BIBLIOTECA Faculdade de Ciências Farmacêuticas Universidade de São Paulo

'Kechido em payour-

## UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS

Programa de Pós-Graduação em Fármaco e Medicamentos

Área de Insumos Farmacêuticos

PLANEJAMENTO RACIONAL DE TUBERCULOSTÁTICOS COM BASE NA ESTRUTURA DOS ÁCIDOS MICÓLICOS

DA PAREDE CELULAR DO Mycobacterium tuberculosis

Kerly Fernanda Mesquita Pasqualoto

Tese para a obtenção do grau de

DOUTOR

Orientadora:

Profa. Dra. Elizabeth Igne Ferreira

São Paulo

2003

175bd

# DEDALUS - Acervo - CQ



Ficha Catalográfica Elaborada pela Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP.

Pasqualoto, Kerly Fernanda Mesquita
P284p Planejamento racional de tubeculostáticos com base na estrutura dos ácidos micólicos da parede celular do Mycobacterium tuberculosis / Kerly Fernanda Mesquita
Pasqualoto. -- São Paulo, 2003.
122p.

Tese (doutorado) - Faculdade de Ciências Farmacêuticas da Universidade de São Paulo. Departamento de Farmácia Orientador: Ferreira, Elizabeth Igne

1. Fármaco : Planejamento : Química farmacêutica 2. Tuberculostáticos : Medicina I. T. II. Ferreira, Elizabeth Igne, orientador.

615.19 CDD

BIBLIOTECA Faculdade da Ciéncias Farmacéuticas Universidade de São Paulo

#### Kerly Fernanda Mesquita Pasqualoto

## PLANEJAMENTO RACIONAL DE TUBERCULOSTÁTICOS COM BASE NA ESTRUTURA DOS ÁCIDOS MICÓLICOS

### DA PAREDE CELULAR DO Mycobacterium tuberculosis

Comissão Julgadora

da

Teses para obtenção do grau de Doutor

Profa. Dra. Elizabeth Igne Ferreira

orientador/presidente

Prof. Dr. Carlos Maurício Rabello de Sant'Anna

Profa. Dra. Antônia Tavares do Amaral

Prof. Dr. Leoberto Costa Tavares

Profa. Dra. Carla Maria de Souza Menezes

São Paulo, 20 de março de 2003.

"NAVEGADORES ANTIGOS tinham uma frase gloriosa: 'Navegar é preciso; viver não é preciso.'

Quero para mim o espírito desta frase, transformada a forma para a casar com o que eu sou: Viver não é necessário; o que é necessário é criar. Não conto gozar a minha vida; nem em gozá-la penso. Só quero tomá-la grande, ainda que para isso tenha de ser o meu corpo e a (minha alma) a lenha desse fogo.

Só quero torná-la de toda a humanidade; ainda que para isso tenha de a perder como minha.

Cada vez mais assim penso. Cada vez mais porho na essência anímica do meu sangue o propósito impessoal de engrandecer a pátria e contribuir para a evolução da humanidade.

É a forma que em mim tomou o misticismo de nossa Raça."

Esta nota solta e não assinada foi publicada, pela primeira vez, na primeira edição do volume FERNANDO PESSOA – OBRA POÉTICA, 1960, organizada por Maria Aliete Galhoz.

Dedico este trabalho à minha alma gêmea, EVANDRO, aos meus pais CLEUSA e FÁBIUS, aos meus irmãos KAREN, EMERSON e MADALENA e aos meus avós, in memorian, LAURINDA e ALFREDO

#### AGRADECIMENTOS

À Profa. Dra. Elizabeth Igne Ferreira, orientadora e amiga, pela oportunidade de desenvolver este estudo, pelo apoio de todas as horas, pelo estímulo e exemplo profissional e humano.

Ao Prof. Dr. Tony Hopfinger pela oportunidade de trabalhar no Laboratory of Molecular Modeling and Design - University of Illinois at Chicago (UIC), pela orientação na aplicação das técnicas de modelagem molecular utilizadas neste estudo, pelo apoio emocional e estímulo científico.

Ao Prof. Dr. Aparício Brittes Funck, meu orientador do mestrado e grande amigo, por me introduzir no mundo científico e ampliar minha visão de ensino.

À Profa. Dra. Antonia Tavares do Amaral (IQ-USP) e seu grupo de pesquisa pela disponibilidade em compartilhar informações referentes ao programa de modelagem molecular Sybyl<sup>®</sup> - Tripos, Inc, utilizado na fase inicial deste trabalho.

À Profa. Dra. Maria Amélia Barata da Silveira e à Profa. Dra. Veni Maria Felli pelo apoio ao longo do curso de pós-graduação.

Aos colegas do Laboratório de Planejamento e Síntese de Quimioterápicos da Profa. Elizabeth Igne Ferreira pelo bom convívio e carinho.

Aos colegas Xuan, Manisha, Craig, Dahua, Jianzhong, Xiaguang, Jane, Yi-Li (UIC/EUA) pela convivência agradável e carinho.

Ao meu colega e amigo Dr. Osvaldo Santos Filho (UIC/EUA) pelos ensinamentos sobre as técnicas de modelagem molecular aplicadas neste estudo, pela força e estímulo científico.

À Dra. Nelilma Romeiro (UFRJ) pela amizade, disponibilidade e incentivo.

Aos meus amigos Carla, Diogo, Kátia "Bee", Michelle, Michelle Ortiz e Roberto pelo carinho de todas as horas e por tornarem este período da minha vida mais divertido!

Aos meus amigos '*chicagoans*' Neli, Christopher, Cirlei, Joseph, Ariani e Ezecquiet pela compreensão e carinho.

Às secretárias Elisabete Claro de Souza Nascimento (FCF/USP) e Patricia Woods (UIC/EUA) pela disponibilidade e empenho de colaboração.

À CAPES e ao Departamento de Química Medicinal e Farmacognosia da Faculdade de Farmácia da UIC pelo auxílio financeiro.

A todos aqueles que, direta ou indiretamente, contribuíram para a realização deste estudo.

Ao Evandro, pelo amor, amizade e companheirismo de sempre!

Aos meus 'peludos' Spencer e Crystal pela companhia constante e prazerosa na fase de redação deste estudo.

À Milene, minha grande amiga, pela compreensão, carinho, estímulo e muitas risadas!

À minha família, por tudo! Em especial, à minha vozinha Laurinda (in memorian) por toda a orientação e incentivo, pelo amor, dedicação, força e exemplo de vida!

vii

## SUMÁRIO

LISTA D	E FIGURASix
LISTA D	E TABELASxl
LISTA D	E QUADROSxiv
LISTA D	E ABREVIATURAS, SIGLAS E SÍMBOLOSxvii
RESUM	0xx
ABSTRA	\CTxxii
1. INT	RODUÇĂO2
1.1. TVE	BERCULOSE
1.2. PLA	ANEJAMENTO RACIONAL DE NOVAS MOLÉCULAS E MODELAGEM MOLECULAR7
1.3. Ác	CIDOS MICÓLICOS E ENOIL-ACP REDUTASE (InhA)9
2. JUS	TIFICATIVA E OBJETIVO23
3. ME	TODOLOGIA
3.1. DA	DOS BIOLÓGICOS25
3.1.1.	DEFINIÇÃO DOS CONJUNTOS DE TREINAMENTO (TRAINING SET) E DE AVALIAÇÃO
	(TEST SET)
3.2. CC	ONSTRUÇÃO E MINIMIZAÇÃO DAS ESTRUTURAS
3.3. ME	TODOLOGIA DE QSAR-4D (IR, Independente do Receptor)
3.3.1.	AMOSTRAGEM CONFORMACIONAL POR DINÂMICA MOLECULAR
3.3.2.	DEFINIÇÃO DOS ELEMENTOS DE INTERAÇÃO FARMACOFÓRICA
3.3.3.	DEFINIÇÃO DO TAMANHO DAS CÉLULAS
3.3.4.	DEFINIÇÃO DOS ALINHAMENTOS TESTADOS
3.3.5.	DEFINIÇÃO DOS TIPOS DE OCUPAÇÃO
3.3.6.	OBTENÇÃO DOS DESCRITORES DE OCUPAÇÃO DAS CÉLULAS (GCODS)
3.3.7.	PRÉ-TRATAMENTO DOS DESCRITORES (GCODS) – REDUÇÃO DO NÚMERO TOTAL DE
	GCODs

