se em média não mais do que 18 unidades percentuais. As exceções são os compostos **5**, **11**, **23**, **37**, e **47**, que apresentam erro ligeiramente maior do que 18 unidades. Com um erro mais significativo, se encontram os compostos **19** e **51**, para os quais os valores preditos apresentam resíduos de 20,79 e 24,95 unidades percentuais de PPB, respectivamente. A razoável concordância entre os valores experimentais e preditos de PPB para os compostos do conjunto teste indica a utilidade do modelo preditivo de HQSAR.

O gráfico das relações entre os valores experimentais e preditos de PPB dos compostos dos conjuntos treinamento e teste é apresentado na Figura 15.

Figura 15. Relação entre os valores experimentais e preditos de PPB para os 304 compostos do conjunto de dados.

Uma análise visual permite verificar a consistência interna e externa do modelo. O modelo de HQSAR possui utilidade no processo de avaliação da PPB de substâncias químicas candidatas a novos fármacos.

6. Método VolSurf: Modelo de Ligação às Proteínas Plasmáticas

O conjunto de dados de PPB para este estudo com o método VolSurf é o mesmo empregado no desenvolvimento dos modelos de HQSAR (Tabela 13) do item 5 deste Capítulo. O processo de modelagem foi conduzido de forma similar ao descrito anteriormente para a propriedade biodisponibilidade (item 3 deste Capítulo). Os resultados da avaliação da combinação de sondas na geração dos modelos VolSurf são apresentados na Tabela 17.

Tabela 17. Análise VolSurf para as várias combinações de sondas empregando o espaçamento padrão da caixa (0,5 Å).

Modelo	Sondas	Parâmetros Estatísticos				
		q^2	r^2	SEE	N	2D
1	OH2/DRY/N1	0,65	0,81	12,76	6	94
2	OH2/DRY/O	0,75	0,82	12,43	6	94
3	OH2/DRY/O::	0,73	0,82	12,43	6	94
4	OH2/DRY/N1/O	0,70	0,82	12,41	6	110
5	OH2/DRY/N1/O::	0,75	0,82	12,40	6	110
6	OH2/DRY/O/O::	0,76	0,82	12,26	6	110
7	OH2/DRY/N1/O/O::	0,76	0,82	12,21	6	126

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; 2D, número de descritores.

Os modelos **2**, **5**, **6** e **7** apresentaram diferenças muito discretas em relação aos parâmetros estatísticos. O modelo **2** ($q^2 = 0,75$ e $r^2 = 0,82$ com $SEE = 12,43$) foi selecionado para a investigação da influência de diferentes espaçamentos da caixa 3D e para as demais análises do método VolSurf (de acordo com as etapas descritas para o modelo **3** de biodisponibilidade do item 3 deste Capítulo). Os resultados estatísticos para os diferentes espaçamentos utilizados (0,2; 0,5; 0,7 e 1 Å) são apresentados na Tabela 18.

Tabela 18. Influência do espaçamento da caixa nos parâmetros estatísticos do modelo VolSurf para PPB.

Espaçamento da Caixa	Parâmetros Estatísticos			
	q^2	r^2	SEE	N
0,2	0,44	0,60	18,55	5
0,5	0,75	0,82	12,43	6
0,7	0,73	0,81	12,74	6
1	0,68	0,79	13,41	6

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.

A variação do espaçamento não levou a um modelo mais significativo. Este resultado indica que as interações mais importantes para a obtenção do modelo preditivo foram consideradas no espaçamento padrão de 0,5 Å.

A redução no número de descritores do modelo **2**, de 94 para 39, foi obtida empregando-se a análise de FFD (taxa de combinação = 5; número de componentes (N) = 6). Os resultados da comparação entre o modelo **2** e o modelo **2/FFD** são mostrado na Tabela 19.

Tabela 19. Comparação dos modelos **2** e **2/FFD** nos parâmetros estatísticos do modelo VolSurf (PLS-LOO).

