**ESTUDOS DAS RELAÇÕES QUANTITATIVAS ENTRE A ESTRUTURA E
ATIVIDADE DE UMA SÉRIE DE INIBIDORES DA PROTEASE DO VÍRUS HIV-1**

Leonardo Luiz Gomes Ferreira

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

Orientador: Prof. Dr. Adriano Defini Andricopulo

São Carlos, Março de 2007

Ferreira, Leonardo Luiz Gomes

“Estudos das Relações Quantitativas entre a Estrutura e Atividade de Uma Série de Inibidores da Protease do Vírus HIV-1”

Leonardo Luiz Gomes Ferreira - São Carlos, 2007

Dissertação (Mestrado) – Área de Física Aplicada – Opção: Física Biomolecular, do Instituto de Física de São Carlos da Universidade de São Paulo

2007 – Páginas: 75

Orientador: Prof. Dr. Adriano Defini Andricopulo

1. Protease; 2. QSAR; 3. HIV-1

I. Título

AGRADECIMENTOS

- Ao professor Adriano Defini Andricopulo, pela excelência na orientação deste trabalho.
- À Juracy Aparecida Rocha e Luciana Corrêa Vellardi, pela fundamental ajuda, na realização deste trabalho.
- Aos colegas do Grupo de Cristalografia pelo apoio, auxílio e pelo ambiente de trabalho que me proporcionaram.
- Aos colegas Andrei Leitão, Káthia Honório, Lívia Salum, Marcelo Gindri e Tiago Moda pelas contribuições na realização deste trabalho.
- Aos professores e funcionários do Instituto de Física de São Carlos pela colaboração e suporte oferecidos.
- Ao Centro de Biotecnologia Molecular Estrutural – CEPID/FAPESP.
- À CAPES pelo suporte financeiro.

SUMÁRIO

Lista de Abreviaturas	vi
Lista de Figuras	vii
Lista de Tabelas	viii
Resumo	ix
Abstract	x

CAPÍTULO I

INTRODUÇÃO	2
1. Fármacos e Saúde Humana	2
2. O Papel da Química Medicinal no Planejamento de Fármacos	2
3. Indústria Farmacêutica, Pesquisa e Desenvolvimento	3
4. O Processo de Desenvolvimento de Novos Fármacos	4
5. A Síndrome da Imunodeficiência Adquirida	7
6. Aspectos Clínicos da AIDS	10
7. Fármacos e Terapia da AIDS	13
8. A Protease do Vírus HIV-1	16
9. Inibidores da HIV-1 PR e o Desenvolvimento de NCEs	19
10. Química Medicinal e as Técnicas de QSAR	20
11. Holograma QSAR	21
12. Análise Comparativa dos Campos Moleculares (CoMFA)	24

CAPÍTULO II

OBJETIVOS	28
------------------	----

CAPÍTULO III

PARTE EXPERIMENTAL	30
1. Modelagem de QSAR	30
2. Conjunto de Dados	30
3. Alinhamento Estrutural	31
4. Desenvolvimento de Modelos de HQSAR	32
5. Geração de Modelos CoMFA	32

CAPÍTULO IV

RESULTADOS E DISCUSSÃO	35
1. Conjunto de Dados	35
2. Holograma QSAR	49
3. QSAR 3D: CoMFA	56

CAPÍTULO V

CONCLUSÕES	69
-------------------	----

CAPÍTULO VI

REFERÊNCIAS BIBLIOGRÁFICAS	71
-----------------------------------	----

LISTA DE ABREVIATURAS

ADME	Administração, Distribuição, Metabolismo e Excreção
AIDS	Síndrome da Imunodeficiência Adquirida
CDC	Centros para o Controle de Doenças
CoMFA	Análise Comparativa dos Campos Moleculares
DNA	Ácido Desoxirribonucleico
EC	<i>Enzyme Commission</i>
FDA	<i>Food and Drug Administration</i>
HAART	Terapia Anti-retroviral Altamente Ativa
HIV-1	Vírus da Imunodeficiência Humana Tipo 1
HIV-1 PR	Protease do Vírus HIV-1
HQSAR	Holograma QSAR
HTS	Ensaio em Larga Escala
IND	Investigação de Novo Fármaco
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
LBDD	Planejamento de Fármacos Baseado no Ligante
NCE	Nova Entidade Química
NIH	<i>National Institute of Health</i>
NNRTI	Inibidor Não Nucleosídico da Transcriptase Reversa
NRTI	Inibidor Nucleosídico da Transcriptase Reversa
NDA	Novo Fármaco
OMS	Organização Mundial da Saúde
P&D	Pesquisa e Desenvolvimento
PDB	<i>Protein Data Bank</i>
PLS	Mínimos Quadrados Parciais
QSAR	Relações Quantitativas Estrutura-Atividade
QSAR 2D	QSAR Bidimensional
QSAR 3D	QSAR Tridimensional
RNA	Ácido Ribonucleico
SAR	Relações Estrutura-Atividade
SBDD	Planejamento de Fármacos Baseado na Estrutura do Receptor
SNC	Sistema Nervoso Central
VS	Ensaio Virtual

LISTA DE FIGURAS

Figura 1	Representação ilustrativa da estrutura da protease do vírus HIV-1	18
Figura 2	Geração de um modelo de HQSAR	23
Figura 3	Geração de um modelo CoMFA	26
Figura 4	Estrutura representativa da diversidade química do conjunto de dados	46
Figura 5	Representação gráfica da distribuição dos valores de pIC_{50}	48
Figura 6	Gráfico HQSAR dos valores experimentais e preditos para as 82 moléculas do conjunto de dados	54
Figura 7	Mapa de contribuição de HQSAR para o inibidor 52	55
Figura 8	Esquema geral das interações entre os inibidores do conjunto de dados e a HIV-1 PR	57
Figura 9	Alinhamento molecular estrutural do conjunto de dados	58
Figura 10	Gráfico CoMFA dos valores experimentais e preditos para as 82 moléculas do conjunto de dados	64
Figura 11	Mapa de contorno 3D CoMFA para o inibidor 52	65
Figura 12	Mapa de contorno estereoquímico do inibidor 52 no sítio ativo da enzima HIV-1 PR	66
Figura 13	Mapa de contorno eletrostático do inibidor 52 no sítio ativo da enzima HIV-1 PR	67

LISTA DE TABELAS

Tabela 1	Principais manifestações clínicas da AIDS	12
Tabela 2	Medicamentos aprovados pelo FDA para o tratamento da AIDS	15
Tabela 3	Parâmetros de distinção de fragmentos no método HQSAR	22
Tabela 4	Conjunto padrão de dados de inibidores da HIV-1 PR	35
Tabela 5	Resultados de HQSAR em função da variação dos parâmetros de distinção de fragmentos	50
Tabela 6	Análise de HQSAR em função da variação do tamanho dos fragmentos	51
Tabela 7	Valores experimentais e preditos de pIC_{50} associados aos resíduos para as moléculas do conjunto teste no modelo HQSAR	53
Tabela 8	Modelos CoMFA obtidos em função da variação em SDC	60
Tabela 9	Resultados CoMFA obtidos em função da variação em GS	61
Tabela 10	Valores experimentais e preditos de pIC_{50} associados aos resíduos para as moléculas do conjunto teste no modelo CoMFA	63

RESUMO

A Protease do Vírus da Imunodeficiência Humana Tipo 1 (HIV-1 PR, EC 3.4.23.16) é um alvo macromolecular de grande importância no desenvolvimento de fármacos na terapia da Síndrome da Imunodeficiência Adquirida (AIDS). As maiores indústrias farmacêuticas do mundo concentram inúmeros esforços em estudos acerca desta proteína, que desde a introdução do saquinavir (Invirase®) na terapêutica em 1995, tem se mantido como um alvo fundamental para a descoberta de novos fármacos anti-HIV. A protease do vírus HIV-1 possui uma história rica de enorme sucesso no processo de descoberta e desenvolvimento de fármacos.

A Química Medicinal moderna, de forte caráter multidisciplinar, fornece um arsenal de alternativas e estratégias racionais úteis no processo de planejamento de novos fármacos. Uma das tecnologias muito empregadas com sucesso é o estudo das relações quantitativas entre a estrutura e atividade (QSAR) para conjuntos de dados padrões. Os estudos de QSAR visam identificar e quantificar no complexo campo de modelagem as relações entre a estrutura e atividade de uma série de moléculas.

Na presente dissertação de mestrado, foram realizados estudos de QSAR bi- (2D) e tridimensionais (3D) empregando, respectivamente, as técnicas holograma QSAR (HQSAR) e a análise comparativa dos campos moleculares (CoMFA), visando à geração de modelos preditivos para um conjunto de inibidores da protease do HIV-1. Os modelos gerados, associados às informações obtidas pelos mapas de contribuição 2D e de contorno 3D, são guias químico-medicinais úteis no planejamento de novos inibidores mais potentes e seletivos da protease do HIV-1.

ABSTRACT

The Human Immunodeficiency Virus Type 1 Protease (HIV-1 PR, EC 3.4.23.16) is a macromolecular target of great importance for the therapy of the Acquired Immunodeficiency Syndrome (AIDS). Major pharmaceutical companies around the world concentrate several efforts on studies concerning this enzyme, which since saquinavir (Invirase[®]) reached the market in 1995, has maintained its status as a fundamental target for anti-HIV drug discovery. HIV-1 protease has a rich history of enormous success in the drug discovery and development process.

The strong multidisciplinary character of modern Medicinal Chemistry supplies an arsenal of useful rational strategies for the design of new drugs. One such technology is quantitative structure-activity relationships (QSAR), which has been successfully applied in a number of settings. QSAR studies aim to identify and quantify the relations between structure and activity of series of bioactive molecules organized within standard data sets.

In the present master's dissertation, 2D and 3D QSAR studies were performed employing the hologram QSAR (HQSAR) and comparative molecular field analysis (CoMFA) techniques, respectively, in order to generate predictive models for a large set of HIV-1 PR inhibitors. The final models along with the information obtained from the 2D contribution and 3D contour maps should be useful in the design of new inhibitors with increased potency and selective within the chemical diversity of the data set.

CAPÍTULO I

INTRODUÇÃO

INTRODUÇÃO

1. Fármacos e Saúde Humana

Os fármacos desempenham um papel essencial na melhoria da qualidade de vida das diversas populações no mundo. Pesquisas demonstram o seu papel eficaz no aumento da expectativa de vida em diversos países entre 1986 e 2000. Além disso, os fármacos auxiliam a prevenir, tratar ou curar doenças outrora fatais ou altamente debilitantes, contribuindo para a diminuição dos casos e riscos envolvidos em internações e cirurgias.^{1,2}

2. O papel da Química Medicinal no Planejamento de Fármacos

A Química Medicinal tem contribuído de forma notável no processo de descoberta e desenvolvimento de novos fármacos. Seu caráter multidisciplinar é marcante nas interfaces da química e biologia, em compasso com o avanço de novas tecnologias, particularmente, nas áreas da quimioinformática e bioinformática.³

A Química Medicinal abrange diversas especialidades em interfaces bem definidas, como a bioquímica, farmacologia, biologia molecular e estrutural, cristalografia, espectroscopia, além das tecnologias baseadas em métodos computacionais. O conhecimento da estrutura tridimensional de alvos moleculares, proporcionado pela biologia molecular e estrutural, tem permitido a aplicação de diversas técnicas de modelagem molecular no estudo das interações fármaco-receptor e no planejamento de novas

entidades químicas (NCEs, do inglês, *new chemical entities*) candidatas a fármacos.