VIII

3.3.8. OBTENÇÃO DOS MODELOS DE QSAR-4D (IR)41
3.3.9. IDENTIFICAÇÃO DOS MELHORES MODELOS (EQUAÇÕES) DE QSAR-3D43
3.3.10. CONJUNTO DE AVALIAÇÃO - VALIDAÇÃO EXTERNA E PREDIÇÃO DOS
MODELOS45
3.3.11. SELEÇÃO E VISUALIZAÇÃO DA CONFORMAÇÃO 'BIOATIVA'
3.4. METODOLOGIA DE FEFF ("FREE ENERGY FORCE FIELD", campo de força de energio
livre) QSAR-3D (DR, Dependente do Receptor)49
3.4.1. OBTENÇÃO E OTIMIZAÇÃO DO MODELO DA ENZIMA (MODELO DO
RECEPTOR)
3.4.2. DINÂMICA MOLECULAR DO MODELO REDUZIDO DA ENZIMA
3.4.3. DOCKING DO LIGANTE MAIS VOLUMOSO
3.4.4. DINÂMICA MOLECULAR DO COMPLEXO ENZIMA-LIGANTE MAIS VOLUMOSO53
3.4.5. DOCKING DOS LIGANTES DO CONJUNTO DE TREINAMENTO E DM DOS
COMPLEXOS ENZIMA-LIGANTE
3.4.6. DINÂMICA MOLECULAR DOS LIGANTES E DA ENZIMA NÃO-LIGADOS
3.4.7. CONSTRUÇÃO DOS MODELOS DE FEFF QSAR-3D
3.4.8. IDENTIFICAÇÃO DOS MELHORES MODELOS DE FEFF QSAR-3D
3.4.9. CONJUNTO DE AVALIAÇÃO - VALIDAÇÃO E PREDIÇÃO DO MODELO
4. RESULTADOS E DISCUSSÃO
4.1. ANÁLISE DE QSAR-4D
4.2. ANÁLISE DE FEFF QSAR-3D
5. CONCLUSÕES
6. PERSPECTIVAS110
7. REFERÊNCIAS BIBLIOGRÁFICAS

#### LISTA DE FIGURAS

Face la religio Méncias Farmacêuticas

X

Figura 8 – Ligante INHd18 da análide d	e QSAR-4D, com seus respectivos GCODs,
ancorado no sítio ativo da Inh.	A ( <b>1zid</b> )79

Figura	9	-	Ligante	INHd49	da	análise	de	QSAR-4	D, com	seus	respectivos	GCODs,
		(	ancorac	lo no sític	o at	ivo da Ir	hA	(1zid)				

## LISTA DE TABELAS

Tabela I – Definição da seqüência de três átomos, designados pelas letras "a-I",
para cada um dos sete alinhamentos testados, utilizando a estrutura do
composto INH1 como exemplo37
Tabela II – Seqüência de etapas de uma análise de QSAR-4D47
Tabela III - Definição dos termos do campo de força de energia livre (FEFF) utilizados
na construção dos modelos de FEFF QSAR-3D58
Tabela IV – Medidas estatísticas e número de descritores encontrados para os dez
melhores modelos de cada um dos alinhamentos testados66
Tabela V – Medidas estatísticas, número de descritores e número de outliers para os
dez melhores modelos do alinhamento 467
Tabela VI Matriz de correlação linear dos resíduos de ajuste para os dez melhores
modelos do alinhamento 468
Tabela VII - Matriz de correlação-cruzada linear dos GCODs para o modelo 1 do
alinhamento 469
Tabela VIII – Índices de coordenadas dos descritores para cada um dos compostos
do conjunto de avaliação (Test_Set)70

Tabela XIV – Distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os descritores do ligante INH1 (mais ativo)......80

Tabela XV – Distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os descritores do ligante INHd18 (atividade intermediária)......80

Tabela XVII – Volume das conformações 'ativas' previstas dos compostos do conjunto de treinamento, resultantes da análise de QSAR-4D IR......90

Tabela	XIX	-	Matriz	de	correlação	linear	dos	resíduos	de	ajuste	entre	pares	de
			model	os d	a análise de	FEFF G	SAR	3D					102

Tabela	XX –	Predição	de	atividade	biológica	para	OS	compostos	do	conjunto	de
		avaliação	o (Te	est_Set)					•••••		106

xiv

## LISTA DE QUADROS

Quadro 2 - Variação no tamanho dos ácidos micólicos (BARRY et al., 1998).....12

Quadro 5 - Valores de MIC e pMIC das trinta e sete hidrazidas selecionadas......25

.

Quadro 17 – Valores dos termos de energia resultantes das simulações de DM de 50
ps a 310 K das conformações de energia mínima dos ligantes do
conjunto de avaliação104

Quadro 18 – Propriedades d	los ligantes do conjunto d	de avaliação no	estado não-
ligado			105

BIBLIOTECA Faculdade de Ciências Farmacêuticas Universidade de São Paulo

•

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

- TB = tuberculose
- OMS = Organização Mundial da Saúde
- MDR = multidrug-resistance (multi-resistência)
- INH = isoniazida
- RIF = rifampicina
- PZA = pirazinamida
- SM = estreptomicina
- EMB = etambutol
- DOTS = Directly Observed Treatment, Short-Course (tratamento de curta duração com observação direta)
- CADD = Computer-Assisted Drug Design (planejamento de fármacos auxiliado por computador)
- QSAR = Quantitative Structure-Activity Relationship (relação quantitativa entre estrutura química e atividade biológica)
- GC = bases guanina-citosina
- CMN = Corynebacterium-Mycobacterium-Nocardia
- FAS = Fatty Acid Synthase (ácido graxo sintetase)
- Acp = acyl carrier protein (proteína acil-carregadora)
- InhA = enoil-acp redutase de M. tuberculosis
- NAD = nicotinamida adenina dinucleotídeo
- Fabl = enoil-acp redutase de Escherichia coli
- FEFF = Free Energy Force Field (campo de força de energia livre)
- IR = independente do receptor
- DR = dependente do receptor
- MIC = Minimum inhibitory concentration (concentração inibitória mínima, CIM)

XVIII

- T = temperatura
- K= escala Kelvin de temperatura
- PM = peso molecular
- DM = dinâmica molecular
- CEP = Conformational Ensemble Profile (perfil de amostragem conformacional)
- 3D = tridimensional ou terceira dimensão
- 4D = quadridimensional ou quarta dimensão
- IPEs = Interaction Pharmacophore Elements (elementos de interação farmacofórica)
- Ao = absolute-occupancy (ocupação absoluta)
- Jo = joint-occupancy (ocupação associada)
- So = self-occupancy (auto-ocupação)
- GCODs = Grid Cell Occupancy Descriptors (descritores de ocupação da célula da grade ou caixa 3D virtual)
- CoMFA = Comparative Molecular Field Analysis (análise comparativa de campos moleculares)
- PLS = Partial Least Squares (regressão de quadrados mínimos parciais)
- GA = Genetic Algorithm (algoritmo genético)
- GFA = Genetic Function Approximation (aproximação da função genética)
- LOF = Lack-of-fit (fator de desajuste de Friedman)
- LSE = Least Square Error (erro de mínimos quadrados)
- DP = desvio padrão
- ClogP = valor calculado do coeficiente de partição *n*-octanol/água, por Ghose, Pritchett e Crippen (1998) (Hyperchem – Hybercube, Inc.)
- CAMD = Computer-Assisted Molecular Design (planejamento molecular auxiliado por computador)
- BA = Biological Activity (atividade biológica, AB)

xix

- HOMO = Highest Occupied Molecular Orbital (energia do orbital molecular ocupado de maior energia)
- LUMO = Lowest Unoccupied Molecular Orbital (energia do orbital molecular desocupado de menor energia)
- L = ligante
- R = receptor
- LR = complexo ligante-receptor
- M = meio aquoso de solvatação
- r.m.s. = root mean square deviation (desvio da raiz dos mínimos quadrados)
- u.m.a. = unidade de massa atômica
- ps = picossegundo
- fs = fentossegundo
- $\dot{A} = Angstrom$
- r<sup>2</sup> = coeficiente de correlação linear
- q<sup>2</sup> = coeficiente de correlação cruzada
- ΔE = variação de energia

xx

#### RESUMO

Os ácidos micólicos, principais componentes da parede celular do M. tuberculosis, são alvos específicos para o planejamento de novos agentes potenciais contra a tuberculose (TB) e o estudo de seu processo biossintético é de importância fundamental no desenvolvimento de inibidores mais seletivos. À vista de tal fato, desenvolveu-se estudo utilizando técnicas de modelagem molecular (CADD) com o objetivo de contribuir para o planejamento racional de protótipos de atividade tuberculostática potencial, com base em análogos estruturais da isoniazida (INH), cujo mecanismo de ação está ligado à inibição da síntese de ácidos micólicos. Neste estudo, 37 hidrazidas, divididas em conjunto de treinamento (30 hidrazidas) e conjunto de avaliação (7 hidrazidas), submetidas ao mesmo protocolo farmacológico e com mecanismo de ação possivelmente comum ao do fármaco INH, foram investigadas por meio das técnicas de modelagem molecular (CADD): QSAR-4D IR (relação quantitativa quadridimensional entre estrutura química e atividade biológica independente do receptor) e FEFF QSAR-3D (relação quantitativa tridimensional entre estrutura química e atividade biológica com campo de força de energia livre). Na análise de QSAR-4D IR, 5.000 conformações de cada ligante foram geradas para se obter perfis de amostragem conformacional (CEP) provenientes de simulações de dinâmica molecular (DM) de 100.000 estados de trajetória registrados a cada 20 passos. Foram testados sete alinhamentos e a freqüência de ocupação de cada célula da grade (1Å) foi computada para cada um dos sete tipos de elementos de interação farmacofórica (IPEs) para cada ligante. Estes descritores de ocupação das células da grade (GCODs) foram utilizados como variáveis independentes na construção dos modelos de QSAR-3D após o processo de redução de dados. Os modelos de QSAR-3D foram gerados e avaliados por meio de um processo de otimização que