Modelo	Número de Descritores	Parâmetros Estatísticos			
		q^2	r^2	SEE	N
2	94	0,75	0,82	12,43	6
2/FFD	39	0,78	0,82	12,52	6

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.

O método de validação cruzada LMO foi aplicado ao modelo com 39 descritores. A retirada de 15 moléculas não levou a uma variação considerável nos resultados estatísticos ($q^2 = 0,77$), confirmando a robustez interna do modelo gerado para o conjunto treinamento.

Os modelos **5**, **6** e **7** também foram sujeitos a mesma sequência de análises, contudo, o modelo **2/FFD** foi o que apresentou os melhores resultados, justificando a sua escolha.

Em uma etapa posterior, o modelo **2/FFD** foi empregado no processo de validação externa de um conjunto teste contendo 54 moléculas. Os resultados são apresentados na Tabela 20.

Tabela 20. Valores experimentais e preditos de PPB, associados aos resíduos para os compostos do conjunto teste.

Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c		Código	Composto	Lit ^a (%)	Pred ^b (%)	Res ^c
Conjunto Teste										
01	Ácido Valpróico ⁸³	14,25	10,47	3,78		29	Lorazepam ⁸⁴ (R)	91,00	86,21	4,79
02	Alosetrona ⁸³	82,00	104,76	-22,76		30	Nalbufina ⁸⁴	50,00	59,22	-9,22
03	Anlodipina ⁸³ (S)	93,00	76,47	16,53		31	Olmesartan ⁸³	99,00	100,20	-1,20
04	Atorvastatina ⁸³	98,00	102,68	-4,68		32	Pemolina ⁸³ (S)	50,00	37,17	12,83
05	Atovaquona ⁸³	99,00	108,18	-9,18		33	Pimozida ⁸⁴	99,00	110,29	-11,29
06	Benazeprilat ⁸³	95,30	83,08	12,22		34	Pioglitazona ⁸³ (R)	99,00	83,40	15,60
07	Bicalutamida ⁸³	96,00	85,84	10,16		35	Prazosin ⁸⁴	95,00	87,14	7,86
08	Biotina ⁸³	20,00	32,85	-12,85		36	Proguanil ⁸³	75,00	71,92	3,08
09	Calcitrol ⁸³	99,00	85,21	13,79		37	Quinidina 3-Hidroxi ⁸⁴	60,00	56,44	3,56
10	Captopril ⁸³	27,50	32,50	-5,00		38	Quinidina ⁸⁴	87,00	75,84	11,16
11	Carbidopa ⁸³	36,00	24,90	11,10		39	Ropinirola ⁸³	40,00	58,28	-18,28
12	Cidofovir ⁸⁴	6,00	19,37	-13,37		40	Rosiglitazona ⁸³	99,80	94,93	4,87
13	Citalopram ⁸³ (R)	80,00	66,16	13,84		41	Sibutramina ⁸⁴ (S)	97,00	103,39	-6,39
14	Clorazepate ⁸⁴ (S)	97,50	85,28	12,22		42	Sirolimus ⁸³	92,00	109,73	-17,73
15	Clorofeniramine ⁸⁴ (S)	70,00	82,58	-12,58		43	Sufentanila ⁸⁴	93,00	97,08	-4,08
16	Clortalidone ⁸³ (S)	72,00	58,27	13,73		28	Tansulosina ⁸³	96,50	81,58	14,92
17	Desacetil Rifapentina ⁸³	97,70	116,22	-18,52		44	Metadona ⁸⁴ (S)	89,00	82,55	6,45
18	Diltiazem ⁸³	78,00	91,78	-13,78		45	Tetraciclina ⁸⁴	65,00	58,52	6,48
19	Etosuximida ⁸⁴ (S)	1,00	17,15	-16,15		46	Talidomida ⁸³ (R)	55,00	68,50	-13,50
20	Exemestana ⁸³	90,00	79,68	10,32		47	Theofilina ⁸³	40,00	51,88	-11,88
21	Flecainida ⁸⁴ (S)	61,00	65,75	-4,75		48	Timolol ⁸³	10,00	28,47	-18,47
22	Hidroclorotiazida ⁸³	67,90	58,82	9,08		49	Tolterodina ⁸³	96,30	88,95	7,35
23	Imipramina ⁸⁴	90,10	100,93	-10,83		50	Trandolapril ⁸³	80,00	89,06	-9,06
24	Irbesartan ⁸³	90,00	97,82	-7,82		51	Tretinoína ⁸³	95,00	101,77	-6,77
25	Lamotrigina ⁸³	55,00	70,47	-15,47		52	Valdecóxib ⁸³	98,00	84,99	13,01
26	Lansoprasol ⁸³ (R)	97,00	93,85	3,15		53	Vardenafil ⁸³	95,00	78,65	16,35
27	Loratadina ⁸⁴	97,00	108,17	-11,17		54	Verapamil ⁸³ (R)	94,00	85,48	8,52