Segundo definido pela IUPAC, a química medicinal se refere à descoberta, ao desenvolvimento, à identificação e à interpretação do modo de ação de compostos biologicamente ativos, em nível molecular. Neste contexto podemos estabelecer que a Química Medicinal é uma disciplina fundamentada em química, envolvendo aspectos das ciências biológicas, médicas e farmacêuticas. Além da descoberta de novos compostos bioativos, também trata do estudo do metabolismo, do modo de ação em nível molecular e da criação de relações entre a estrutura química e atividade biológica.⁴

3. Indústria Farmacêutica, Pesquisa e Desenvolvimento

O mercado farmacêutico é um setor extremamente competitivo e caracterizado por altos níveis de investimentos e elevado risco. O desenvolvimento de uma NCE, desde a sua concepção até a introdução no mercado farmacêutico, leva em média de 12 a 15 anos, com investimentos da ordem de US\$ 800-880 milhões.^{5,6}

Neste amplo contexto, a sobrevivência e desenvolvimento da indústria farmacêutica requerem um forte comprometimento com a inovação científica e tecnológica através de programas bem estabelecidos de pesquisa e desenvolvimento (P&D). O sucesso destas iniciativas, deve resultar na introdução de fármacos pioneiros no mercado, ou ao menos em novas gerações de NCEs capazes de competir no mercado com fármacos já comercializados em suas respectivas classes terapêuticas. Dentre todas as

indústrias competitivas do mundo, a indústria farmacêutica é aquela que mais investe recursos em P&D.⁷

4. O processo de Desenvolvimento de Novos Fármacos

O processo de descoberta e desenvolvimento de novos fármacos é dividido em duas grandes fases: (i) pré-clínica (descoberta) e (ii) clínica (desenvolvimento). Nos estágios iniciais da fase pré-clínica, as pesquisas se concentram na identificação de moléculas pequenas capazes de modular a atividade de alvos macromoleculares, com elevado potencial de uso terapêutico.⁴

A estrutura tridimensional do alvo escolhido pode ser conhecida, permitindo o planejamento baseado na estrutura do receptor (SBDD, do inglês, *structure-based drug design*). Graças a grande evolução das técnicas de biologia estrutural, como a cristalografia de raios-x, um considerável número de alvos moleculares tem suas estruturas tridimensionais disponíveis.⁸ Caso a estrutura do receptor biológico não seja conhecida, o processo pode ser conduzido através do planejamento baseado na estrutura do ligante (LBDD, do inglês, *ligand-based drug design*). Em muitos casos, o uso combinado das duas estratégias pode ser útil no processo de planejamento, gerando conhecimentos novos através da sinergia entre as técnicas.

Uma vez identificado um alvo molecular promissor, é fundamental a sua validação, ou seja, é preciso verificar qual é a relação deste alvo com a patologia em questão. É necessário investigar as vias bioquímicas envolvidas e estabelecer se a modulação seletiva da proteína-alvo, através

de moléculas pequenas, é capaz de gerar a resposta esperada no controle da doença ou desordem.^{4,9}

No processo de identificação e seleção de compostos capazes de interagir com o alvo selecionado, é possível explorar o imenso espaço químico com o auxílio de técnicas em química medicinal, como a organização de bases padrões de dados, aplicação de filtros moleculares, e emprego de métodos modernos como os ensaios biológicos automatizados em larga escala (HTS, do inglês, *high-throughput screening*) e o ensaio virtual (VS, do inglês, *virtual screening*), além do estudo das relações quantitativas entre a estrutura e atividade (QSAR). Em todos estes casos, o gerenciamento qualificado da informação facilita a geração de conhecimento apropriado com base nas informações disponíveis. Neste estágio da pesquisa, busca-se identificar moléculas que apresentem propriedades desejadas para um fármaco, predominantemente potência e seletividade, mas que ainda precisam ser otimizadas em relação a outras propriedades. Os compostos que se destacam nesta fase são selecionados como compostos-líderes (do inglês, *lead compounds*).^{9,10}

A fase seguinte visa à otimização dos compostos-líderes a cerca de suas propriedades farmacodinâmicas e farmacocinéticas. Neste sentido a química medicinal moderna oferece várias estratégias capazes de gerar coleções dirigidas de compostos. Os estudos de SAR e QSAR são importantes para guiar a síntese de novos compostos, minimizando o universo de moléculas a serem consideradas nos programas de triagens em P&D.¹¹ Idealmente, parâmetros como potência, seletividade e afinidade, além de propriedades farmacocinéticas como ADME (Absorção, Distribuição, Metabolismo e Excreção), devem ser considerados conjuntamente, a fim de

eliminar candidatos com propriedades inadequadas nos estágios iniciais, reduzindo os custos do processo de P&D.¹² Em suma, na fase pré-clínica de pesquisa básica, são realizados testes que visam avaliar o potencial terapêutico de um composto e sua segurança em relação aos testes em humanos que serão realizados na fase clínica.^{4,10}

Após a conclusão da fase pré-clínica, as NCEs selecionadas são submetidas à aprovação do FDA (do inglês, *Food and Drug Administration*) nos Estados Unidos para autorização de realização de testes em seres humanos através de licença para investigação de um novo candidato a fármaco (IND, do inglês, *investigational new drug*).⁶ Nas fases clínicas são realizados diversos estudos para determinar a eficácia e a segurança das NCEs para uso clínico. Na fase clínica I, o composto é testado em um pequeno grupo de indivíduos saudáveis para determinar a faixa de dosagem na qual o composto pode ser administrado com segurança. Estudos farmacocinéticos examinam propriedades importantes de ADME. A fase II tem o propósito de determinar se o composto é eficaz em pacientes com a doença-alvo. A segurança e os possíveis efeitos adversos continuam sendo avaliados, além de serem realizados estudos relacionados à dose e regime de tratamento. Na fase clínica III, o composto é testado em um número grande de pacientes (de mil a cinco mil) de diversas regiões geográficas, com o objetivo de gerar dados suficientes em termos de uma avaliação final sobre a eficácia e segurança do novo candidato a fármaco.^{9,10} Ao término da fase III, todos os dados são organizados e submetidos ao FDA para solicitação da aprovação de um novo fármaco (NDA, do inglês, *New Drug Application*). Após a aprovação da NDA, o fármaco é lançado no mercado, mas continua sendo monitorado. Isso ocorre na fase clínica IV (pós-

lançamento), onde os testes prosseguem visando pesquisar aspectos específicos sobre o uso do fármaco, como segurança e eficácia em longo prazo, além do acompanhamento de efeitos colaterais e toxidez em grupos específicos de indivíduos.¹⁰

5. A Síndrome da Imunodeficiência Adquirida

A Síndrome da Imunodeficiência Adquirida (AIDS, do inglês, *Acquired Immune Deficiency Syndrome*), causada pelo vírus da imunodeficiência humana (HIV), é caracterizada pela perda progressiva de linfócitos T auxiliares CD4⁺, levando a uma severa depressão do sistema imunológico, infecções oportunistas, complicações neurológicas e neoplasias. O Centro de Controle de Doenças (CDC, do inglês, *Centers for Disease Control*) define como portadores da AIDS, os indivíduos que possuem uma contagem de células T CD4⁺ menor que 200 células por milímetro cúbico de sangue, infecção pelo HIV e a presença de pelo menos uma das condições associadas à AIDS, como por exemplo o Sarcoma de Kaposi.¹³⁻¹⁶

Mais de 25 milhões de pessoas perderam suas vidas em função da AIDS desde que ela foi reconhecida nos Estados Unidos em 1981, fazendo desta uma das mais destrutivas epidemias da história. Apesar dos recentes avanços e melhoramentos nas terapias anti-retrovirais, 2,9 milhões de pessoas morreram somente no ano de 2006 em todo o mundo. O número total de pessoas vivendo com o vírus HIV é estimado em 39,5 milhões, sendo que 4,3 milhões de novos casos foram detectados em 2006.¹³

De acordo com a Organização Mundial da Saúde (OMS), a descoberta de novos medicamentos é uma das questões prioritárias no combate eficaz à doença. A resistência e a ineficácia de alguns fármacos é uma questão cada vez mais preocupante. O combate à AIDS envolve esforços simultâneos no tratamento e prevenção em diversas regiões geográficas, através de programas governamentais sustentáveis.¹⁴ Na América Latina o número de pessoas infectadas pelo vírus HIV é estimado em 1,7 milhões. Países como Argentina, Colômbia e Brasil concentram o maior número de casos, sendo que o Brasil responde por um terço dos 1,7 milhões de casos.¹³

A transmissão do vírus HIV ocorre sob condições que proporcionam a troca de sangue ou líquidos corporais contendo o vírus ou células infectadas. Portanto, as vias principais de transmissão são através do ato sexual, inoculação parenteral e passagem do vírus de mães infectadas para os recém-nascidos. A transmissão sexual é claramente o modo predominante de transmissão no mundo inteiro, respondendo por cerca de 75% dos casos. A transmissão parenteral ocorre principalmente entre viciados em drogas administradas por via intravenosa. A transmissão através de mães infectadas pode se dar por via transplacentária, durante o nascimento através do canal de parto, ou ainda após o nascimento por ingestão de leite materno.^{16,17}

O vírus HIV, um retrovírus pertencente à família dos lentivírus, é dividido em duas formas geneticamente diferentes, mas relacionadas, o HIV-1 e HIV-2. O HIV-1 é o tipo mais comum nos Estados Unidos e Europa, enquanto o HIV-2 é mais encontrado na África.¹⁸

À semelhança da maioria dos retrovírus, o HIV é esférico, sendo envolvido por um envelope composto por uma bicamada lipídica derivada da membrana celular de células do hospedeiro. O envelope é salpicado por duas glicoproteínas virais denominadas gp120 e gp41. Abaixo do envelope existe uma matriz protéica que circunda o capsídeo, constituída pela proteína p17. O capsídeo tem a forma de um cone, sendo composto pela proteína p24. No interior do capsídeo, está contido o material genético do vírus, que consiste em duas cópias de RNA, e as enzimas protease, transcriptase reversa e integrase.¹⁹ Seu genoma é composto pelos três genes retrovirais comuns, *env*, *gag* e *pol*, além de vários outros genes: *tat*, *rev*, *vif*, *nef*, *vpr* e *vpu*, que regulam a síntese e a articulação das partículas virais. Os produtos dos genes *gag* e *pol* são traduzidos inicialmente em proteínas precursoras que devem ser clivadas pela protease viral para fornecer as proteínas maduras.²⁰