RESUMO

combina algoritmo genético (GA) com regressão de quadrados mínimos parciais (PLS). Na análise de FEFF QSAR-3D, utilizou-se a estrutura cristalizada da enzima enoil-acp redutase do M. tuberculosis, InhA (1zid, código de acesso no PDB), com 2,7 Å de resolução, como biomacromolécula alvo do conjunto de hidrazidas investigado. As variáveis dependentes corresponderam aos valores de concentração inibitória mínima (- log MIC = pMIC) contra cepas de M. tuberculosis var. bovis à temperatura de 310 K (37 °C). As variáveis independentes dos modelos de QSAR-3D (descritores) consistem em termos de energia classificados utilizando campo de força AMBER combinado a um modelo de solvatação "hydration shell". Outros descritores como energia de HOMO e de LUMO, ClogP, momento de dipolo, volume e área superficial foram computados para o estado não-ligado dos ligantes estudados. Simulações de DM em temperaturas variadas ("warming up") e o método de aproximação da função genética (GFA) foram empregados utilizando regressão de quadrados mínimos parciais (PLS) e regressão linear multidimensional como funções de ajuste para desenvolver os modelos de FEFF QSAR-3D. De acordo com a análise de QSAR-4D IR, pôde-se constatar que a presença de grupos substituintes apolares na porção acil dos ligantes diminui consideravelmente a atividade biológica. Substituintes nesta porção impediriam a interação do tipo  $\pi$ stacking com a cadeia lateral do resíduo Phe149 do sítio de interação. Todavia, modificações na porção NAD dos ligantes (grupos apolares, doadores e/ou aceptores de ligação de hidrogênio) permitiriam melhor interação com resíduos de aminoácido Tyr158, Met103, Pro193, lle194 e Thr196, Leu197, Thr17, Ser20, contribuindo à atividade biológica. Conforme a análise de FEFF QSAR-3D, o modelo selecionado apresentou como descritores que contribuem à atividade biológica a área superficial e o ClogP dos ligantes no estado não-ligado. Contudo, neste estudo, a análise de FEFF QSAR-3D foi limitada devido à natureza do sistema estudado e dos dados de atividade biológica semi-quantitativos.

xxii

#### ABSTRACT

Mycolic acids, the major components of the cell wall of M. tuberculosis, are attractive targets for the rational design of new antituberculosis agents. The knowledge of their biosynthetic pathways is really important for developing inhibitors more potent to treat tuberculosis (TB). With the purpose of contributing for rational drug design of tuberculostatic leads, a mechanism-based study using molecular modeling (CADD) was performed with analogs of isoniazid (INH), drug that acts on mycolic acid biosynthesis. In this study, a set of 37 hydrazides (a training set of 30 hydrazides, and a test set of 7 hydrazides), which were evaluated with the same biological essay, and, probably, would act like the lead INH, were investigated using the molecular modeling formalisms (CADD): RI 4D-QSAR (receptor-independent fourdimensional quantitative structure-activity relationship) and RD FEFF 3D-QSAR (receptor-dependent free energy force field three-dimensional quantitative structure-activity relationship). Considering the 4D-QSAR analysis, 5.000 conformations of each molecule of training set were sampled to generate a conformational ensemble profile (CEP), from a molecular dynamic simulations (MDS) of 100,000 steps trajectory states, recorded every 20 steps. Seven trial alignments were tested, and the frequency of occupation of each grid cell space (1Å) was computed for each of seven types of interaction pharmacophore elements (IPEs) of each ligand. These grid cell occupancy descriptors (GCODs) were used as independent variables in constructing (3D)-QSAR models after data reduction. The 3D-QSAR models were generated and evaluated by a scheme that combines a genetic algorithm (GA) optimization with partial least squares (PLS) regression, FEFF QSAR-3D analysis was used to construct ligand-receptor binding models for a set of hydrazides investigated, and the molecular target used was a crystallized structure of the encyl-acp reductase enzyme from M. tuberculosis, InhA, (1zid, entry PDB

xxiii ABSTRACT

code), at 2.7 Å resolution. The dependent variable of the 3D-QSAR models is the minimum inhibitory concentration (-log MIC = pMIC) against strains of M. tuberculosis var. bovis at 310 K (37 °C) temperature. The independent variables of the 3D-QSAR models (the descriptors) are scaled energy terms of a modified first-generation AMBER force field combined with a hydration shell aqueous solvation model. Other descriptors as the HOMO, LUMO, ClogP, dipole, volume, and superficial area were computed for the unbound state of the ligands studied. Multiple temperature molecular dynamics simulation (MDS) (warming up technique) and the genetic function approximation (GFA) were employed using partial least square (PLS) and multidimensional linear regression as the fitting functions to develop FEFF 3D-QSAR models for the binding process. According to RI 4D-QSAR analysis, the presence of nonpolar groups in the acyl moiety of the ligands decreased the biological activity considerably. Groups of atoms in this position would prevent an  $\pi$ -stacking interaction with the side chain of Phe149 of the active site. However, molecular modifications in the NAD moiety of the ligands (nonpolar groups, hydrogen bond donor and/or acceptor groups) would improve the interaction with the amino acid residues Tyr158, Met103, Pro193, Ile194, and Thr196, Leu197, Thr17, Ser20, increasing the biological activity. Considering FEFF 3D-QSAR analysis, the model selected presented the superficial area and ClogP of the unbound ligands as descriptors which improved the biological activity. Nevertheless, in this study, FEFF 3D-QSAR was limited by the nature of the system studied and the biological activity data, which were semi-quantitative data.

.

.

.

INTRODUÇÃO

## 1. INTRODUÇÃO

## 1.1. TUBERCULOSE

A tuberculose (TB) é doença infecciosa crônica causada por micobactérias do "complexo tuberculose", incluindo Mycobacterium bovis, Mycobacterium africanum e principalmente o Mycobacterium tuberculosis. A TB é exemplo clássico de doença causada por parasita intracelular. Na maioria dos casos, a infecção é transmitida de pessoa para pessoa por meio da inalação de gotículas infectadas. A TB pode envolver qualquer órgão sistêmico, mas o pulmão é o órgão mais comum da lesão primária e o principal comprometido. A estrutura e a evolução das lesões causadas pelo bacilo são determinadas por sistema de defesa específico do hospedeiro, pela resposta imunológica e por fatores genéticos (WOLINSKY, 1992; SESSI, GRASSI, 1996). De acordo com a definição da Sociedade Torácica Americana, estar infectado não significa estar doente. O estado de doença é definido pelo aparecimento de sinais e sintomas da infecção comprovados por meio de exames clínicos, radiológicos e bacteriológicos (SESSI, GRASSI, 1996).

Em 1993, a TB foi declarada pela Organização Mundial da Saúde (OMS) como emergência global de saúde pública, devido as suas proporções. Cerca de um terço da população mundial está infectada pelo *M. tuberculosis*. Cada pessoa doente pode infectar de 10 a 15 pessoas somente em um ano. Em 1996, houve aproximadamente oito milhões de novos casos da doença, com três milhões de mortes (WHO 1998, 2002a). Noventa e cinco porcento dos casos de TB e noventa e oito porcento das mortes ocorrem nos países em desenvolvimento. Setenta e cinco porcento dos indivíduos com a doença encontram-se na faixa etária de 15 a 49 anos, representando homens e mulheres no período mais produtivo do desenvolvimento (WHO 1998, 2002b, 2002c).

A incidência regional e global de TB está próxima a dois milhões de casos/ano na África, cerca de três milhões de casos/ano no Sudeste da Ásia, mais de um quarto de milhão na Europa Oriental e cerca de 161.800 novos casos/ano no Brasil (WHO, 1997, 2002a). Foi estimado que entre 2002 e 2020, cerca de um bilhão a mais de pessoas estarão infectadas, cento e cinqüenta milhões estarão doentes e trinta e seis milhões morrerão de TB, se o controle não for fortalecido (WHO, 2002a).

A AIDS agravou significativamente a situação mundial da TB. Um terço do aumento na incidência dessa doença, nos últimos cinco anos, pode ser atribuído ao HIV. Este vírus enfraquece o sistema imunológico e a probabilidade de se tornar doente aumenta cerca de trinta vezes para um indivíduo HIV-positivo e infectado com o bacilo da TB (WHO, 2002a). O diagnóstico é difícil porque os sinais e sintomas e o aspecto radiográfico característicos freqüentemente estão ausentes (WOLINSKY, 1992). Na África, o HIV tem sido fator determinante para o aumento na incidência de casos de TB nos últimos dez anos (WHO, 2002a). No Brasil, acredita-se que duzentos mil indivíduos estejam co-infectados (TB e HIV). Esta combinação letal é mais marcante no Rio de Janeiro e em São Paulo, cidades com índices elevados tanto de TB quanto de HIV/AIDS (WHO, 1997, 2002d).