^a Literatura. ^b Preditos. ^c Resíduo: diferença entre o valor da literatura e o predito.

Em geral, os valores preditos desviam dos experimentais por não mais que 16 unidades percentuais. As exceções são os compostos **2**, **17**, **48** e **39**, para os quais os valores preditos apresentam erro mais considerável (entre 18,28 e 22,76 unidades percentuais). A boa concordância entre os valores de PPB experimental e predito para os compostos do conjunto teste indica a robustez do modelo preditivo obtido com o

programa VolSurf. O gráfico das relações entre os valores experimentais e preditos de PPB dos compostos dos conjuntos treinamento e teste é apresentado na **Figura 16**.

Figura 16. Gráfico dos valores experimentais e preditos de PPB para os 304 compostos do conjunto de dados.

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

A análise PLS discrimina os descritores com relação à PPB conforme mostra a Figura 17. Estão diretamente relacionados V_{OH2} , S_{OH2} , $D1,2,3,4_{DRY}$, $HB1,2_O$, MW e $\log P$. São inversamente relacionados a PPB, $Iw1,8_{OH2}$, $Cw3,5,6_{OH2}$ e $ID5,6_{DRY}$. Esses descritores apresentam contribuição detrimental para uma maior PPB dos compostos.

Figura 17. Coeficientes PLS dos descritores VolSurf.

CAPÍTULO V
CONCLUSÕES

V. CONCLUSÕES

- ✓ Os conjuntos de dados gerados e padronizados neste trabalho, tanto para biodisponibilidade oral quanto para ligação às proteínas plasmáticas, encontram os objetivos e padrões estabelecidos inicialmente.
- ✓ As propriedades farmacocinéticas alvo de nossos estudos são muito complexas e a sua determinação experimental não é trivial. A variabilidade inerente na geração de dados experimentais é um fator que foi analisado na coleta dos dados. Embora com essa reconhecida limitação, os conjuntos padronizados neste trabalho apresentaram as qualificações esperadas para o desenvolvimento de modelos *in silico*.
- ✓ A diversidade química e a distribuição dos valores da propriedade são pontos extremamente atrativos dos conjuntos de dados criados, proporcionando excelente cobertura do espaço químico-biológico a ser explorado.
- ✓ O emprego inédito do método HQSAR na geração de modelos preditivos de ADME foi bem sucedido e sugere a aplicação desta técnica em outros conjuntos de dados de propriedades farmacocinéticas.
- ✓ O emprego do método VolSurf em nossos estudos confirmou a utilidade desta técnica na geração de modelos de propriedades de ADME e serviu como base comparativa para os modelos de HQSAR criados.