O vírus tem como alvo linfócitos T auxiliares, macrófagos e células dendríticas, ou seja, células do sistema imunológico que possuem o receptor de superfície CD4.^{15,17} Primeiramente, ocorre a adsorção do vírus à superfície destas células através da ligação entre a proteína do envelope viral gp120 e o receptor de membrana CD4. Contudo, a ligação ao CD4 não é suficiente. A proteína gp120 também deve se ligar a outras moléculas da superfície celular, denominadas co-receptores. Duas moléculas da superfície celular, CCR5 e CXCR4, desempenham a função de co-receptores.¹⁹ A etapa seguinte envolve alterações na conformação da gp41. Estas alterações resultam na introdução de um peptídeo de fusão através da membrana celular, resultando na fusão entre o envelope viral e a membrana celular. Após a fusão, o vírus injeta na célula seu material genético constituído por

RNA, juntamente com as enzimas transcriptase reversa, protease e integrase.^{16,17} A transcriptase reversa então catalisa a biossíntese de DNA, utilizando como molde o RNA viral. Este DNA recém-sintetizado é transportado até o núcleo celular onde é integrado ao genoma celular através da ação da integrase, resultando na formação de um pró-vírus.²¹ Neste estágio a infecção da célula se torna permanente. O pró-vírus pode permanecer latente por um longo tempo, que varia de alguns meses a muitos anos. Por outro lado, o DNA pró-viral pode ser transcrito, resultando na formação de partículas virais.^{19,20} A transcrição dos genes virais leva à biossíntese de duas classes de moléculas de RNA mensageiro (RNAm), as que sofrem união (do inglês, *splicing*) e aquelas que não sofrem este processo. O RNAm processado, é transportado para o citoplasma, onde ocorre o processo de tradução, com a biossíntese de três proteínas reguladoras: tat, rev e nef. O RNAm que não sofre união é retido no núcleo até que a proteína rev, se ligue a este RNA e o transporte até o citoplasma. Este RNA é traduzido dando origem às proteínas estruturais e às enzimas virais. A partir deste ponto, várias partículas virais são estruturadas e liberadas, causando a morte das células infectadas.^{20,22}

6. Aspectos Clínicos da AIDS

A AIDS é uma doença caracterizada por imunossupressão profunda que leva a infecções oportunistas, neoplasias secundárias e manifestações neurológicas. Há dois alvos principais do vírus HIV, o sistema imune e o sistema nervoso central (SNC).²³ Uma característica marcante da AIDS é a imunossupressão profunda que afeta principalmente a imunidade celular,

como resultado da infecção e perda acentuada de células T CD4⁺, com influência direta em todo o conjunto celular do sistema imune.^{23,24}

A infecção de monócitos e macrófagos também é extremamente importante na patogenia da infecção pelo HIV, levando a implicações importantes. Os monócitos e macrófagos representam uma verdadeira fábrica e reservatório do vírus, cuja produção permanece protegida das defesas do hospedeiro, e nos estágios avançados da infecção, quando o número de células T declina intensamente, os macrófagos podem ser o principal local de replicação viral.²⁵ As células dendríticas também são alvos importantes para o desencadeamento e manutenção da infecção. As células dendríticas capturam o vírus e o transportam para linfonodos regionais, onde as células T são infectadas.^{22,23}

O sistema nervoso central também é um alvo importante na infecção pelo HIV. O vírus é levado ao cérebro por monócitos infectados. Os neurônios não são infectados pelo vírus, portanto os danos neurológicos são causados indiretamente por produtos virais e fatores solúveis produzidos pelos macrófagos.^{15,19,24}

O perfil típico dos pacientes com AIDS caracteriza-se pela presença de febre, perda ponderal, diarreia, linfadenopatia generalizada, infecções oportunistas múltiplas, doenças neurológicas e em muitos casos, neoplasias secundárias.¹⁹ Estas manifestações são apresentadas de forma resumida na Tabela 1.

Tabela 1. Principais manifestações clínicas da AIDS.

INFECÇÕES OPORTUNISTAS, NEOPLASIAS E MANIFESTAÇÕES NEUROLÓGICAS DEFINIDORAS DA AIDS	
Infecções por Protozoários	
Criptosporidiose (enterite)	
Toxoplasmose (pneumonia ou infecção do SNC)	
Infecções Fúngicas	
Pneumocistose (pneumonia ou infecção disseminada)	
Candidíase (esofágica, traqueal ou pulmonar)	
Criptococose (infecção do SNC)	
Coccidioidomicose (disseminada); Histoplasmose (disseminada)	
Infecções Bacterianas	
Micobacteriose (tuberculose pulmonar ou extrapulmonar)	
Nocardiose (pneumonia ou meningite)	
Salmonelose	
Infecções Virais	
Citomegalovírus (infecção pulmonar, intestinal, retinite ou do SNC)	
Vírus herpes simples (localizado ou disseminado)	
Vírus varicela-zoster (localizado ou disseminado)	
Neoplasias	
Sarcoma de Kaposi	
Linfoma não-Hodgkin de células B	
Câncer invasivo do colo uterino	
Manifestações Neurológicas	
Demência da AIDS	
Linfoma do SNC	

7. Fármacos e a Terapia da AIDS

O tratamento de pacientes com AIDS consiste na administração de combinações de medicamentos anti-retrovirais, conhecidos como “coquetéis” anti-HIV, além de outros medicamentos capazes de minimizar as complicações decorrentes, como antibióticos, antifúngicos, antivirais e antineoplásicos.^{26,27} Embora várias pesquisas estejam em andamento, não existem vacinas para prevenir a infecção pelo vírus HIV e a AIDS.

Não há cura para a AIDS, embora os fármacos sejam capazes de reduzir de forma significativa a carga viral, eles não eliminam completamente o vírus do organismo, devido ao reservatório de células que apresentam infecção latente.²⁸ Portanto, a terapia medicamentosa tem como objetivos: (i) a redução da morbidade e mortalidade associada à AIDS; (ii) a melhoria da qualidade de vida dos indivíduos infectados; (iii) a recuperação e preservação do sistema imunológico; e, (iv) a redução máxima da carga viral pelo maior intervalo de tempo possível.²⁶

Devido à sua alta taxa de mutação, existe o surgimento de diferentes linhagens do vírus, algumas delas resistentes aos fármacos disponíveis. Portanto, o tratamento padrão consiste na combinação de vários fármacos, o que é conhecido como Terapia Anti-retroviral Altamente Ativa (HAART, do inglês, *Highly Active Antiretroviral Therapy*). Esta estratégia, que tipicamente combina fármacos de pelo menos duas classes diferentes, tem se mostrado eficaz na redução da carga viral.²⁹ Porém, a emergência de resistência aos medicamentos existentes tem gerado uma grande expectativa por novos fármacos.^{28,30}

Existem atualmente 28 medicamentos anti-HIV aprovados pelo FDA, divididos em 3 grandes classes.³¹

i. Inibidores da Transcriptase Reversa: interferem em um estágio crítico do ciclo viral, denominado transcrição reversa, quando a transcriptase reversa converte o RNA viral em DNA. Existem 2 tipos de inibidores nesta classe. Os inibidores nucleosídicos da transcriptase reversa (NRTI, do inglês, *nucleoside reverse transcriptase inhibitors*), que mimetizam nucleotídeos e quando incorporados ao DNA viral, interrompem o processo de polimerização; e os inibidores não-nucleosídicos da transcriptase reversa (NNRTI, do inglês, *non-nucleoside reverse transcriptase inhibitors*), que se ligam a um sítio alostérico da enzima, provocando mudanças conformacionais importantes que impedem a ligação dos nucleotídeos.

ii. Inibidores da Protease: ligam-se ao sítio ativo da enzima, impedindo que poliproteínas precursoras sejam clivadas originando as proteínas virais maduras.

iii. Inibidores de fusão: interagem com a proteína de fusão gp41, impedindo que ocorra a fusão entre o envelope viral e a membrana celular.

A Tabela 2 apresenta uma lista dos medicamentos anti-HIV, divididos em suas respectivas classes.³¹

Tabela 2. Medicamentos aprovados pelo FDA para o tratamento da AIDS.

Medicamentos Combinados			
Nome comercial	Nome Genérico	Companhia Farmacêutica	Aprovado
Atripla	Efavirenz, emtricitabina e tenofovir	Bristol-Myers Squibb e Gilead Sciences	07/2006
<i>Inibidores Nucleosídicos da Transcriptase Reversa</i>			
Combivir	Lamivudina e zidovudina	GlaxoSmithKline	09/1997
Emtriva	Emtricitabina	Gilead Sciences	07/2003
Epivir	Lamivudina	GlaxoSmithKline	11/1995
Epzicom	Abacavir e lamivudina	GlaxoSmithKline	08/2004
Hivid	Zalcitabina e dideoxycitidina	Hoffmann-La Roche	06/1992
Retrovir	zidovudina, azidotimidina	GlaxoSmithKline	03/1987
Trizivir	Abacavir, zidovudina, e lamivudina	GlaxoSmithKline	11/2000
Truvada	Tenofovir e emtricitabina	Gilead Sciences, Inc.	08/2004
Videx EC	Didanosina	Bristol Myers-Squibb	10/2000
Videx	Didanosina e dideoxyinosina	Bristol Myers-Squibb	10/1991
Viread	Tenofovir	Gilead	10/2001
Zerit	Stavudina	Bristol Myers-Squibb	06/1994
Ziagen	Abacavir	GlaxoSmithKline	12/1998
<i>Inibidores não Nucleosídicos da Transcriptase Reversa</i>			
Rescriptor	Delavirdina	Pfizer	04/1997
Sustiva	Efavirenz	Bristol Myers-Squibb	09/1998
Viramune	Nevirapina	Boehringer Ingelheim	06/1996
Inibidores da Protease			
Agenerase	Amprenavir	GlaxoSmithKline	04/1999
Aptivus	Tipranavir	Boehringer Ingelheim	06/2005
Crixivan	Indinavir	Merck	03/1996
Invirase	Saquinavir	Hoffmann-La Roche	12/1995
Kaletra	Lopinavir e ritonavir	Abbott	09/2000
Lexiva	Fosamprenavir	GlaxoSmithKline	10/2003
Norvir	Ritonavir	Abbott	03/1996
Prezista	Darunavir	Tibotec Inc.	06/2006
Reyataz	Atazanavir	Bristol-Myers Squibb	06/2003
Viracept	Nelfinavir	Agouron Pharmaceuticals	03/1997
Inibidores de Fusão			
Fuzeon	Enfuvirtide	Hoffmann-La Roche & Trimeris	03/2003

Devido aos incômodos, complexos regimes e falta de informação, muitos pacientes apresentam dificuldades em aderir ao tratamento de forma correta. Geralmente, a administração de alguns medicamentos várias vezes ao dia provoca efeitos adversos, como náusea e vômitos, além de outros efeitos relacionados a alterações metabólicas, como distribuição anormal de gordura, metabolismo anormal de carboidratos e lipídios, e perda de massa óssea.²⁸