Outro fator que contribui para aumentar a incidência de TB e é responsável pelo aumento da mortalidade consiste no aparecimento de novas cepas do M. *tuberculosis* resistentes aos fármacos anti-TB convencionalmente utilizados. Estas cepas podem apresentar resistência a um fármaco isolado ou à combinação de fármacos. Neste caso, são denominadas de multi-resistentes (MDR – "multidrug-resistant"). De acordo com dados da OMS, mais de cinqüenta milhões de pessoas podem estar infectadas por cepas resistentes (WHO, 2002a). A resistência pode ser causada, principalmente, pelo tratamento parcial ou inconsistente, pela prescrição de fármacos inadequados (fármaco isolado ou em combinação) e por um sistema falho de distribuição de medicamentos (WHO, 2002a; RAITAN, KALIA, AHMAD,

3

1998). A resistência também pode ser conseqüência natural à infecção por HIV (RATTAN, KALIA, AHMAD, 1998). Os casos de resistência são de difícil tratamento, além de economicamente dispendiosos. Nos países industrializados, o tratamento da TB custa cerca de dois mil dólares por paciente e aumenta mais de 100 vezes em pacientes com cepas MDR-TB (duzentos e cinqüenta mil dólares por paciente) (WHO, 2002a).

A quimioterapia de curta duração constitui o fundamento da terapia anti-TB. Esta estratégia de tratamento é recomendada pela OMS, por meio dos regimes de DOTS (Directly Observed Treatment, Short-course – tratamento de curta duração com observação direta), para garantir o controle efetivo da TB. Isoniazida (hidrazida do ácido isonocotínico) (INH), rifampicina (RIF), pirazinamida (PZA), estreptomicina (SM) e etambutol (EMB) são os fármacos constituintes da quimioterapia anti-TB de curta duração. Estes fármacos são administrados por 6 a 8 meses, de acordo com os esquemas terapêuticos preconizados pela OMS (WHO, 1999). A estreptomicina foi introduzida na quimioterapia anti-TB na década de 40, a isoniazida e a pirazinamida na década de 50, o etambutol na década de 60 e a rifampicina na década de 70 (SESSI, GRASSI, 1996; KOROLKOVAS, 1998) (Quadro 1).



Quadro 1 - Estrutura química dos fármacos utilizados na estratégia de DOTS da OMS





SM

KERLY FERNANDA MESQUITA PASQUALOTO

A estratégia de DOTS impede novas infecções e o desenvolvimento de MDR-TB, combinando cinco elementos: comprometimento político, servicos microscópicos, distribuição de medicamentos, monitoração do sistema e observação direta do tratamento. De acordo com dados da OMS (WHO, 1999, 2002a), o suprimento de medicamentos para 6 meses, no regime de DOTS, tem um custo de 11 dólares por paciente, constituindo-se em investimento extremamente viável. No entanto, somente doze porcento dos pacientes estimados receberam DOTS em 1996; no início de 1997, 95 dos 212 países adotaram a estratégia de DOTS e 63 dos 95 países implementaram DOTS em todas as regiões. Em 2000, 148 países adotaram a estratégia de DOTS e vinte e sete porcento dos casos de TB no mundo foram tratados por meio desta estratégia (WHO, 2002d). No Brasil, a estratégia de DOTS foi concentrada em 250 municípios, selecionados pela alta incidência de TB, alto índice de mortalidade, baixo índice de cura e disseminação da AIDS. O desafio está em expandir esta estratégia de tratamento tão rápido quanto possível, sem arriscar a perda de qualidade (WHO, 1997). O plano traçado para o controle da TB no Brasil, no período de 2001 a 2005, envolve a mobilização política e social e a expansão da estratégia de DOTS para 329 municípios, cujos índices de TB representam oitenta porcento dos casos notificados (WHO, 2002d).

Embora os esquemas terapêuticos preconizados para a TB sejam amplamente aceitos, grandes esforços devem ser realizados para adaptá-los às diferentes situações sócio-econômicas e às várias apresentações da doença. É necessário, também, pesquisar por novos agentes antimicobacterianos, considerando-se que os organismos infectantes tornam-se resistentes aos fármacos disponíveis, alguns deles apresentam efeitos colaterais graves e o tratamento convencional das micobacterioses em pacientes imunodeprimidos é insatisfatório (SESSI, GRASSI, 1996). As estratégias baseadas em otimizar a inibição de alvos conhecidos requerem entendimento detalhado do mecanismo de ação dos agentes antimicobacterianos convencionais, enquanto que as estratégias baseadas em identificar novos alvos necessitam do esclarecimento acerca dos mecanismos bioquímicos específicos da micobactéria e organismos relacionados. Muitos processos metabólicos específicos da micobactéria ocorrem durante a biossíntese dos componentes da parede celular e novos alvos interessantes têm emergido para o planejamento racional de novos agentes antimicobacterianos (BARRY, 1997). Os ácidos micólicos, componentes principais da parede celular do *M. tuberculosis*, são exemplos destes alvos e o estudo de sua estrutura química associado ao de seu processo biossintético constitui ferramenta importante para o planejamento de novos agentes potenciais contra tuberculose (PASQUALOTO, FERREIRA, 2001).

#### 1.2. PLANEJAMENTO RACIONAL DE NOVAS MOLÉCULAS E MODELAGEM MOLECULAR

Apesar dos avanços tecnológicos, a informação principal que permite o planejamento racional de novas moléculas é o conhecimento da etiologia de determinada doença, que se deseja combater, ou, pelo menos, dos processos bioquímicos que são alterados (WERMUTH, 1996; COHEN, 1996).

De acordo com os glossários de termos utilizados em Química Medicinal (WERMUTH et al., 1998) e no planejamento de fármacos (SANT'ANNA, 2002), recomendações da IUPAC para 1997, modelagem molecular é a investigação das estruturas e das propriedades moleculares usando a química computacional e as técnicas de visualização gráfica visando fornecer uma representação tridimensional, sob um dado conjunto de circunstâncias.

O CADD (Computer-Assisted Drug Design, planejamento de fármacos auxiliado por computador) pode ser aplicado a qualquer molécula ativa cuja interação com um alvo molecular específico (receptor, enzima, canais de íons) seja postulada. A modelagem molecular utiliza duas abordagens: planejamento direto e indireto. Quando a estrutura tridimensional da biomacromolécula alvo (receptor, enzima) está disponível, o planejamento direto pode ser efetuado. Por outro lado, quando a estrutura tridimensional da biomacromolécula alvo não está disponível, o planejamento indireto é a única abordagem possível (WERMUTH, 1996; COHEN, 1996).

A contribuição do CADD na descoberta de novos protótipos possui valor limitado devido ao fato de que os estudos de modelagem molecular permitiriam somente a identificação de novos ligantes provenientes de sistemas biológicos já conhecidos. Do mesmo modo, métodos de relação quantitativa entre estrutura química e atividade biológica (QSAR, Quantitative Structure-Activity Relationship) possibilitariam a otimização somente de conjuntos de compostos já descobertos (WERMUTH, 1996). O CADD não é, entretanto, uma rota direta para a descoberta de novos fármacos, mas fornece informações mais detalhadas para alcançar este objetivo (OOMS, 2000).

A utilização da modelagem molecular como ferramenta para o planejamento de fármacos apresenta aspectos positivos e até mesmo certo poder criativo. Entre os aspectos positivos encontram-se (WERMUTH, 1996):

 (1) redução do número de compostos a serem sintetizados, através do reconhecimento de parâmetros estruturais e eletrônicos que resultam em compostos pouco ativos ou inativos, por exemplo;

(2) diferenciação de estereoisômeros – a estereoespecificidade é um dos principais atributos de receptores farmacológicos e uma perfeita complementaridade estereoquímica entre ligante e sítio ativo é critério essencial para alta afinidade e seletividade; (3) possibilidade de distinguir entre agonistas e antagonistas – isto é relativamente fácil para a categoria específica de antagonistas que, de acordo com a teoria de Ariëns (1954), derivam de agonistas simplesmente por meio da adição de grupos volumosos, como anéis aromáticos, que funcionariam como pontos adicionais na interação fármaco-receptor (passagem de agonistas muscarínicos para antagonistas muscarínicos, por exemplo). No entanto, a diferenciação entre as duas categorias torna-se menos evidente quando a transformação de agonista para antagonista reside em mudanças relativamente sutis, como as observadas nos antagonistas benzodiazepínicos, por exemplo;

(4) possibilidade de explicar observações aparentemente paradoxais, como por exemplo, a inversão inesperada de afinidade encontrada em enantiômeros R e
\$ de um conjunto de análogos da sulpirida (antagonista seletivo de dopamina D<sub>2</sub>)
quando se alteram de N-alquil(etil) para N-benzil derivados (ROGNAN et al., 1990);

(5) possibilidade de demonstrar algum poder de predição e permitir o planejamento de compostos novos e mais potentes, ou melhor, de estruturas químicas totalmente novas.