- ✓ Em geral, o melhor resultado obtido pelos modelos HQSAR, em comparação com VolSurf, é devido à independência do método HQSAR das conformações 3D dos compostos, que são variáveis através dos diferentes compartimentos biológicos.
- ✓ Todos os modelos finais HQSAR e VolSurf apresentam alta consistência interna e externa. Os modelos foram validados externamente e apresentam substancial poder preditivo.
- ✓ Os modelos *in silico* desenvolvidos são simples, robustos e consistentes para serem usados na triagem *in silico* de séries de moléculas candidatas a novos fármacos. A diversidade química a ser explorada pode ser alta e não há restrições quanto a faixas de distribuição das propriedades alvo.
- ✓ Os modelos gerados se integram perfeitamente às diversas etapas do processo de descoberta e desenvolvimento de novos fármacos. Os modelos podem também serem usados como filtros na criação de conjuntos padrões de dados para aplicação de técnicas em larga escala como o ensaio virtual e o HTS.
- ✓ Os conjuntos e modelos gerados podem ser úteis em estratégias em Química Medicinal na identificação, seleção e otimização de moléculas bioativas candidatas a novos fármacos.

CAPÍTULO VI
REFERÊNCIAS

VI. REFERÊNCIAS

1. Caprino, L.; Russo. P. Developing a Paradigm of Drug Innovation: An Evaluation Algorithm. *Drug Discov. Today*, **2006**, *11*, 999-1006.
2. Yago, G.; Amram, M.; Magula, T. Finacial Innovations for Funding Early-Stage. *Drug Discov. Dev.* **2006**, *12*, 22-28.
3. Pfizer (2006) Princípios Básicos da Farmacoeconomia. <http://www.pfizer.com.br/pdf/farmacoeconomia.pdf>
4. Russo. E. Chemistry Plans a Structural Iverhaul. *Naturejobs*, **2002**, *Nov*, 4-7.
5. Jónsdóttir, S. Ó.; Jorgensen, F. S.; Brunak, S. Prediction Methods and Databases Within Chemoinformatics: Emphasis on Drugs and Drugs Candidates. *Bioinformatics*, **2005**, *21*, 2145-2160.
6. Andricopulo, A. D.; Montanari, C. A. Structure-Activity Relationships for the Design of Small-Molecule Inhibitors. *Mini Rev. Med. Chem.* **2005**, *5*, 585-593.
7. Berndt, E. R.; Gottschalk, A. H. B.; Philipson, T. J.; Strobeck, M. W. Industry Funding of the FDA: Effects of PDUFA on Approval Times and Withdrawal rates. *Nat. Rev. Drug Discov.* **2005**, *4*, 545-554.
8. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Costs. *J. Health Econ.* **2003**, *22*, 151-185.
9. Lombardino, J. G.; Lowe III, J. A. The Role of the Medicinal Chemist in Drug Discovery – Then and Now. *Nat. Rev. Drug Discov.* **2004**, *3*, 853-862.

10. Boston Consulting Group. A Revolution in R&D: How Genomics and Genetics are Transforming the Biopharmaceutical Industry. (Boston Consulting Group, Boston, Massachusetts, 2001)
11. Netterwald, J. Will the Sun Come out Tomorrow? *Drug. Discov. Dev.* **2006**, *12*, 17-20.
12. Couzin, J. The Brains Behind Blockbusters. *Science*, **2005**, *309*, 728-730.
13. Avorn, J. Sending Pharma Better Signals. *Science*, **2005**, *309*, 669.
14. McGee, P. Clinical Trials on the Move. *Drug Discov. Dev.* **2006**, *9*, 16-22.
15. Hoag, H. Seeing the Big Picture. *Nature*, **2006**, *440*, 1084-1085.
16. Mervis, J. Inside the Pipeline: Pharma Goes to Work. *Science*, **2005**, *309*, 721.
17. van de Waterbeemd, H.; Gifford, E. ADMET in silico Modelling: Towards Prediction Paradise? *Nat. Rev. Drug Discov.* **2003**, *2*, 192-204.
18. Fischer, S. Outlook 2007: We Heard an awful lot about the advent of biomarkers. *Drug Discov. Dev.* **2006**, *12*, 30-32.
19. Ashburn, T. T.; Thor, K. B. Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673-683.
20. Weiner, D. Modeling, Simulation, and the Critical Path. *Drug Discov. Dev.* **2006**, *12*, 54.
21. Bajorath, J. Positioning ADMET in silico Tools in Drug Discovery. *Drug Discov. Today*, **2004**, *9*, 14-15.
22. FDA (2004) Innovation stagnation: challenge and opportunity on the critical path to new medicinal products <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf>