O desenvolvimento de novos agentes terapêuticos para o emprego eficaz e seguro no tratamento da AIDS é um dos principais alvos da indústria farmacêutica moderna. Essa tendência continuará em evidência por muitos anos, pois a demanda será alta devido à resistência e aos efeitos colaterais associados à HAART.^{14,28,32}

8. A Protease do Vírus HIV-1

As enzimas são consideradas por muitas indústrias farmacêuticas os alvos moleculares mais atrativos para o desenvolvimento de novos fármacos, devido ao seu papel essencial em muitos processos fisiopatológicos.³³ Este aspecto é evidenciado pelos números que demonstram que cerca de metade dos fármacos atualmente em uso clínico são inibidores enzimáticos.^{33,34}

Na metade da década de 80, estudos acerca do genoma do vírus HIV-1, e comparações com seqüências genômicas de outros retrovírus, levaram à descoberta de que o vírus produzia uma peptidase. Estudos de inativação desta enzima, através de técnicas de mutação e inibição enzimática, levaram à constatação de que partículas virais imaturas e não-infecciosas

eram produzidas. Concluiu-se, portanto, que esta peptidase era essencial para a maturação adequada do vírus.³⁵ Desde então, a protease do vírus HIV-1 (HIV-1 PR, (EC 3.4.23.16) tem sido alvo de inúmeros estudos, tanto em relação à sua estrutura quanto ao seu mecanismo de ação.^{34,36} Diversas estruturas da HIV-1 PR estão disponíveis, bem como vários bancos de dados contendo informações sobre estrutura-atividade. Este grande volume de conhecimento gerado acerca desta enzima impulsiona as pesquisas em química medicinal moderna na busca por novos fármacos para a terapia da AIDS.³⁷

A HIV-1 PR é traduzida como uma poliproteína precursora codificada pelos genes *gag-pol*. Foi a primeira enzima do vírus HIV-1 a ser estruturalmente caracterizada. Consiste em um homodímero de 22kDa, onde cada monômero possui 99 aminoácidos. Sua estrutura é relacionada àquela das proteases aspárticas celulares, como a pepsina. O sítio ativo possui a forma de um canal, situado na interface do dímero, com cada monômero contribuindo com um dos resíduos catalíticos de aspartato.^{36,38} O sítio ativo é dividido em subsítios que por convenção são denominados S1, S2 e S3 em uma das direções a partir dos aspartatos catalíticos, e S1', S2' e S3' na direção oposta. Apesar da presença de resíduos de aspartato, o sítio ativo da HIV-1 PR apresenta caráter predominantemente hidrofóbico.³⁴ A porção superior do sítio ativo é formada por estruturas denominadas *flaps* que sofrem alterações conformacionais essenciais à ligação do substrato.³⁹ Uma molécula de água, que corresponde à água catalítica presente nesta família de proteases, é encontrada entre os resíduos de aspartato, onde realiza ligações hidrogênio com o grupo carboxilato destes resíduos.³⁷ A Figura 1 ilustra a estrutura da HIV-1 PR.

Figura 1. Representação ilustrativa da estrutura da protease do vírus HIV-1.

A protease do vírus HIV-1 é responsável pela clivagem das poliproteínas precursoras codificadas pelos genes *gag* e *gag-pol*, para liberar as proteínas estruturais e as enzimas virais.⁴⁰ Em relação à hidrólise das ligações peptídicas, a HIV-1 PR utiliza o mecanismo comum às proteases aspárticas, ou seja, uma molécula de água posicionada entre os dois resíduos de aspartato catalíticos é utilizada para o ataque nucleofílico ao carbono carbonílico da amida da ligação peptídica a ser clivada.⁴¹ Este ataque nucleofílico gera um intermediário tetraédrico que é rapidamente quebrado, originando as extremidades C-terminal e N-terminal das proteínas que serão utilizadas na maturação vírus HIV-1.⁴²

9. Inibidores da HIV-1 PR e o Desenvolvimento de NCEs

De acordo com o *National Institute of Health* (NIH) dos Estados Unidos, o uso de inibidores da protease na terapêutica a partir de 1995, combinado com outros fármacos anti-retrovirais, foi responsável por uma queda de aproximadamente 70% nas mortes causadas pela AIDS, dados que ilustram a grande importância desta classe de fármacos.⁴³

Os inibidores da HIV-1 PR representam um exemplo de extremo sucesso do planejamento de fármacos baseado na estrutura do receptor (SBDD).³⁵ Estes fármacos (Tabela 2) impedem que poliproteínas precursoras sejam clivadas dando origem às proteínas funcionais do vírus, interrompendo o ciclo de reprodução viral.⁴⁰ A estratégia utilizada no desenvolvimento de inibidores da protease é baseada na substituição da ligação peptídica por grupos não hidrolisáveis, que de alguma forma mimetizem o intermediário de reação tetraédrico produzido na reação proteolítica. Vários grupos químicos são comumente utilizados nesta estratégia, como por exemplo as hidroxietilaminas e os hidroxietilenos. O primeiro inibidor da HIV-1 PR, o saquinavir (Invirase®), da Hoffmann-La Roche, foi introduzido no mercado farmacêutico em 1995. Este fármaco pioneiro possui um grupo hidroxietilamina em sua estrutura.³⁴

Invariavelmente, os inibidores da HIV-1 PR são encontrados nos complexos enzima-inibidor sob a forma estendida, refletindo a forma em canal do sítio ativo da enzima. Uma característica comum a essas moléculas são as ligações hidrogênio formadas entre os inibidores e a molécula de água existente no sítio ativo da enzima.³⁸ As uréias cíclicas e as hidroxipironas são exemplos de inibidores em fase de desenvolvimento que

possuem grupos que substituem esta molécula de água. Atualmente, existem dez inibidores da protease aprovados pelo FDA. Os principais desafios que precisam ser superados nas próximas gerações de inibidores da protease, incluem maior potência e seletividade frente a cepas resistentes do vírus, maior biodisponibilidade oral, maior tempo de meia-vida e menor toxidez.³⁹

10. Química Medicinal e as Técnicas de QSAR

Uma das técnicas muito empregadas em química medicinal no planejamento de NCEs candidatas a fármacos é o estudo das relações quantitativas entre a estrutura e atividade (QSAR, do inglês, *Quantitative Structure-Activity Relationships*).⁴⁴

As técnicas de QSAR são empregadas com a finalidade de se encontrar moléculas com propriedades superiores em relação às aquelas apresentadas pelo conjunto de dados que deu origem ao processo de modelagem. A propriedade alvo, geralmente potência ou afinidade, deve ser padronizada e validada para ser útil em estudos de QSAR. Propriedades moleculares são calculadas e empregadas como descritores. Métodos estatísticos são usados para produzir uma relação que descreve as variações na propriedade alvo, em função dos descritores. Esta relação é denominada QSAR, a qual pode ser utilizada em programas de planejamento para orientar a síntese de novas moléculas com propriedades otimizadas. O processo cíclico de formulação da QSAR, predição da propriedade alvo, síntese e ensaios biológicos, prossegue até que moléculas qualificadas, candidatas a NCEs sejam selecionadas.⁴⁵

11. Holograma QSAR

O método holograma QSAR (HQSAR, do inglês, *Hologram Quantitative Structure-Activity Relationships*), baseia-se no conceito de impressão digital molecular (do inglês, *fingerprint*). A premissa da HQSAR é que a estrutura molecular pode ser convertida em notações lineares, ou seja, seqüências numéricas que codificam a estrutura bidimensional das moléculas do conjunto de dados. Baseando-se no conceito central da QSAR, de que a atividade biológica é função da estrutura molecular, é possível fazer predições a respeito da atividade de uma molécula a partir das informações codificadas nos hologramas moleculares.^{46,47}

Os hologramas moleculares empregados na técnica de HQSAR são uma forma diferente de impressão digital. O holograma é gerado a partir da fragmentação da estrutura bidimensional das moléculas, onde todos os fragmentos possíveis são obtidos e posteriormente distribuídos pelo holograma molecular. Cada posição do holograma contém o número de fragmentos alocados nesta determinada posição, sendo que fragmentos iguais que ocorrem mais de uma vez são sempre colocados na mesma posição. Desta forma, o holograma contém informações a respeito dos tipos e da quantidade de fragmentos moleculares. Isto distingue o holograma da impressão digital tradicional, onde a notação é simplesmente binária, contendo apenas caracteres 0 e 1.⁴⁸

Durante o processo de modelagem de HQSAR, há a possibilidade de se definir vários parâmetros relevantes na obtenção dos hologramas, como o comprimento do holograma, o tamanho dos fragmentos e os parâmetros utilizados para distinguir diferentes fragmentos.⁴⁸ O tamanho dos

fragmentos é definido pelo número máximo (N) e mínimo (M) de átomos que cada fragmento gerado poderá conter. Os parâmetros apresentados na Tabela 3 podem ser utilizados para definir as diferenças entre os fragmentos. Há a liberdade de se gerar modelos selecionando todos os parâmetros, ou combinações de apenas alguns parâmetros, escolhendo ao final aquela combinação que produz o resultado estatístico mais relevante.

Tabela 3. Parâmetros de distinção de fragmentos no método HQSAR.

Parâmetro de distinção de fragmento	Definição
Átomos (A)	Permite que os fragmentos sejam distinguidos com base nos tipos de átomos.
Ligações (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 átomos quirais.
Doador e Receptor (DA)	Esta opção permite que os fragmentos sejam distinguidos com base na presença de átomos aceptores ou doadores de ligações hidrogênio.

A partir da geração dos hologramas moleculares, o método de regressão por mínimos quadrados parciais (PLS, do inglês, *partial least squares*) correlaciona a atividade biológica com a informação contida nos hologramas, gerando um modelo estatístico de QSAR. Uma nova molécula que ainda não foi testada pode então ter sua atividade predita quando submetida a este modelo.⁴⁷ O processo é ilustrado na Figura 2.

Figura 2. Geração de um modelo de HQSAR.

Um importante papel de um modelo de QSAR, além de prever a atividade de moléculas não testadas, é fornecer indicações sobre quais fragmentos moleculares são importantes para a atividade biológica. Estas informações podem orientar a síntese de novas moléculas com propriedades otimizadas. No método de 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 fragmentos moleculares para a atividade das

moléculas.⁴⁸ Esta metodologia, que atualmente faz parte de um leque de tecnologias disponíveis em química medicinal, é capaz de gerar modelos de QSAR com alta capacidade de interpolação e extrapolação em diversos conjuntos de dados.