Portanto, a modelagem molecular tem auxiliado de forma relevante no processo de descoberta de novos fármacos. Avanços tecnológicos nas áreas de caracterização estrutural de biomacromoléculas, de ciência computacional e de biologia molecular têm contribuído para que o planejamento de novas moléculas se torne factível (OOMS, 2000).

#### 1.3. ÁCIDOS MICÓLICOS E ENOIL-ACP REDUTASE (INHA)

As micobactérias pertencem a um grupo grande de bactérias Grampositivas, cujo DNA é rico em bases guanina-citosina (GC). Este grupo é denominado Actinomycete. Neste grupo, as micobactérias estão classificadas na subdivisão Corynebacterium-Mycobacterium-Nocardia (CMN), caracterizada pela

10

presença de paredes celulares de constituição química peculiar, chamada parede celular tipo IV (BRENAN, NIKAIDO, 1995; BARRY *et al.*, 1998, PASQUALOTO, FERREIRA, 2001). Os microrganismos que apresentam a parede celular tipo IV estão incluídos em uma categoria distinta denominada *Mycolata* (BRENAN, NIKAIDO, 1995; BARRY *et al.*, 1998).

As paredes celulares das bactérias Gram-positivas são geralmente compostas por grandes quantidades de peptidoglicano, com alguns polissacarídeos ou polímeros poliol fosfato adicionais. A parede celular do tipo IV apresenta ácido meso-diaminopimélico como ácido diamínico na porção peptidoglicano e, em contraste à N-acetilação encontrada nas demais bactérias, o resíduo de ácido murâmico é N-glicolilado nos gêneros Mycobacterium e Nocardia. Uma característica importante da parede celular do tipo IV é a presença de um único polissacarídeo, arabinogalactano, que se encontra conectado a ácidos graxos de cadeia longa, denominados ácidos micólicos (BRENNAN, NIKAIDO, 1995; PASQUALOTO, FERREIRA, 2001) (Figura 1).



Figura 1 – Esquema da parede celular do M. tuberculosis (adaptado de Linda M. Stannard, 1996 [www.uct.ac.za/depts/mmi/lsteyn/cellwall.html]).

KERLY FERNANDA MESQUITA PASQUALOTO

11

Os ácidos micólicos são  $\alpha$ -alquil- $\beta$ -hidroxiácidos de elevado peso molecular, covalentemente ligados à porção arabinogalactano da parede celular do *M. tuberculosis* (BRENAN, NIKAIDO, 1995; BARRY *et al.*, 1998; PASQUALOTO, FERREIRA, 2001). A pirólise dos ácidos micólicos libera a porção distal ao C3 ou  $\beta$  (grupo 3 ou  $\beta$ -hidroxi), gerando ácido meromicólico. Isto permite a análise da estrutura dos ácidos micólicos pela sua separação em duas partes: porção meromicolato e  $\alpha$ ramificação (BRENAN, NIKAIDO, 1995; BARRY *et al.*, 1998).

Nos gêneros Corynebacterium, Rhodococcus e Nocardia, os ácidos micólicos são constituídos no máximo por 40, 50 e 60 átomos de carbono, respectivamente. No entanto, no gênero Mycobacterium apresentam, usualmente, 70 a 90 carbonos. Diferenças na estrutura dos ácidos micólicos podem afetar a fluidez e permeabilidade da bicamada lipídica assimétrica do Mycobacterium, o que explicaria os diferentes níveis de sensibilidade a inibidores lipofílicos entre as várias espécies do microrganismo (BRENAN, NIKAIDO, 1995; BARRY *et al.*, 1998; PASQUALOTO, FERREIRA, 2001).

Os ácidos micólicos da parede celular do M. tuberculosis são distintos dos encontrados em outros gêneros (Corynebacterium, Nocardia, Rhodococcus) devido às seguintes características: (1) apresentam maior número de carbonos (C<sub>70</sub> a C<sub>90</sub>); (2) o tamanho das  $\alpha$ -ramificações é maior (C<sub>20</sub> a C<sub>26</sub>); (3) na cadeia carbônica principal, denominada porção meromicólica, apresentam um ou dois grupos (que podem ser ligação dupla ou anel ciclopropano) capazes de produzir "nós" na molécula, diminuindo sua flexibilidade; (4) podem apresentar funções oxigenadas adicionais ao grupo  $\beta$ -hidroxi; e, (5) podem apresentar ramificações constituídas por grupos metila na cadeia carbônica principal (BRENAN, NIKAIDO, 1995; BARRY et al., 1998; PASQUALOTO, FERREIRA 2001).

No Quadro 2 estão apresentadas as variações no tamanho dos ácidos micólicos para os gêneros Corynebacterium, Nocardia, Rhodococcus e Mycobacterium (BARRY et al., 1998).

high a start for the second		কৰণ উদ্ধৰ সাম কাৰ্যসূচ		्यात्वा देवा क्रिस्टाइड चुक्स
Gênero Nº to	al de carbonos	Nº de ligações d	iupia 👘 a-Ro	mificações
같아요. 그는 것은			liberac	las na pirólise
and the second state of th	and an and the second and a second with		i de la calación de l	And Block and the star
Connebacterium	22-36	0-2	Call Carlot A.	14-18
Colynebacterion		<b>U</b> - <b>Z</b>		14-10
Nocardia	44-60	0-3		12-18
		•••		12 10
Rhodococcus	34 - 48	0-4		12-18
Mycobacterium	70 - 90	1-2	1.1	20-26

Quadro 2 - Variação no tamanho dos ácidos micólicos (BARRY et al., 1998)

As modificações estruturais nos ácidos micólicos ocorrem em dois locais da porção meromicólica ou cadeia carbônica principal, referidos como distal (local mais próximo à porção ω-final da cadeia) e proximal (local mais próximo à porção β-hidroxi ácido da cadeia). Modificações polares se restringem à porção distal e incluem grupos funcionais como metila, éteres, cetonas e epóxidos. Modificações apolares ocorrem nos dois locais da cadeia meromicólica, distal e proximal, e incluem ligações duplas *cis* ou *trans* e grupos ciclopropano *cis* ou *trans*. Os grupos funcionais *trans* apresentam uma ramificação metila adjacente (BARRY *et al.*, 1998; PASQUALOTO, FERREIRA, 2001).

α-Alquil-micolatos, metoximicolatos e cetomicolatos são as três classes de ácidos micólicos produzidas pelo M. *tuberculosis* que diferem entre si, principalmente, na presença e na natureza dos substituintes contendo oxigênio localizados na porção distal da cadeia principal (YUAN, BARRY, 1996; BARRY *et al.*, 1998; PASQUALOTO, FERREIRA, 2001) (Figura 2).

Faculdade de Ciências Farmacêuturat Universidade de São Paulo


Figura 2 - Classes de micolatos encontradas no M. tuberculosis (BARRY et al., 1998).

Os α-alquil-micolatos podem apresentar insaturações (ligações duplas) ou grupos ciclopropano nas configurações cis ou trans, mas quando trans também possuirão uma ramificação metila adjacente (LARSSON et al., 1989; LUQUIN et al., 1991; BARRY et al., 1998). Conforme estudos de George e colaboradores (1995), o diciclopropil micolato foi identificado como o principal ácido micólico produzido pelo *M. tuberculosis*, sugerindo que a ciclopropanação contribui para a integridade da complexa parede celular.

A oxigenação da cadeia meromicólica é encontrada em metoximicolatos e cetomicolatos. Os micolatos que apresentarem uma ramificação metila adjacente a um grupo metoxila, na porção meromicólica distal, em adição à insaturação (cis ou trans) ou à ciclopropanação (cis ou trans), são denominados metoximicolatos, ao passo que os micolatos que apresentarem uma ramificação metila adjacente a um grupo cetona, na porção meromicólica distal, associados com insaturação (cis ou trans) ou ciclopropanação (cis ou trans) são chamados cetomicolatos (LARSSON et al., 1989; LUQUIN et al., 1991; BARRY et al., 1998).

14

O mecanismo biossintético de ácidos graxos, denominado sistema FAS (*Fatty Acid Synthase*), envolve ciclos repetitivos de condensação, ceto redução, desidratação e enoil redução. As enzimas que participam deste processo são consideradas alvos ideais para o planejamento de agentes antibacterianos e antimicobacterianos mais seletivos, devido à diferença na organização molecular entre os sistemas FAS de bactérias/micobactérias e de mamíferos (McCARTHY, HARDIE, 1984; MAGNUSON et al., 1993; BARRY et al., 1998).

Mamíferos, aves e leveduras produzem ácidos graxos via sistema FAS-I, no qual a atividade enzimática reside numa única cadeia polipeptídica de uma enzima multifuncional (KOLATTUKUDY, POULOSE, BUCKNER, 1981; WAKIL, STOOPS, JOSHI, 1983; SMITH, 1994; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000). Em contraste, a maioria das plantas e bactérias utiliza um sistema FAS-II, em que cada atividade enzimática corresponde a um polipeptídio individual (FULCO, 1983; KATER *et al.*, 1994; ROCK, CRONAN, 1996; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000).