23. Butina, D.; Matthew, D. S.; Frankcombe, K. Predicting ADME Properties in silico: Methods and Models. *Drug Discov. Today*, **2002**, 7, S83-S88.
24. Lesney M. S. The Chemistry of ADMET. *Today's Chemist At Work*, **2004**, Jan, 26-30.
25. Beresford, A. P.; Selick, H. E.; Tarbit, M. H. The Emerging Importance of Predictive ADME Simulation in Drug Discovery. *Drug Discov. Today*, **2002**, 7, 109-116.
26. Barreiro, E. J.; Fraga, C. A. M. Química Medicinal As Bases Moleculares da Ação dos Fármacos. Artmed Editora: Brasil, 2001.
27. Williams, A. D.; Lemke, T. L. Foye's Principles of Medicinal Chemistry, 5th ed.; Lippincott Williams & Wilkins: United States of America, 2002.
28. Gardner, C. R.; Walsh, C. T.; Almarsson, O. Drugs as Materials: Valuing Physical Form in Drug Discovery. *Nat. Rev. Drug Discov*, **2004**, 3, 926-934.
29. Tulsi, B.; SAR Revitalized by Huge Compound Repositories. *Drug Discov. Dev.* **2003**, 6, 41-44.
30. Li, A. P. Screening for Human ADME/Tox Drug Properties in Drug Discovery. *Drug Discov. Today*, **2001**, 6, 357-366.
31. Navia, M. A.; Chaturvedi, P. R. Design Principles for Orally Bioavailable Drugs. *Drug Discov. Today*, **1996**, 1, 179-189.
32. Egan, W. J.; Lauri, G. Prediction of Intestinal Permeability. *Adv. Drug Deliv. Rev.* **2002**, 54, 273-289.
33. Smith, D. A.; van de Waterbeemd H.; Walker, D. K. Methods and Principles in Medicinal Chemistry Vol. 13.; Pharmacokinetics and Metabolism in Drug Design. Wiley-VCH: Germany, 2001.

34. 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. *Adv. Drug Disc. Revs.* **1997**, *23*, 3-25.
35. Martin, Y. C. A Bioavailability Score. *J. Med. Chem.* **2005**, *48*, 3164-3170.
36. 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. *J. Med. Chem.* **2002**, *45*, 2615-2623.
37. Vieth, M.; Siegel, M. G.; Higgs, R. E.; Watson, I. A.; Robertson, D. H.; Savin, R. K.; Durst, G. L.; Hipsking P. A. Characteristic Physical Properties and Structural Fragments of Marketed Oral Drugs. *J. Med. Chem.* **2004**, *47*, 224-232.
38. Hardman, J. G.; Limbird, L. E.; Gilman, A. G. Goodman & Gilman's The Pharmacological Basis of Therapeutics, 10th ed.; McGraw-Hill: United States of America, 2001.
39. de Graaf, C.; Vermeulen, N. P. E.; Feenstra, K. A. Cytochrome P450 in Silico: An Integrative Modeling Approach. *J. Med. Chem.* **2005**, *48*, 2725-2755.
40. Wermuth, C. G. The Practice of Medicinal Chemistry, 2th ed.; Academic Press: Great Britain, 2003.
41. de Groot, M. J. Designing Better Drugs: Predicting Cytochrome P450 Metabolism. *Drug Disc. Today*, **2006**, *11*, 601-606.
42. Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391-3400.