12. Análise Comparativa dos Campos Moleculares (CoMFA)

O método da análise comparativa dos campos moleculares (CoMFA, do inglês, *Comparative Molecular Field Analysis*) é a técnica de QSAR 3D mais utilizada. O conceito no qual se baseia, é o de que a atividade biológica está diretamente relacionada com as interações fármaco-receptor. Estas interações apresentam caráter tridimensional, portanto, este método considera as moléculas em três dimensões, utilizando campos moleculares de interação como variáveis independentes. Estes campos de interação são obtidos através do cálculo da energia de interação entre as moléculas e sondas virtuais, que consistem em determinado tipo de átomo. As diferenças na atividade biológica (variável dependente) são desta forma, descritas em função das diferenças nestes campos de interação. O método CoMFA emprega predominantemente os campos de interações eletrostáticas e interações estereoquímicas, baseando-se na hipótese de que as interações fármaco-receptor são extensivamente dirigidas por estas forças de natureza não-covalente.⁴⁹

O conjunto de moléculas precisa ser submetido a um alinhamento estrutural, ou seja, as moléculas precisam apresentar conformações semelhantes e mesma orientação no espaço cartesiano. Esta é a principal dificuldade inerente a este método, quando comparado aos métodos de

QSAR 2D, como o HQSAR. Obviamente, cargas atômicas devem ser incluídas nas moléculas, desde que interações eletrostáticas serão calculadas.⁴⁵ As moléculas alinhadas e superpostas, são colocadas em uma grade tridimensional reticulada. Em cada ponto de interseção desta grade, são calculadas as energias de interação estereoquímica e eletrostática entre uma sonda virtual, que pode ser por exemplo, um átomo de carbono, e os átomos próximos a este ponto. Este processo é realizado até que todos os pontos sejam escaneados pela sonda. A partir destes cálculos, é gerada uma tabela contendo em uma das colunas, o valor da atividade biológica (variável dependente), e nas outras colunas os descritores moleculares (variáveis independentes). Cada ponto da grade é representado por duas colunas, uma delas contendo o valor da energia de interação eletrostática e a outra o valor para a interação estereoquímica.⁸ A partir desta tabela, o método de regressão PLS, gera um modelo estatístico correlacionando a propriedade alvo com os descritores. Este modelo permite que novas moléculas que ainda não foram sintetizadas ou testadas, possam ter o valor da propriedade biológica em questão predito.^{8,49} Este processo é esquematizado na Figura 3.

Figura 3. Geração de um modelo CoMFA.

Como no HQSAR, o método CoMFA permite a avaliação visual de modificações na estrutura das moléculas que levem a um aumento da atividade biológica. Para tanto, poliedros que apresentam um código de cores, denominados mapas de contorno, são construídos em torno das moléculas.⁴⁵

A metodologia CoMFA é uma das técnicas de QSAR mais empregadas pela indústria farmacêutica na construção de modelos preditivos para uma série de propriedades biológicas e moleculares.

CAPÍTULO II

OBJETIVOS

OBJETIVOS

Os objetivos principais desta dissertação de mestrado são:

- ✓ Organizar um conjunto de dados padrão de inibidores da protease do vírus HIV-1 para uso na modelagem de QSAR.
- ✓ Desenvolver modelos de QSAR 2D e QSAR 3D, empregando as técnicas HQSAR e CoMFA, respectivamente, para um conjunto padrão de inibidores da protease do vírus HIV-1.
- ✓ Avaliar a robustez estatística dos modelos gerados, o seu poder de correlação nos conjuntos treinamento e predição externa nos conjuntos teste.
- ✓ Obter informações a partir dos mapas de contribuição 2D, gerados pelo método de HQSAR, e dos mapas de contorno 3D, gerados pelo método CoMFA, e utilizá-las como guias em futuros trabalhos em Química Medicinal no planejamento de novos inibidores da protease do vírus HIV-1 mais potentes e seletivos.

CAPÍTULO III

PARTE EXPERIMENTAL

PARTE EXPERIMENTAL

1. Modelagem de QSAR

Todos os trabalhos de modelagem molecular, análises estatísticas e produção dos modelos de QSAR 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 da Universidade de São Paulo. Estas estações operam no sistema operacional Linux, utilizando-se, no desenvolvimento deste trabalho, as interfaces computacionais SYBYL 7.2[®] (Tripos, St. Louis, EUA) e GOLD 3.0[®] (Cambridge Crystallographic Data Centre, Cambridge, UK)

2. Conjunto de Dados

Os inibidores da HIV-1 PR empregados no desenvolvimento dos modelos de QSAR foram selecionados através de pesquisas na literatura.^{50,51} O conjunto de dados final consiste em uma série de 82 moléculas estruturalmente relacionadas, associadas aos valores de potência (IC₅₀) correspondentes (Tabela 4). As estruturas dos inibidores foram criadas utilizando-se as ferramentas disponíveis no programa computacional SYBYL 7.2[®], sendo posteriormente submetidas aos procedimentos de minimização de energia e cálculo de cargas atômicas parciais através do método AM1, no programa MOPAC 6.0. As estruturas 3D geradas foram

empregadas no processo de alinhamento estrutural, necessário para a realização do método CoMFA.

3. Alinhamento Estrutural

O alinhamento estrutural tridimensional é uma das etapas cruciais na modelagem de QSAR 3D CoMFA. A estratégia empregada para a obtenção do alinhamento molecular foi a docagem (acoplamento) das moléculas do conjunto de dados no sítio ativo da enzima, utilizando-se o programa GOLD 3.0[®]. Primeiramente, foi realizada uma pesquisa no banco de dados PDB (do inglês, *Protein Data Bank* – <http://www.rcsb.org/pdb>) para a seleção de uma estrutura cristalográfica da HIV-1 PR. Na indisponibilidade de uma estrutura da enzima em complexo com um dos inibidores da base de dados, os critérios empregados na análise foram a resolução das estruturas e a conservação dos aminoácidos de ligação. Com base nos critérios adotados, foi selecionada uma estrutura selvagem da enzima em complexo com uma aminohidroxissulfona (código PDB: 1XL5), com resolução de 1,73 Å.⁵²

O processo de preparo da estrutura selecionada (alvo) para a modelagem envolveu o seguinte procedimento: (i) verificação da conservação do sítio ativo; (ii) análise dos aminoácidos principais; (iii) remoção do ligante do sítio ativo da enzima; (iv) eliminação das moléculas de água; e, (v) adição de átomos de hidrogênio.

Depois de finalizados os processos de preparação do alvo e do conjunto de dados, foram estabelecidos os parâmetros de docagem. O programa GOLD 3.0[®] foi utilizado para a docagem dos inibidores no sítio

ativo da enzima. Os resultados são apresentados no Capítulo IV (item 3, QSAR 3D: CoMFA) desta dissertação.

4. Desenvolvimento de Modelos HQSAR

O módulo de HQSAR da plataforma SYBY 7.2[®] foi empregado na produção de um modelo de HQSAR para a série de inibidores da protease do vírus HIV-1 apresentada na Tabela 4.⁵³

As estruturas das moléculas do conjunto de dados e os valores correspondentes de pIC_{50} ($-\log IC_{50}$) foram transferidos para uma planilha molecular (MSS, do inglês, *molecular spreadsheet*), onde fica 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 consistem em: comprimento do holograma, tamanho dos fragmentos e diversos parâmetros de distinção de fragmentos. Várias análises de QSAR, empregando inúmeras combinações de parâmetros, foram realizadas com o objetivo de otimizar os modelos até produzir um modelo final de HQSAR com elevada consistência interna. Esse modelo foi então utilizado no processo de validação externa de um conjunto teste. Os resultados são apresentados no Capítulo IV desta dissertação.

5. Geração de Modelos CoMFA

O módulo CoMFA da plataforma SYBYL 7.2[®] foi empregado no desenvolvimento dos modelos de QSAR 3D, para o mesmo conjunto de

dados da HIV-1 PR utilizado na modelagem de HQSAR.⁵⁴ No método CoMFA, se faz necessário que as moléculas estejam alinhadas no espaço tridimensional (descrito anteriormente no item 3 deste capítulo). A atribuição de cargas atômicas parciais foi necessária para os cálculos das interações eletrostáticas e geração dos campos correspondentes através do método semi-empírico AM1.

Uma planilha molecular (MSS) foi criada para a realização das análises de QSAR 3D, contendo toda a informação necessária em relação à estrutura e atividade das moléculas do conjunto de dados. A seguir, foram especificados os parâmetros iniciais referentes à geração dos descritores e posteriormente foram realizadas várias análises de QSAR, introduzindo-se variações na geração dos descritores e nas análises estatísticas com a finalidade de refinar os modelos e produzir um modelo CoMFA com alta capacidade de correlação e predição para o conjunto de dados gerado. Os resultados são apresentados no Capítulo IV desta dissertação.

CAPÍTULO IV

RESULTADOS E DISCUSSÃO

RESULTADOS E DISCUSSÃO

1. Conjunto de Dados

O conjunto de dados da HIV-1 PR empregado no desenvolvimento dos modelos de QSAR 2D e 3D possui 82 moléculas associadas aos respectivos valores de potência. As estruturas químicas e os valores da propriedade biológica (IC_{50}) são apresentados na Tabela 4.

Tabela 4. Conjunto padrão de dados de inibidores da HIV-1 PR.

Conjunto Treinamento		
Código	Estrutura	IC_{50} (nM)
1		2,7
2		1,8
3		4,1
4		18,3

5

8,7

6

2,6

7

2,6

8

13,0

9

9,7

10

1,5

11

3,7

12

12,3

13

6,3

14

1,9

15

8,6

16

3,6

17

3,6

18

3,0

19

3,5

20

1,9

21

2,2

22

4,5

23

21,8

24

2,4

25

5,0

26

3,5

27

9,8

28

131

29

29,0

30

8,0

31

2,2

32

3,0

33

6,7

34

14,0

35

1,3

36

4,1

37

18,5

38

3,6

39

1,7

40

3,4

41

4,9

42

3,9

43

4,3

44

8,9

45

0,43

46

1,2

47

0,94

48

1,5

49

0,78

50

0,34

51

0,35

52

0,17

53

0,57

54

1,3

55

0,24

56

0,41

57

0,58

58

0,19

59

1,9

60		1,3
61		1,6
62		0,8
63		0,95
64		0,98
65		1,4

Conjunto Teste

Código	Estrutura	IC ₅₀ (nM)
66		4,0

75

2,0

76

21,0

77

3,6

78

2,2

79

3,3

80

0,28

81

0,45

82

0,72

A propriedade utilizada, IC_{50} , refere-se à concentração de inibidor requerida para diminuir em 50% a atividade enzimática. Os valores de IC_{50} variam de 0,17 a 131 nM, um fator de potência de aproximadamente 800 vezes. Todos os valores de IC_{50} foram determinados nas mesmas condições experimentais, o que é um fator essencial para o desenvolvimento de modelos confiáveis de QSAR.^{50,51}

O conjunto de dados apresenta razoável diversidade química. A estrutura representativa da diversidade química é apresentada na Figura 4. Pode-se verificar a presença do fragmento farmacofórico constituído pelos anéis hidroxipirona e 2-*terc*-butil-5-metilfenila conectados por um átomo de enxofre. Este sistema possui papel fundamental no processo de reconhecimento molecular desta série de inibidores frente à HIV-1 PR. A conservação da unidade farmacofórica é requerida para a manutenção da potência inibitória, sendo que a diversidade estrutural desta série é conferida pelos variados substituintes presentes nos fragmentos A, B e C da Figura 4.