O sistema FAS-II melhor caracterizado é o da Escherichia coli, que inclui as seguintes enzimas-acp (acyl carrier protein): β-cetoacil-acp sintetases (FabB, FabF e FabH), β-cetoacil-acp redutase (FabG), β-hidróxi-acil-acp desidrases ou desitratases (FabA e FabZ) e enoil-acp redutase (FabI, conhecida antigamente como EnvM) (MAGNUSON et al., 1993; ROZWARSKI et al., 1999).

A enzima enoil-acp redutase catalisa a reação principal, que regula o ciclo de alongamento de ácidos graxos (BERGLER *et al.*, 1996; STEWART *et al.*, 1999). A reação consiste na redução estereoespecífica, NADH-dependente, de ácidos graxos  $\alpha$ , $\beta$ -insaturados ligados à proteína acil-carregadora (ROZWARSKI *et al.*, 1999; STEWART *et al.*, 1999).

O sistema FAS-II da Escherichia coli inicia a biossíntese de ácidos graxos pela condensação dos substratos acetil e malonil; continua o processo com a etapa de

15

alongamento do produto resultante da condensação pela adição de unidades consecutivas de  $C_2$  até que a cadeia alquílica do ácido graxo atinja aproximadamente 16 carbonos (ROZWASRKI *et al.*, 1999).

As micobactérias possuem ambos os sistemas de biossíntese de ácidos graxos, FAS-I e FAS-II (BLOCH, 1977; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000).

O sistema FAS-I das micobactérias produz ácidos graxos com 16 e 24-26 carbonos (BLOCH, 1977; KIKUCHI, RAINWATER, KOLATTUKUDY, 1992; BARRY *et al.* 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000), enquanto que o sistema FAS-II prefere substratos de partida com 16 carbonos (BLOCH, 1977; BARRY *et al.*, 1998; MARRACHI, LANÉELLE, QUÉMARD, 2000) e pode alongá-los até 56 carbonos (QURESHI, SATHYAMOORTHY, TAKAYAMA, 1984; BARRY *et al.*, 1998; MARRACHI, LANÉELLE, QUÉMARD, 2000). Isto indica que o sistema FAS-II das micobactérias utiliza os produtos do sistema FAS-I como base para o alongamento mais uniforme da cadeia de ácido graxo. Os produtos de cadeia mais longa do sistema FAS-II são os precursores dos ácidos micólicos, principais componentes da parede celular micobacteriana (BRENAN, NIKAIDO, 1995; LEE, BRENNAN, BESRA, 1996; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000) (Figura 3).

# INTRODUÇÃO

16



CoA = coenzima A AcpM = proteína acil carregadora KAS = beta-ceto acil sintetase MabA = ceto-acil redutase InhA = enoil redutase

Figura 3 – Sistema micobacteriano FAS II responsável pela produção de ácidos micólicos – enzimas envolvidas na biossíntese da cadeia meromicólica (BARRY et al., 1998).

Três classes de inibidores de enoil-acp redutase complexados com seus respectivos alvos enzimáticos foram caracterizados: isoniazida, que inibe a InhA, enoil-acp redutase do *M. tuberculosis* (ROZWARSKI *et al.*, 1998); diazoborinas (BALDOCK *et al.*, 1996; BALDOCK *et al.*, 1998) e triclosana (McMURRY, OETHINGER, LEVY, 1998; STEWART *et al.*, 1999), compostos que inibem a Fabl, enoil-acp redutase da *E. coli*. Estudos recentes relataram que a triclosana também inibe a InhA (PARIKH, XIAO, TONGE *et al.*, 2000). As estruturas destes inibidores estão apresentadas no Quadro 3.



A isoniazida (hidrazida do ácido isonicotínico) é um bioprecursor, utilizada como fármaco de primeira escolha nos esquemas terapêuticos de TB. Sua forma ativa, que pode ser um ânion ou radical, liga-se covalentemente ao átomo C4 da nicotinamida do cofator (NADH) complexado à enzima, resultando na formação de acilpiridina:NAD (ROZWARSKI et al., 1998; ROZWARSKI et al., 1999) (Quadro 4).



Quadro 4 - Mecanismo proposto para a formação do inibidor da enoil-acp redutase do M.

R = restante da estrutura do cofator NAD (nicotinamida adenina dinucleotídeo)

-0

As diazoborinas constituem uma família experimental de compostos, que têm propriedades antibacterianas contra variedade de bactérias Gram-negativas. Estes compostos também formam ligação covalente com o cofator (NADH)

H

0

19

complexado à enzima. Neste caso, a ligação ocorre entre o átomo de boro do inibidor e o grupo 2'-hidroxila da ribose do NAD (BALDOCK *et al.*, 1996; BALDOCK *et al.*, 1998). Estudos de diazoborinas de diferentes classes complexadas à enoil-acp redutase de *E. coli* foram relatados na tentativa de elucidar as características destas interações que modulam a afinidade enzimática (LEVY *et al.*, 2001).

A comparação de estruturas cristalizadas dos inibidores, diazoborina e triclosana, complexados à Fabl revelou que as porções adenosina do cofator são superponíveis e que há diferenças nas porções nicotinamida. Ambos os inibidores interagem somente com a nicotinamida, mas a triclosana forma ligações de hidrogênio com o grupo 2'-hidroxila da ribose do NAD (STEWART *et al.*, 1999).

A triclosana é um agente antibacteriano de amplo espectro, ativo contra vários microrganismos: bactérias Gram-positivas (Bacillus subtilis, M. smegmatis, Staphylococcus aureus), bactérias Gram-negativas (E. coli, Salmonella typhimurium, Shigella flexneri), fungos e leveduras. Devido ao seu perfil de segurança, esse fármaco tem sido utilizado como aditivo antibacteriano em muitos produtos de higiene pessoal (BHARGAVA, LEONARD, 1996; STEWART et al., 1999).

A enoil-acp redutase do M. tuberculosis, InhA, apresenta seqüência total de aminoácidos com 28% e 23% de homologia às enoil-acp redutases de E. coli (Fabl) e da planta Brassica napus, respectivamente (ROZWASRKI et al., 1999). Na região do sítio ativo, a semelhança estrutural entre InhA/Fabl e InhA/enoil-acp Brassica napus é de 40% e 33%, respectivamente (BALDOCK et al., 1996). As enoil-acp redutases de E. coli e Brassica napus compartilham 35% de homologia e os resíduos de aminoácidos que formam o sítio ativo são muito conservados nas duas proteínas (KATER et al., 1991; BALDOCK et al., 1996).

As mutações enzimáticas, próximas ao sítio ativo (BALDOCK et al., 1996), resultantes da substituição de um único aminoácido, como por exemplo Ser94Ala, na InhA (DESSEN et al., 1995; QUÉMARD et al., 1995; BALDOCK et al., 1996; ROZWARSKI et al., 1998) e Gly93Ser, na Fabl (TURNOWSKY et al., 1989; BERGLER, HÖGENAUER, TURNOWSKY, 1992; BALDOCK et al., 1996; BALDOCK et al., 1998) são suficientes para conferir resistência à isoniazida e às diazoborinas, respectivamente.

Cinco mutações na InhA, além de Ser94Ala, foram observadas em isolados clínicos de M. *tuberculosis* isoniazida-resistente: Ile16Thr, Ile21Val, Ile47Thr, Val78Ala e Ile95Phe. As enzimas InhA fármaco-resistentes apresentam afinidade reduzida pelo cofator NADH (BASSO *et al.*, 1998; PARIKH *et al.*, 1999).

Em contraste com a Fabl, a enoil-acp da Brassica napus é insensível às diazoborinas e a presença de um resíduo alanina na posição 138 (que equivale estruturalmente à posição 93 da Fabl) parece ser o fator determinante para esta resistência (ROUJEINIKOVA et al., 1999; DE BOER et al., 1999).

Três mutações na Fabl resultam em resistência à triclosana: Gly93Val, Met159Thr e Phe203Leu. A presença do resíduo valina na posição 93 da Fabl também resulta em resistência às diazoborinas, sugerindo modo de ação comum para estes dois tipos de inibidores (HEATH *et al.*, 1998; STEWART *et al.*, 1999).

Estudos recentes sobre a inibição da InhA pela triclosana e pela isoniazida (McMURRY, OETHINGER, LEVY, 1998) relataram que a substituição do resíduo tirosina por fenilalanina na posição 158 reduz a afinidade da triclosana pela enzima. O grupo hidroxila da tirosina 158 é importante para a interação da triclosana com a enzima InhA, da mesma forma que o grupo hidroxila da correspondente tirosina 156, na Fabl. Estes resíduos (Tyr158 e Tyr156) contribuem de forma relevante no mecanismo catalítico das enoil redutases. Isto sugere que a triclosana interage de modo semelhante com ambas as enzimas, InhA e Fabl (McMURRY, OETHINGER, LEVY, 1998; PARIKH *et al.*, 1999). Foi demonstrado, também, que as mutações 11e47Thr e 11e21Val na InhA, que ocorrem em isolados clínicos de *M. tuberculosis* isoniazida-resistente, não prejudicaram a inibição pela triclosana. Esta observação parece Indicar que inibidores da InhA, que tenham como alvo o sítio de ligação do

enoil-substrato, podem ser efetivos contra cepas de *M. tuberculosis* isoniazidaresistente (McMURRY, OETHINGER, LEVY, 1998).