43. Aureli, L.; Cruciani, G.; Mariconi, A. Predicting Human Serum Albumin Affinity of Interleukin-8 (CXCL8) Inhibitors by 3D-QSPR Approach. *J. Med.Chem.* **2005**, *48*, 2469-2479.
44. Sheppard, G. S.; Bouska, J. J. Why Optimize Cancer Drugs for ADMET? *Drug Discov. Today*, **2005**, *2*, 343-349.
45. Kratochwil, N. A.; Gerber, P. R. Predicting Plasma Protein Binding of Drugs: a New Approach. *Biochem. Pharma.* **2002**, *64*, 1355-1374.
46. Modi, S. Computacional Approaches to the Understantiding of ADMET Properties and Problems. *Drug Discov. Today*, **2003**, *8*, 621-623.
47. Wolohan, P. R. N.; Clark, R. D. Predicting Drug Pharmacokinetic Properties Using Molecular Interaction Fields and SIMCA. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 65-76.
48. WOMBAT-PK (2007) Database for Clinical Pharmacokinetics. <http://sunsetmolecular.com/>
49. PharmGKB (2007) The Phamacogenetics and Pharmacogenomics Knowledge Base. <http://www.pharmgkb.org/>
50. Ekins, S.; Nikolsky, Y.; Nikolskaya T. Techniques: Application of Systems biology to Absorption, Distribution, Metabolism, Excretion and Toxicity. *Trends Pharmacol. Sci.* **2005**, *26*, 202-209.
51. Hou, T.; Xu, Z. ADME Evaluation in Drug Discovery. *J. Mol. Model*, **2002**, *8*, 337-349.
52. Oprea, T. I. Virtual Screening in Lead Discovery: A Viewpoint. *Molecules*, **2002**, *7*, 51-62.

53. Rose, S.; Stevens, A. Computacional Design Strategies for Combinatorial Libraries. *Curr. Opin. Chem. Biol.* **2003**, *7*, 331-339.
54. Saunders, K. C. Automation and Robotics in ADME Screening. *Drug Discov. Today: Techno.* **2004**, *1*, 373-380.
55. Yu, H.; Adedoyin, A. ADME-Tox in Drug Discovery: Integration of Experimental and Computacional Technologies. *Drug Discov. Today.* **2003**, *8*, 852-861.
56. Ekins, S.; Boulanger, B.; Swaan, P. W.; Hupcey, M. A. Z. Towards a New Age of Virtual ADME/Tox and Multidimensional Drug Discovery. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 381-401.
57. Gasteiger, J. The Central Role of Chemoinformatics, *Chemom. Intell. Lab. Syst.* **2006**, *82*, 200-209.
58. Crivori, P.; Zamora, I.; Speed, B.; Orrenius, C.; Poggesi, I. Model Based on GRID-Derived Descriptors for Estimating CYP3A4 Enzyme Stability of Potential Drug Candidates. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 155-166.
59. Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME Evaluation in Drug Discovery. *J. Chem. Inf. Model*, **2006**, *7*, 1-11.
60. Wang, J.; Krudy, G.; Xie, X. Q.; Wu, C.; Holland, G. Genetic Algorithm-Optimized QSPR Models for Bioavailability, Protein Binding, and Urinary Excretion. *J. Chem. Inf. Model.* **2006**, *46*, 2674-2683.
61. Votano, J. R.; Parham, M.; Hall, M. L.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A. QSAR Modeling of Human Serum Protein Binding with Several Modeling Techniques Utilizing Structure-Information Representation. *J. Med. Chem.* **2006**, *49*, 7169-7181.