Figura 4. Estrutura representativa da diversidade química do conjunto de dados.

Conforme pode ser observado na Tabela 4, as moléculas do conjunto de dados foram divididas em dois conjuntos, denominados conjunto treinamento e conjunto teste. O conjunto treinamento é formado pelas moléculas que são utilizadas para a geração dos modelos de QSAR, ou seja, é a partir desta informação sobre a estrutura e atividade das moléculas que são gerados os descritores e são realizadas as análises estatísticas. O conjunto teste é utilizado na avaliação externa dos modelos em relação a sua capacidade de prever a propriedade alvo de novas moléculas que não foram incluídas durante o processo de modelagem no conjunto treinamento.

Do total de 82 inibidores do conjunto de dados, 17 (20%) foram separados para o conjunto teste, sendo que os outros 65 (80%) permaneceram no conjunto treinamento (Tabela 4). O conjunto treinamento é formado pelos inibidores de **1** a **65**, enquanto o conjunto teste é formado pelos inibidores de **66** a **82**.

Os valores de IC_{50} das moléculas do conjunto de dados foram convertidos nos valores correspondentes na escala logarítmica de pIC_{50} ($-\log IC_{50}$). Portanto, para uma análise comparativa, quanto maior o valor de pIC_{50} mais potente será o composto. A distribuição dos valores de pIC_{50} é mostrada no histograma da Figura 5. Os valores variam de 6,88 a 9,77, cerca de 3 ordens de magnitude. É necessário salientar a boa distribuição existente nos conjuntos treinamento e teste em relação ao conjunto de dados. O conjunto teste, em especial, apresenta uma distribuição representativa em comparação com o conjunto treinamento. Essa distribuição apropriada ao longo do espectro da propriedade biológica representa um aspecto muito importante na avaliação da consistência interna e externa dos modelos de QSAR.

Figura 5. Representação gráfica da distribuição dos valores de pIC₅₀

Por todas estas razões, o conjunto padrão de dados organizado para os inibidores da HIV-1 PR é particularmente apropriado para o desenvolvimento dos modelos preditivos de QSAR 2D e 3D propostos nesta dissertação.

2. Holograma QSAR

O conjunto treinamento de dados empregado no trabalho de modelagem de HQSAR é apresentado na Tabela 4. Para a geração dos modelos, primeiramente, foram especificados os seguintes parâmetros relativos à obtenção dos descritores:

i. Comprimento do holograma: as análises foram realizadas utilizando-se os seguintes comprimentos de holograma: 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401.

ii. Tamanho dos fragmentos: inicialmente foi utilizado o tamanho de fragmento padrão 4-7.

iii. Distinção de Fragmentos: foram empregadas 13 diferentes combinações dos seguintes parâmetros de distinção: A (átomos), B (ligações), C (conectividade), H (átomos de hidrogênio), Ch (quiralidade) e DA (doadores e aceptores de ligação hidrogênio). Estas combinações são mostradas a seguir: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/C/H/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/Ch/DA, A/B/C/H/DA e A/B/H/Ch/DA.

Diversas análises estatísticas de HQSAR foram conduzidas com base nos parâmetros selecionados. Ao final do processo de modelagem, foram obtidos os resultados apresentados na Tabela 5.

Tabela 5. Resultados de HQSAR em função da variação dos parâmetros de distinção de fragmentos.

Distinção de fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
A/B	0,728	0,886	0,303	151	6
A/B/C	0,745	0,902	0,293	199	6
A/B/C/H	0,673	0,882	0,332	257	6
A/B/C/H/Ch	0,661	0,838	0,338	71	6
A/B/C/H/Ch/DA	0,630	0,846	0,353	151	6
A/B/H	0,668	0,881	0,336	257	6
A/B/C/Ch	0,737	0,892	0,296	199	5
A/B/DA	0,681	0,886	0,328	401	6
A/B/C/DA	0,721	0,903	0,307	199	6
A/B/H/DA	0,660	0,840	0,339	97	6
A/B/C/Ch/DA	0,731	0,907	0,301	199	6
A/B/C/H/DA	0,646	0,863	0,345	257	6
A/B/H/Ch/DA	0,642	0,832	0,348	97	6

q^2 : coeficiente de correlação com validação cruzada; r^2 : coeficiente de correlação; SEE: erro padrão da estimativa; HL: comprimento do holograma; N: número de componentes.

Como pode ser observado, o modelo que produziu os melhores resultados estatísticos ($q^2 = 0,745$ e $r^2 = 0,902$) foi aquele com a combinação de distinção de fragmentos A/B/C, apresentando comprimento de holograma 199 e número ótimo de componentes igual a 6. Este modelo foi selecionado para investigação da influência da variação do tamanho dos fragmentos sobre os parâmetros estatísticos. Desta forma, foram realizadas análises individuais empregando tamanhos de fragmentos distintos: 2-5, 3-6, 5-8, 6-9, 7-10 e 8-11, representados no intervalo pelo número mínimo e

máximo de átomos. A Tabela 6 apresenta os resultados obtidos neste processo.

Tabela 6. Análise de HQSAR em função da variação do tamanho dos fragmentos.

Tamanho dos fragmentos	Parâmetros Estatísticos				
	q^2	r^2	SEE	HL	N
2-5	0,697	0,864	0,320	199	6
3-6	0,712	0,891	0,312	151	6
4-7	0,745	0,902	0,293	199	6
5-8	0,748	0,903	0,292	199	6
6-9	0,732	0,900	0,301	199	6
7-10	0,761	0,905	0,284	151	6
8-11	0,732	0,895	0,306	151	6

Embora as diferenças não tenham sido significativas em alguns casos nesta série comparativa, o modelo gerado com o tamanho de fragmento 7-10 para a combinação A/B/C, foi aquele que apresentou os melhores resultados estatísticos ($q^2 = 0,761$ e $r^2 = 0,905$). O modelo final de HQSAR gerado com o conjunto treinamento possui a validação cruzada como indicativo maior de sua robustez. Este parâmetro demonstra a capacidade de correlação dos dados e a 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 QSAR apresente utilidade no planejamento de fármacos, é necessário que o mesmo passe por um processo de

validação externa para uma avaliação da sua real capacidade preditiva. O processo de validação externa do modelo de HQSAR mais robusto foi feito com o emprego do conjunto teste apresentado na Tabela 4. As moléculas contidas no conjunto teste (**66-82**) não foram utilizadas na derivação do modelo. Portanto, o conjunto teste é apropriado para avaliar o poder preditivo do modelo para novas moléculas dentro da mesma diversidade química. Os resultados do processo de validação externa são apresentados na Tabela 7.

As diferenças modestas entre os valores experimentais e preditos (resíduos da **Tabela 7**) indicam a boa capacidade preditiva do modelo gerado. Das 17 moléculas, 10 apresentaram resíduos abaixo de 0,2 unidades logarítmicas, 5 ficaram abaixo de 0,3, 1 abaixo de 0,4 e 1 ficou acima de 0,5 (valor de 0,592).

Tabela 7. Valores experimentais e preditos de pIC_{50} associados aos resíduos para as moléculas do conjunto teste no modelo HQSAR.

Código	pIC_{50} Experimental	pIC_{50} Predito	Resíduo
66	8,397	8,130	-0,267
67	8,275	8,518	0,243
68	8,657	8,946	0,289
69	8,455	8,393	-0,062
70	8,677	8,821	0,144
71	8,443	8,491	0,048
72	7,982	8,161	0,179
73	8,017	8,221	0,204
74	7,772	8,095	0,323
75	8,698	8,511	-0,187
76	7,677	7,726	0,049
77	8,443	8,346	-0,097
78	8,657	8,785	0,128
79	8,481	8,660	0,179
80	9,552	8,960	-0,592
81	9,346	9,069	-0,227
82	9,142	9,066	-0,076

A Figura 6 apresenta o gráfico dos valores experimentais contra os valores preditos para os conjuntos treinamento e teste. A análise gráfica mostra a boa concordância dos valores preditos e experimentais. O modelo final gerado apresenta, portanto, significativa consistência interna e externa para a diversidade estrutural estudada, possuindo comprovada capacidade preditiva.

Figura 6. Gráfico HQSAR dos valores experimentais e preditos para as 82 moléculas do conjunto de dados.

Além de prever a potência de novas moléculas não incluídas no conjunto treinamento, um modelo de HQSAR fornece importantes indicativos sobre os diversos fragmentos moleculares que estão relacionados com a atividade biológica. Esta informação é codificada em um mapa de contribuição, que apresenta cada átomo de uma determinada molécula e a sua contribuição para a atividade biológica. Neste mapa, um código de cores é atribuído, refletindo a contribuição individual de cada átomo e dos diferentes fragmentos da molécula. Cores na região do vermelho do espectro (vermelho, laranja avermelhado e laranja) refletem

contribuições desfavoráveis ou negativas. Cores na região do verde (amarelo, azul esverdeado e verde) refletem contribuições favoráveis ou positivas. Átomos com contribuições intermediárias são coloridos em branco. Na Figura 7 é mostrado o mapa de contribuição para o inibidor mais potente (**52**) do conjunto de dados.

Figura 7. Mapa de contribuição de HQSAR para o inibidor **52**.

Como pode ser observado, o mapa de contribuição fornece informações importantes, que associadas ao conhecimento em química sintética, podem ser úteis no planejamento da síntese de novas moléculas com propriedades otimizadas. Por exemplo, o anel tiofeno é destacado como um fragmento que contribui de forma positiva para a atividade biológica desta molécula. Por outro lado, o átomo de carbono 1 do fragmento 2-*terc*-butil-5-metilfenila contribui de forma negativa.

3. QSAR 3D: CoMFA

Um dos aspectos mais críticos da modelagem de QSAR 3D é o alinhamento molecular 3D das moléculas do conjunto de dados. O programa GOLD 3.0[®], utilizado neste trabalho para a geração do alinhamento 3D baseado na estrutura do receptor, oferece uma série de parâmetros que podem ser controlados pelo usuário. Dentre os parâmetros mais importantes se destacam: (i) a quantidade de soluções de docagem (acoplamentos) que deve ser produzida para cada ligante; e, (ii) a definição da área do sítio ativo para o acoplamento das diferentes moléculas. Essa área, que consiste em uma esfera, pode ser definida a partir de um ponto com coordenadas cartesianas definidas, ou de um átomo de um ligante ou ainda de um resíduo de aminoácido da proteína. Uma vez definido o sítio ativo, todas as moléculas (com todos os seus átomos constituintes) só poderão ser alocadas no espaço disponível dentro desta esfera. Outro parâmetro se refere às restrições que podem ser impostas. Estas restrições podem especificar interações que devem ocorrer entre a proteína e os ligantes, como por exemplo, ligações hidrogênio e interações hidrofóbicas. A Figura 8 apresenta um esquema geral das interações entre os inibidores do conjunto de dados e a HIV-1 PR. Estas interações foram consideradas no processo de docagem molecular.