O conhecimento das características bioquímicas e estruturais das enzimas enoil-acp redutases supre informações importantes de como as mutações interferem nas interações fármaco-receptor e permite o planejamento de novos agentes terapêuticos eficientes contra as cepas sensíveis e MDR.

۴

JUSTIFICATIVA E OBJETIVO

.

.

.

### 2. JUSTIFICATIVA E OBJETIVO

Face ao quadro alarmante da TB no mundo e, em particular no Brasil, onde são registrados 161.800 novos casos anualmente e considerando-se os índices elevados de resistência do microrganismo aos fármacos convencionalmente utilizados, há necessidade premente de novas e específicas alternativas quimioterápicas.

Os constituintes da parede celular da micobactéria - ácidos micólicos, arabinogalactano, peptidoglicano e micobactinas - são considerados alvos específicos para novos fármacos anti-TB. Agentes que interferem na biossíntese destes constituintes seriam seletivos e menos tóxicos, atuando especificamente na micobactéria, sem alterar a flora bacteriana do hospedeiro (SENSI, 1989; BARRY, 1997).

Os ácidos micólicos representam o principal constituinte da complexa parede celular micobacteriana e encontram-se implicados na sua permeabilidade e na sua fluidez. A baixa permeabilidade da parede celular parece contribuir, de forma relevante, para a menor sensibilidade do bacilo aos antibióticos e quimioterápicos (LIU, 1996).

À vista do exposto, o objetivo do presente estudo é, com base na estrutura dos ácidos micólicos, planejar tuberculostáticos potencialmente ativos na infecção, utilizando como ferramenta as técnicas de modelagem molecular (CADD): QSAR-4D IR (relação quantitativa quadridimensional entre estrutura química e atividade biológica independente do receptor) e FEFF QSAR-3D (relação quantitativa tridimensional entre estrutura química e atividade biológica com campo de força de energia livre), em um conjunto de 37 hidrazidas, com mecanismo de ação. possivelmente comum ao da isoniazida. Espera-se, com esta metodologia, obter mais subsídios para a concepção racional de análogos estruturais da isoniazida.

METODOLOGIA

.

.

,

# 3. METODOLOGIA

#### 3.1. DADOS BIOLÓGICOS

Foram selecionadas da literatura trinta e sete hidrazidas, incluindo o fármaco isoniazida (hidrazida do ácido isonicotínico, INH1), avaliadas pelo mesmo protocolo farmacológico e cujas atividades foram medidas como a concentração mínima capaz de inibir o crescimento de cepas de Mycobacterium tuberculosis var. bovis, à temperatura, T, de 310 K (37 °C) (BERNSTEIN *et al.*, 1952; BERNSTEIN *et al.*, 1953a; BERNSTEIN *et al.*, 1953b; KLOPMAN, FERCU, JACOB, 1996). Conforme apresentado no Quadro 5, os dados de atividade biológica, MIC (minimum inhibitory concentration; concentração inibitória mínima), foram transformados em valores de - log MIC (pMIC, g/L) e, automaticamente, convertidos em concentração molar (mol/L) pelo programa 4D-QSAR 3.0 (HOPFINGER, 1999).

Ácidos que originam hidrazidas	Código	MIC (µg/mL)	p MIC (g/L; 10-4/10-3)
ácido isonicotínico	INHI	0,02	4,70
ácido picolínico	INHd2	0,15	3,82
ácido cinâmico	INHd31	0,40	3,40
ácido piridina-2-flúor-4-carboxílico	INHd43	0,40	3,40
ácido 3-nitrobenzólco	INHd20	0,60	3,22
ácido benzóico	INHd14	0,60	3,22
ácido piridina-3-amino-4-carboxílico	INHd46	1,50	2,82
ácido nicotínico	INHd37	2,00	2,70
ácido 4-clorobenzóico	INHd15	,3,00	2,52
ácido 3-aminobenzóico	INHd23	3,00	2,52
ácido 4-clorofenilacético	INHd29	10,00	2,00
ácido 4-bromobenzóico	INHd16	12,00	1,92
ácido 4-nitrobenzóico	INHd18	12,00	1,92

Quadro 5 - Valores de MIC e pMIC das trinta e sete hidrazidas selecionadas

Quadro 5 (continuação) - Valores de MIC e pMIC das trinta e sete hidrazidas selecionadas

Ácidos que originam hidrazidas	Código	MIC (µg/ml.)	pMIC (g/L; 10-6/10-3)
ácido plridina-2-bromo-4-carboxílico	INHd44	15,00	1,82
ácido 4-hidroxibenzóico	INHd25	12,00	1,92
ácido fenilacético	INHd30	12,00	1,92
ácido pirimidina-4-carboxílico	ldv130	20,00	1,70
ácido 4-hidróxi-3-metoxibenzóico	INHd27	30,00	1,52
ácido 2-aminobenzóico	INHd22	30,00	1,52
ácido mandélico	INHd34	40,00	1,40
ácido 4-quinolina-carboxílico	INHd42	80,00	1,10
ácido 4-pindinilacético	INHd47	100,00	1,00
ácido benzo[b]furanina-2-carboxílico	ldv107	225,00	0,65
ácido piridina- 3-metil-4-carboxílico	INHd45	200,00	0,70
ácido imidazol-2-mercapto-5-carboxílico	ldv126	250,00	0,60
ácido 2-nitrobenzólco	INHd19	300,00	0,52
ácido Imidazol-1-metil-5-carboxílico	ldv125	300,00	0,52
ácido pirimidina-2-metil-4-hidróxi-5- carboxílico	INHd41	400,00	0,40
ácido 1-metil-4-piperidina-carboxílico	INHd49	600,00	0,22
ácido 4-piperidina-carboxílico	INHd48	600,00	0,22
ácido piromúcico	ldv90	0,06	4,22
ácido pirazina-carboxílico	⊡ldv128	1,50	2,82
ácido 5-imidazol-carboxílico	Idv124	2,00	2,70
ácido tiazol-5-carboxílico	ldv131	10,00	2,00
ácido tiazol-4-carboxílico	ldv132	15,00	1,82
ácido (2-amino-1,3,4-tiadiazol-4-il)acético	ldv136	300,00	0,52
ácido 1-dimetilaminocarbonil-4-piperidina-	INHd51	600,00	0.22
carboxílico			

INHd - hidrazidas aromáticas, heteroaromáticas e com substituição no anel, derivadas da isoniazida;

Idv - hidrazidas de ácidos heterocíclicos e derivados.

KERLY FERNANDA MESQUITA PASQUALOTO

27

Cabe ressaltar que a atividade biológica dos compostos foi testada *in vitro* contra cepas BCG de M. *tuberculosis* utilizando o meio modificado de Kirchner e em testes padronizados *in vivo* contra cepas Ravenel de M. *tuberculosis* var. *bovis* em ratos (RAKE *et al.* 1949). Os valores de equivalente de isoniazida para os compostos sintetizados e testados por Bernstein e colaboradores (1952, 1953a, 1953b) foram calculados por meio da razão entre as doses molares efetivas de isoniazida e dos respectivos compostos, na faixa de concentração mínima efetiva. A dose diária mínima efetiva de isoniazida foi menor ou igual a 2 mg/kg e as doses máximas toleradas foram de 64 mg/Kg (administração oral, dieta) e 125 mg/kg (administração intravenosa). Diferenças nos valores de equivalente de isoniazida de um fator de dois foram consideradas não significativas.

### 3.1.1. DEFINIÇÃO DOS CONJUNTOS DE TREINAMENTO (TRAINING SET) E DE AVALIAÇÃO (TEST SET)

A seleção dos trinta e sete compostos da literatura (BERNSTEIN et al., 1952; BERNSTEIN et al., 1953a; BERNSTEIN et al., 1953b; KLOPMAN, FERCU, JACOB, 1996) foi efetuada considerando-se a semelhança estrutural com a isoniazida, os valores de atividade biológica (MIC), provenientes de ensaio biológico comum, bem como, a possibilidade de apresentarem mecanismo de ação similar ao do fármaco protótipo, conforme sugerido por ROZWARSKI et al. (1998).

Assumiu-se, no presente estudo, a hipótese de que todos os compostos agiriam como o protótipo, INH1. Após a 'perda' do grupo hidrazida, a forma ativa (ânion ou radical) formaria ligação covalente com o cofator nicotinamida adenina dinucleotídeo (NAD, carbono C<sub>4</sub> do anel nicotinamida), resultando em um inibidor fortemente ligado (ROZWARSKI *et al.*, 1998) (Quadro 4).

Dos trinta e sete compostos selecionados (BERNSTEIN et al., 1952; BERNSTEIN et al., 1953; BERNSTEIN et al., 1953; KLOPMAN, FERCU, JACOB, 1996), 13 foram considerados ativos (valores de MIC menores que 10 µg/mL), 12 foram considerados

METODOLOGIA

28

de atividade intermediária (valores de MIC entre 10 e 50  $\mu$ g/mL) e 12 foram considerados inativos (valores de MIC maiores de 50  $\mu$ g/mL).