62. Lessigiarska, I.; Pajeva, I.; Cronin, M. T. D.; Worth, A. P. 3D QSAR Investigation of The Blood-Brain Barrier Penetration of Chemical Compounds. *SAR QSAR Environ. Res.* **2005**, *16*, 79-91.
63. Doddareddy, M. R.; Cho, Y. S.; Koh, H. Y.; kim, D. H.; Pae, A. N. In Silico Renal Clearance Model Using Classical VolSurf Approach. *J. Chem. Inf. Model*, **2006**, *46*, 1312-1320.
64. Colmenarejo, G.; Pedraglio, A. A.; Lavandera, J. L. Cheminformatic Models to Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001**, *44*, 4370-4378.
65. Ekins, S. In Silico Approaches to Predicting Metabolism, Toxicology and Beyond. *Biochem. Soc. Trans.* **2003**, *31*, 611-614.
66. Yoshida, F.; Topliss, J. G. QSAR Model for Drug Human Oral Bioavailability. *J. Med. Chem.* **2000**, *43*, 2575-2585.
67. Abraham, D. J. Burger's Medicinal Chemistry and Drug Discovery, 6th ed. Vol. 1.; Drug Discovery. John & Sons, Inc: United States of America 2003.
68. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms. *Chem. Rev.* **2001**, *101*, 619-672.
69. Hansch, C.; Leo, A. Exploring QSAR: Fundamental and Applications in Chemistry and Biology. ACS: United States of America, 1995.
70. Hansch, C.; Hoekman, D.; Gao, H. Comparative QSAR: Toward a Deeper Understanding of Chemicobiological Interactions. *Chem. Rev.* **1996**, *96*, 1045-1075.

71. Stouch, T. R.; Kenyon, J. R.; Johnson, S. R.; Chen, X. Q.; Doweyko, A.; Li, Y. In Silico ADME/Tox: Why Models Fail. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 83-92.
72. HQSARTM, User's Manual, Tripos Inc., St. Louis, 2006.
73. Lowis, D. R. HQSAR: A new, Highly Predictive QSAR Technique. Tripos Technical Notes, 1, 1997.
74. Cruciani, G.; Pastor, M.; Guba, W.; VolSurf: A New Tool for the Pharmacokinetic Optimization of Lead Compounds, *Eur. J. Pharm. Sci.* **2000**, *11 Suppl. 2*, S29-S39.
75. VolSurf 7.2 (2006) Molecular Discovery, LD, www.moldiscovery.com
76. Avery, M. A.; Alvim-Gaston, M.; Rodrigues, C. R.; Barreiro, E. J.; Cohen, F. E.; Sabnis, Y. A.; Woolfrey, J. R. Structure-Activity Relationships of the Antimalarial Agent Artemisin. 6. The Developing of Predictive *in vitro* Potency Models using CoMFA and HQSAR Methodologies. *J. Med. Chem.* **2002**, *45*, 292-303.
77. Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. Hologram Quantitative Structure-Activity Relationships for a Series of Farnesoid X Receptor Activators. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119-3125.
78. Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules, *J. Med. Chem.* **1985**, *28*, 849-857.
79. Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting Blood-Brain Barrier Permeation from Three-Dimensional Molecular Structure *J. Med. Chem.* **2000**, *43*, 2204-2216.
80. Laivingstone, D. Data Analysis for Chemists: Applications to QSAR and Chemical Product Design. Oxford University Press: Great Britain, 1995.

81. Zamora, I.; Oprea, T.; Cruciani, G. Surface Descriptors for Protein-Ligand Affinity Prediction. *J. Med. Chem.* **2003**, *46*, 25-33.
82. Cruciani, G.; Crivori, P.; Carrupt, P. A.; Testa, B. Molecular Fields in Quantitative Structure-Permeation Relationships: The VolSurf Approach. *Theochem-J. Mol. Struct.* **2000**, *503*, 17-30.
83. Physicians' Desk Reference (PDR), 59th ed.; Thomson, 2005
84. 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 1924-2023.
85. Sietsema, W. K. The Absolute Oral Bioavailability of Selected Drugs. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1989**, *27*, 179-211.
86. de Paula, C. B. V. Estudos Cinéticos e das Relações Quantitativas Entre a Estrutura e Atividade de Inibidores da Purina Nucleosídeo Fosforilase Bovina e de *Schistosoma mansoni*. Dissertação de Mestrado – Instituto de Física de São Carlos, Universidade de São Paulo, 2005.