Figura 8. Esquema geral das interações entre os inibidores do conjunto de dados e a HIV-1 PR.

Com base na análise estrutural das interações intermoleculares responsáveis pelo processo de reconhecimento molecular, algumas características foram estabelecidas no processo de docagem. O sítio ativo foi definido com um raio de 10 Å em torno do resíduo de aspartato 25. Em relação às restrições mais importantes, foi definido que: (i) o átomo de hidrogênio da hidroxila do anel hidroxipirona faria uma ligação hidrogênio com o átomo de oxigênio da carboxila do resíduo de aspartato 25; (ii) o átomo de oxigênio da carbonila do anel hidroxipirona faria uma ligação hidrogênio com o átomo de hidrogênio da amina do resíduo de isoleucina 50. Foi estabelecido ainda que para cada ligante deveriam ser produzidas 20 soluções.

Após realizado o processo de alinhamento, foram geradas 1640 soluções (conformações), sendo 20 para cada molécula do conjunto de

dados. Cada conformação foi avaliada quanto à satisfação da hipótese farmacofórica apresentada na Figura 8. Desta maneira, verificou-se se os grupos químicos dos inibidores foram alocados de forma correta nos subsítios da proteína. Também foi avaliado o grau de similaridade entre as conformações das diferentes moléculas e o seu acoplamento relativo dentro do sítio de ligação. Baseando-se nestes critérios, foi escolhida uma única conformação para cada molécula. A Figura 9 apresenta as 82 moléculas alinhadas para o processo de modelagem de QSAR 3D. Deve ser enfatizado que diversos outros alinhamentos também foram testados antes da escolha final deste alinhamento gerado com o programa GOLD 3.0[®]. Os outros alinhamentos descartados do processo de modelagem não serão discutidos.

Figura 9. Alinhamento molecular estrutural do conjunto de dados.

Os modelos CoMFA foram desenvolvidos com base na mesma combinação de conjuntos treinamento e teste utilizada na modelagem de HQSAR (item 2 deste capítulo da dissertação). Inicialmente, foram

especificados os seguintes parâmetros para a obtenção dos descritores moleculares 3D: triplos padrão (*standard*) como classe de campo; campos eletrostático e estereoquímico; distância como dielétrico; valores de corte eletrostático e estereoquímico de 30 Kcal/mol; grade com espaçamento de 2 Å entre os pontos e com 4 Å de distância entre as moléculas e os limites externos da grade; átomo de carbono sp^3 como sonda, com carga +1.

Uma vez gerados os descritores, foram especificados os seguintes parâmetros para a análise estatística com o método PLS para a obtenção do número ótimo de componentes para o modelo de QSAR: método de validação com “deixe um de fora” (LOO, do inglês, *Leave-One-Out*), para um número máximo de 6 componentes, empregando o escalonamento CoMFA padrão (*standard*).

O modelo foi gerado com 4 componentes e valores de $q^2 = 0,586$ e $SEP = 0,353$. A seguir, com o número ótimo de componentes foi realizada a análise sem validação cruzada, para a obtenção de um modelo com os seguintes parâmetros, $r^2 = 0,724$ e $SEE = 0,288$. Evidentemente, este modelo inicial foi considerado insatisfatório, com base nos parâmetros estatísticos. Desta maneira foi utilizado o recurso de focagem de região “QSAR CoMFA Region Focus”, na tentativa de melhorar os resultados. Este recurso consiste em atribuir um peso maior para aqueles pontos da grade que contribuem mais para o modelo. Várias formas são utilizadas para a realização desta operação. Uma delas é o coeficiente de desvio padrão (SDC, do inglês, *standard deviation coefficient*), que a partir de uma análise previamente realizada, atribui o peso que determinado ponto terá em uma nova análise, multiplicando a variação dos valores dos campos de interação deste ponto pelo coeficiente que foi atribuído a este ponto pela análise de

PLS anterior. O modelo CoMFA anterior foi, portanto, empregado na derivação de três outros modelos com valores de SDC diferentes. Os resultados deste processo são apresentados na Tabela 8.

Tabela 8. Modelos CoMFA obtidos em função da variação em SDC.

Modelo	q^2	SEP	r^2	SEE	N	F	Fração	
							S	E
1	0,640	0,349	0,961	0,110	5	103,573	0,473	0,527
2	0,818	0,248	0,968	0,104	6	293,770	0,514	0,486
3	0,732	0,301	0,877	0,203	6	69,810	0,477	0,523

q^2 : coeficiente de correlação com validação cruzada; SEP: erro padrão de predição; r^2 : coeficiente de correlação; SEE: erro padrão de estimativa; N: número ótimo de componentes; F: valor do teste F; S: contribuição estereoquímica; E: contribuição eletrostática.

No primeiro modelo (modelo 1) derivado foi utilizado o valor de SDC = 0,3, em conjunto com os outros parâmetros (validação com LOO e número máximo de componentes igual a 6) para a obtenção do número ótimo de componentes. Este modelo foi gerado com 5 componentes e valores de $q^2 = 0,640$ e SEP = 0,349. Na produção do modelo final sem validação cruzada com o número ótimo de componentes, foram obtidos valores de $r^2 = 0,961$ e SEE = 0,110. No modelo 2, de maneira similar, foi utilizado o valor de SDC = 0,6. Este modelo foi gerado com 6 componentes e com valores de $q^2 = 0,818$ (SEP = 0,248) e $r^2 = 0,968$ (SEE = 0,104). No modelo 3, com SDC = 0,9, foi gerado um modelo com 6 componentes e com valores de $q^2 = 0,732$ (SEP = 0,301) e $r^2 = 0,877$ (SEE = 0,203).

Como pode ser observado, a utilização do recurso *CoMFA Focus* com o método SDC resultou em uma melhora significativa nos parâmetros estatísticos. Este resultado pode ser justificado porque existem muitos pontos da grade que não contribuem de forma positiva para o modelo, sendo considerados pontos de redundância ou baixa variância. Como o escalonamento CoMFA padrão (*standard*) atribui a todos os pontos o mesmo peso, muitos destes pontos podem ser incluídos na análise, resultando em um modelo de baixa (ou menor) qualidade. O recurso SDC potencializa os pontos com maior contribuição, resultando na melhoria dos resultados. O modelo 2 (SDC = 0,6) foi o que apresentou os melhores resultados, com um valor significativo de $q^2 = 0,818$.

Além do SDC, foi utilizado outro recurso do *CoMFA Focus* para verificação da influência da variação do espaçamento entre os pontos da grade (GS, do inglês, *grid spacing*) na qualidade final dos modelos. O espaço padrão entre os pontos é de 2 Å, o que significa que as energias de interação são medidas somente a cada intervalo de 2 Å. Portanto, foram aplicadas variações do espaçamento que permitiram avaliar este parâmetro nos resultados CoMFA. Para tanto, foram utilizados os valores de GS de 1, 3 e 4 Å. Os resultados são apresentados na Tabela 9.

Tabela 9. Resultados CoMFA obtidos em função da variação em GS.

Modelo	q^2	SEP	r^2	SEE	N	F	Fração	
							S	E
4	0,633	0,341	0,944	0,133	6	210.380	0,469	0,531
5	0,523	0,348	0,674	0,313	2	81.691	0,489	0,511
6	0,570	0,360	0,648	0,325	2	72.764	0,493	0,507

No modelo 4 (seguindo a seqüência), com valor de GS de 1 Å, foram obtidos valores de $q^2 = 0,633$ (SEP = 0,341), e $r^2 = 0,944$ (SEE = 0,133), com 6 componentes. O modelo 5 (GS = 3 Å) produziu valores de $q^2 = 0,523$ (SEP = 0,348) e $r^2 = 0,674$ (SEE = 0,313), com 2 componentes. Por fim, o modelo 6 (GS = 4 Å) produziu valores de q^2 de 0,570 (SEP = 0,360) e $r^2 = 0,648$ (SEE = 0,325). Como pode ser observado, verifica-se que a variação nos valores de GS não resultou em melhora adicional dos resultados, quando comparados aos resultados obtidos apenas pela variação do valor de SDC.

Na análise final de todos os dados obtidos, o modelo 2 foi selecionado para ser testado quanto ao seu potencial em prever a atividade biológica das moléculas da série teste apresentadas na Tabela 4. Os resultados são apresentados na Tabela 10.

Tabela 10. Valores experimentais e preditos de pIC_{50} associados aos resíduos para as moléculas do conjunto teste no modelo CoMFA.

Molécula	pIC_{50} Experimental	pIC_{50} Predito	Resíduo
66	8,397	8,173	-0,224
67	8,275	8,324	0,049
68	8,657	8,508	-0,149
69	8,455	8,683	0,228
70	8,677	8,640	-0,037
71	8,443	8,082	-0,361
72	7,982	7,987	0,005
73	8,017	8,499	0,482
74	7,772	7,889	0,117
75	8,698	8,473	-0,225
76	7,677	7,376	-0,301
77	8,443	8,117	-0,326
78	8,657	8,516	-0,141
79	8,481	8,369	-0,112
80	9,552	9,211	-0,341
81	9,346	9,412	0,066
82	9,142	9,062	-0,080

A diferença entre os valores experimentais e preditos da Tabela 10 indica uma elevada consistência externa do modelo, com boa capacidade de predição. Dentre as 17 moléculas do conjunto teste, 9 apresentaram resíduos abaixo de 0,2 unidades logarítmicas, 7 ficaram abaixo de 0,4 e apenas 1 acima de 0,4. A Figura 10 apresenta o gráfico dos valores experimentais contra os valores preditos para todo o conjunto de dados, onde pode ser observada a boa concordância entre estes dados.

O processo de validação externa comprova que este modelo é capaz de prever a potência biológica de outras moléculas dentro desta classe química, representada pela diversidade contida dentro do espaço químico explorado.

Figura 10. Gráfico CoMFA dos valores experimentais e preditos para as 82 moléculas do conjunto de dados.

Os mapas de contorno 3D CoMFA permitem a visualização dos campos estereoquímicos e eletrostáticos, e são úteis no trabalho químico medicinal, com ênfase nos estudos de SAR e síntese planejada.

As representações dos campos 3D com cores verde e amarelo referem-se a características estereoquímicas, ao passo que os campos 3D com cores azul e vermelho se referem a características eletrostáticas. Poliedros verdes em torno de determinadas regiões da molécula, sugerem substituições com grupos mais volumosos, ao passo que poliedros amarelos sugerem grupos menos volumosos. A presença de poliedros azuis em determinada região da molécula sugere substituições por grupos com maior carga positiva, ao passo que poliedros vermelhos sugerem grupos mais eletronegativos. O mapa de contorno é mostrado, de forma ilustrativa para a molécula mais potente do conjunto de dados (inibidor **52**), na Figura 11.

Figura 11. Mapa de contorno 3D CoMFA para o inibidor **52**.

As informações mais relevantes obtidas pelo mapa de contorno sugerem, pela presença de poliedros verdes, a introdução de grupos mais volumosos próximos aos carbonos 3 e 4 do anel 2-*terc*-butil-5-metilfenila. Esta informação é coerente com os estudos de docagem realizados para a obtenção do alinhamento estrutural. O posicionamento das moléculas no sítio ativo da enzima permite que grupos sejam adicionados nestas posições

sugeridas pelo mapa de contorno, preenchendo desta forma o subsítio S3' do sítio ativo da enzima. A Figura 12 mostra o inibidor **52** posicionado no sítio ativo da enzima, associada ao mapa de contorno estereoquímico.

Figura 12. Mapa de contorno estereoquímico do inibidor **52** no sítio ativo da enzima HIV-1 PR.

Já a presença de um poliedro azul na região do grupo metila do anel sugere substituições por grupos mais eletropositivos. Esta informação deve ser interpretada com mais reserva, pois esta porção da molécula ocupa o subsítio S2' do sítio ativo da enzima, uma região rica em aminoácidos apolares. Porém, a presença de dois resíduos de aspartato pode, de certa

forma, dar suporte a esta sugestão. A Figura 13 mostra o mapa de contorno eletrostático do inibidor **52** no sítio ativo da enzima.

Figura 13. Mapa de contorno eletrostático do inibidor **52** no sítio ativo da enzima HIV-1 PR.

CAPÍTULO V

CONCLUSÕES

CONCLUSÕES

- ✓ Os modelos finais de HQSAR e CoMFA para o conjunto de 82 inibidores da HIV-1 PR apresentam, como indicado pelos bons coeficientes de correlação e capacidade preditiva, elevada consistência interna e externa.
- ✓ O uso do conjunto teste é importante no processo de validação externa dos modelos. O espaço químico-biológico foi delineado de maneira apropriada neste conjunto de dados.
- ✓ Os mapas de contribuição e de contorno podem ser úteis em estudos de SAR e no planejamento da síntese de novos inibidores mais potentes e seletivos contra a HIV-1 PR.
- ✓ A aplicação de estratégias integradas em química medicinal é importante no planejamento de fármacos. O uso de métodos baseados na estrutura do receptor em conjunto com o emprego de técnicas de QSAR 3D (CoMFA) e QSAR 2D (HQSAR), confirma mais uma vez que é válido associar estratégias procurando uma sinergia útil no planejamento de NCEs. Os modelos finais de QSAR devem ser úteis para guiar futuros trabalhos em química medicinal no planejamento de novos inibidores da HIV-1 PR.

CAPÍTULO VI

REFERÊNCIAS BIBLIOGRÁFICAS

VI. REFERÊNCIAS BIBLIOGRÁFICAS

1. Pharmaceutical Research and Manufacturers of America (Phrma). The Value of Medicines <<http://www.phrma.org>> (Phrma, Washington, 2006).
2. MEDTAP International, Inc., The Value of Investment in Health Care: Better Care, Better Lives, <www.aha.org> (MEDITAP, Bethesda 2003).
3. Merlot, C.; Domine, D.; Church, D. J. Fragment Analysis in Small Molecule Discovery. *Curr. Opin. Drug Discov. Devel.*, 5, 391-399, 2002.
4. Wermuth, C. G. The Practice of Medicinal Chemistry. Academic Press, London, 2003.
5. Lombardino, J. G.; Lowe III, J. A. The role of the Medicinal Chemist in Drug Discovery – Then and Now. *Nat. Rev. Drug. Discov.*, 3, 853-862, 2004.
6. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Cost. *J. Health Econom.*, 22, 151-185, 2003.
7. Cohen, F. J. Macro Trends in Pharmaceutical Innovation. *Nat. Rev. Drug Discov.*, 4, 78-84, 2005.
8. Kubinyi, H. 3D QSAR in Drug Design: Theory, Methods and Applications v. 1. Kluwer Academic Publishers/Escom, Leiden, 1993.
9. Thomas, G. Fundamentals of Medicinal Chemistry. Wiley, Chichester, 2003.

10. Williams, D. A.; Foye, W.O.; Lemke, T. L. Foye's Principles of Medicinal Chemistry. Lippincott Williams and Wilkins, Philadelphia, 2002.
11. Bunin, A. B. Increasing the Efficiency of Small Molecule Drug Discovery. *Drug. Discov. Today*, 18, 823-826, 2003.
12. Andricopulo, A. D.; Montanari, C. A. Structure-Activity Relationships for the Design of Small-Molecule Inhibitors. *Mini Rev. Med. Chem.*, 5, 585-593, 2005.
13. Joint United Nations Programme on HIV/AIDS (UNAIDS). AIDS epidemic update, December 2006 <<http://www.unaids.org>> (UNAIDS, Geneva, 2006).
14. Joint United Nations Programme on HIV/AIDS (UNAIDS). AIDS epidemic update, December 2005 <<http://www.unaids.org>> (UNAIDS, Geneva, 2005).
15. Roitt, I.; Brostoff, J.; Male, D. Imunologia. Manole, São Paulo, 1999.
16. Hutchinson, J. F. The Biology and Evolution of HIV. *Annu. Rev. Anthropol.*, 30, 85-108, 2001.
17. Robbins,. Patologia Estrutural e Funcional. Guanabara Koogan, Rio de Janeiro, 2000.
18. Tiemessen, C. T.; Hunt, G. M.; Papathanasopoulos M. A. Evolution and Diversity of HIV-1 in África – a Review. *Virus Genes*, 26, 151-163, 2003.
19. Knipe, D. M.; Howley P. M.; Griffin, D. E.; Lamb, R. A.; Martin, M. A. Fields Virology. Lippincott Williams and Willkins Publishers, Philadelphia, 2001.
20. Turner, B. G.; Summers, M. F. Structural Biology of HIV. *J. Mol. Biol.*, 285, 1-32, 1999.

21. Fredd E. O.; Mouland A. J. The Cell Biology of HIV-1 and Other Retroviruses. *Retrovirology*, 3, 2006.
22. Fauci, A. S. HIV and AIDS: 20 Years of Science. *Nat. Med.*, 9, 839-843, 2003.
23. National Institute of Allergy and Infectious diseases (NIAID). How HIV Causes AIDS <www.niaid.nih.gov> (NIAID, Bethesda, 2004).
24. Smith, S. M. The Pathogenesis of HIV infection: stupid may not be so dumb after all. *Retrovirology*, 3, 2006.
25. Marcello, A. Latency: the hidden HIV-1 challenge. *Retrovirology*, 3, 2006.
26. Department of Health and Human Services (DHHS). Guidelines for the Use of Antiretroviral Agents in HIV-1 Infected Adults and Adolescents. <www.aidsinfo.nih.gov> (DHHS, Rockville, 2006).
27. Department of Health and Human Services (DHHS). Approved Medications to Treat Complications of HIV/AIDS <www.aidsinfo.nih.gov> (DHHS, Rockville, 2004).
28. National Institute of Allergy and Infectious diseases (NIAID). Treatment of HIV Infection <www.niaid.nih.gov> (NIAID, Bethesda, 2006).
29. Vella, S.; Palmisano, L. Antiretroviral therapy: state of the HAART. *Antiviral Res.*, 45, 1-6, 2000.
30. Richman, D. D. HIV Chemotherapy. *Nature*, 410, 995-1001, 2001.
31. Food and Drug Administration (FDA). Drugs Used in the Treatment of HIV Infection. <www.fda.gov> (FDA, Rockville, 2006).
32. National Institute of Health (NIH). HIV/AIDS Fact Sheet. <www.nih.gov> (NIH, Rockville, 2006).

33. Copeland, R.A. Evaluation of Enzyme Inhibitors in Drug Discovery. Wiley, New Jersey, 2005.
34. Brik, A.; Wong, C. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.*, 1, 5-14, 2002.
35. Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 Protease: A Major Success of Structure-Assisted Drug Design. *Annu. Rev. Biophys. Biomol.*, 27, 249-84, 1998.
36. Tong L. Viral Proteases. *Chem. Rev.*, 102, 4609-4626, 2002.
37. Wlodawer, A.; Gustchina, A. Structural and Biochemical studies of Retroviral Proteases. *Biochim. Biophys. Acta*, 1477, 16-34, 2000.
38. Turner, B., Summers, M. Structural Biology of HIV. *J. Mol. Biol.*, 285, 1-32, 1999.
39. Turner, S. HIV Protease Inhibitors: The Next Generation. *Curr. Med. Chem.*, 1, 141-162, 2002.
40. De Clercq, E. New Anti-HIV Agents and Targets. *Med. Res. Rev.*, 6, 531-565, 2002.
41. Babine, R.; Bender, S. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem. Rev.*, 97, 1359-1472.
42. Vacca, J.; Condra, J. Clinically Effective HIV-1 Protease Inhibitors. *Drug Discov. Today*, 7, 261-272, 1997.
43. Pharmaceutical Research and Manufacturers of America (Phrma). HIV-AIDS. <www.innovation.org> (Phrma, Washington, 2005).
44. Kubinyi, H. QSAR: Hansch Analysis and Related Approaches. VCH, New York, 1993.
45. QSAR and Advanced CoMFA Manual. Tripos Inc., St. Louis, 2005.

46. Doddareddy, M.; Lee, Y. Hologram Quantitative Structure Activity Relationship Studies on 5-HT₆ Antagonists. *Bioorg. Med. Chem.*, **12**, 3815-3824, 2004.
47. Lowis, D. R. HQSAR: A New, Highly Predictive QSAR Technique. Tripos Technical Notes, **1**, 1997.
48. HQSARTM Manual, Tripos Inc., St. Louis, 2006.
49. Waterbeemd, H. Advanced Computer Assisted Techniques in Drug Discovery. VCH, New York, 1994.
50. Boyer, F.; Prasad, J. *et al.* 5,6-Dihydropyran-2-ones Possessing Various Sulfonyl Functionalities: Potent Nonpeptidic Inhibitors of HIV Protease. *J. Med. Chem.*, **43**, 843-848, 2000.
51. Hagen, S.; Domagala, J. 4-Hydroxy-5,6-dihydropyrones as Inhibitors of HIV Protease: The Effect of Heterocyclic Substituents at C-6 on Antiviral Potency and Pharmacokinetic Parameters. *J. Med. Chem.*, **44**, 2319-2332, 2001.
52. Specker, E.; Lilie, H. *et al.* An Old Target Revisited: Two Privileged Skeletons and an Unexpected Binding Mode for HIV Protease Inhibitors. *Angew. Chem. Int. Ed. Engl.*, **44**, 3140-3144, 2005.
53. Honório, K.; Andricopulo, A. Hologram Quantitative Structure-Activity Relationships for a Series of Farnesoid X Receptor Activators. *Bioorg. Med. Chem. Lett.*, **15**, 3119-3125, 2005.
54. Castilho, M.; Andricopulo, A. Two- and Three-Dimensional Quantitative Structure-Activity Relationships for a Series of Purine Nucleoside Phosphorylase Inhibitors. *Bioorg. Med. Chem.*, **14**, 516-527, 2006.