Os compostos foram divididos em dois conjuntos: conjunto de treinamento (trinta compostos, incluindo a isoniazida) e conjunto de avaliação (sete compostos).

O conjunto de treinamento é constituído de compostos selecionados aleatoriamente do conjunto total, que, por definição, serão utilizados em todas as etapas da metodologia, incluindo a obtenção e a validação interna dos modelos (equações de QSAR) (HOPFINGER, 1999).

Considerando-se os valores de atividade biológica, o conjunto de treinamento apresenta dez compostos ativos (INH1, INHd2, INHd31, INHd43, INHd20, INHd14, INHd46, INHd37, INHd15, INHd23), dez compostos de atividade intermediária (INHd29, INHd16, INHd18, INHd44, INHd25, INHd30, Idv130, INHd22, INHd27, INHd34) e dez compostos inativos (INHd42, INHd47, Idv107, INHd45, Idv126, INHd19, Idv125, INHd41, INHd49, INHd48) (Quadro 6).



Quadro 6 - Conjunto de treinamento e seus respectivos grupos de atividade biológica

KERLY FERNANDA MESQUITA PASQUALOTO





Quadro 6 (continuação) - Conjunto de treinamento e seus respectivos grupos de atividade

Q conjunto de avaliação também é constituído de compostos selecionados do conjunto total, porém estes compostos não são incluídos na etapa de obtenção dos modelos, mas participam apenas na etapa de validação externa dos modelos. Os sete compostos que compõem o conjunto de avaliação foram sorteados do total de compostos, considerando-se cada grupo de atividade biológica. Estes compostos foram utilizados para validar o modelo QSAR construído a partir do conjunto de treinamento e para avaliar a capacidade de predição do modelo obtido (HOPFINGER, 1999).

Considerando-se os valores de atividade biológica, o conjunto de avaliação apresenta três compostos ativos (Idv90, Idv128, Idv124), dois compostos de atividade intermediária (Idv131, Idv132) e dois compostos inativos (Idv136, INHd51) (Quadro 7).

KERLY FERNANDA MESQUITA PASQUALOTO



Quadro 7 - Conjunto de avaliação e seus respectivos grupos de atividade biológica

No gráfico 1 está apresentada a distribuição das atividades biológicas (pMIC) dos conjuntos de treinamento e de avaliação. Os valores de pMIC apresentaram um desvio padrão de 1,2 do valor médio.

> BIBLIOTECA Faculdade de Ciéncias Farmacêuticas Universidade de São Paulo



Gráfico 1 – Distribuição da concentração inibitória mínima (pMIC) em cepas de M. tuberculosis var. bovis, 310 K, dos compostos do conjunto de treinamento (30 compostos) e de avaliação (7 compostos, ★).

# 3.2. CONSTRUÇÃO E MINIMIZAÇÃO DAS ESTRUTURAS

As estruturas tridimensionais dos compostos em estudo foram construídas utilizando o módulo "Builder" do programa de modelagem molecular HyperChem 6.0 (Hypercube Inc., 1996). A geometria correspondente à cristalografia de raios-X de INH1 (forma ativa da isoniazida, isonicotínico-acil:NAD) (ROZWARSKI et al., 1998), complexado com a enzima enoil-acp redutase do Mycobaterium tuberculosis, InhA (**1zid**, código de acesso no PDB, Protein Data Bank, resolução 2,7 Å), foi utilizada como referência para criar as geometrias iniciais dos demais compostos. O programa MOLSIM 3.0 (DOHERTY, 1994 - The Chem21 Group, Inc.) foi utilizado para otimizar a geometria de todos os compostos e o método semi-empírico AM1 (Austin

KERLY FERNANDA MESQUITA PASQUALOTO

Model 1- DEWAR et.al., 1985) foi empregado para obter as cargas atômicas parciais dos compostos, necessárias na etapa de dinâmica molecular.

# 3.3. METODOLOGIA DE QSAR-4D (IR, Independente do Receptor)

## 3.3.1. AMOSTRAGEM CONFORMACIONAL POR DINÂMICA MOLECULAR

Esta etapa da metodologia de QSAR-4D (IR) corresponde aos cálculos de dinâmica molecular (DM) dos ligantes, com a finalidade de gerar uma amostragem conformacional ("conformational ensemble") para cada composto do conjunto de treinamento, visando, posteriormente, a escolha da conformação 'ativa'. O objetivo desta etapa é obter a distribuição de Boltzmann do conjunto das possíveis conformações disponíveis para cada molécula do conjunto de treinamento. As conformações geradas durante a simulação de DM de cada ligante constituem o perfil da amostragem conformacional ("conformational ensemble profile", CEP). O CEP permite a investigação da flexibilidade conformacional molecular. Convém lembrar que a quarta dimensão na metodologia de QSAR-4D corresponde à amostragem conformacional relacionada ao tempo de simulação na DM (HOPFINGER, 1999).

Em princípio, qualquer estrutura 3D é aceitável para iniciar a amostragem conformacional. Na prática, principalmente para um conjunto de treinamento de análogos, uma conformação comum, com baixa energia (não necessariamente a conformação de energia mínima), considerando ângulos torsionais comuns nas estruturas do conjunto de treinamento, deveria ser selecionada. Isto seria o ponto de referência para o bom andamento de uma análise QSAR-4D (HOPFINGER, 1999).

Neste estudo, as estruturas minimizadas de cada composto ligado ao cofator (ligante) foram utilizadas como estruturas iniciais em cada simulação de DM. Cinco mil conformações de cada ligante foram amostradas para gerar o CEP a partir de simulações de DM de cem mil estados da trajetória do confôrmero ou

cem mil passos, cada passo de 1 fs. As conformações geradas para cada ligante foram registradas a cada 20 passos no arquivo de trajetória, totalizando 5000 conformações. A temperatura durante as simulações de DM foi mantida constante em 310 K, que corresponde à temperatura do ensaio biológico dos compostos em estudo. As simulações de DM foram realizadas em um banho térmico (BERENDSEN et al., 1984) com tempo de relaxação de 0,1 ps.

Os cálculos de DM foram realizados utilizando-se o programa MOLSIM 3.0 (DOHERTY, 1994), um pacote de mecânica e dinâmica molecular, incluído no programa 4D-QSAR 3.0 (HOPFINGER, 1999 - The Chem21 Group, Inc.), com campo de força MM2 (ALLINGER, 1977), que apresenta parametrizações relatadas por Hopfinger e Pearlstein (1984). Conforme mencionado anteriormente, as cargas atômicas parciais dos ligantes foram computadas utilizando o método semiempírico AM1 (DEWAR *et al.*, 1985).

Foi assumida a condição de estado não-ionizado para os ligantes investigados (hidroxilas dos grupos fosfato da porção NAD com seus respectivos átomos de hidrogênio). De acordo com as observações de Klopmann e colaboradores (1996), as hidrazidas em estudo teriam seus grupos ácido/base nãoionizados no pH celular devido aos seus valores de pKa serem muito distantes do pH fisiológico.

#### 3.3.2. DEFINIÇÃO DOS ELEMENTOS DE INTERAÇÃO FARMACOFÓRICA

O programa de 4D-QSAR 3.0 (HOPFINGER, 1999) permite definir sete tipos de elementos de interação farmacofórica ("Interaction Pharmacophore Elements", IPEs), que correspondem aos tipos de átomos que podem participar de interações específicas com o receptor (biomacromolécula). Portanto, os átomos de cada molécula do conjunto de treinamento serão classificados de acordo com o tipo de interação que eles sejam capazes de realizar com o receptor. Neste estudo foram selecionadas as sete classes de IPEs: átomo polar positivo (p+; nº código = 2), átomo polar negativo (p-; nº código = 3), átomo apolar (np; nº código = 1), átomo aceptor de ligação de hidrogênio (hba; nº código = 5), átomo doador de ligação de hidrogênio (hbd; nº código = 4), átomo em sistema aromático (ar; nº código = 6) e qualquer tipo de átomo (a; nº código = 0).

# 3.3.3. DEFINIÇÃO DO TAMANHO DAS CÉLULAS

As células que compõem a caixa tridimensional (3D) virtual, que mimetiza o espaço no sítio ativo da enzima ou do receptor, são cúbicas e suas dimensões podem ser definidas pelo usuário. O programa de 4D-QSAR 3.0 (HOPFINGER, 1999) não permite números fracionários para a aresta das células da grade. Quanto menor o tamanho da célula, maior será o número de variáveis geradas, tornando mais trabalhosa a etapa de pré-tratamento das variáveis e a de geração dos modelos. Em compensação, os modelos são mais refinados em termos de tipos de átomos (IPEs) selecionados, ou seja, a probabilidade de uma célula ser classificada como qualquer tipo de átomo (a), inespecífica em termos de IPEs, diminui com células de tamanho reduzido.

Após avaliação preliminar, neste estudo, foi utilizada uma caixa 3D virtual de células cúbicas com 1 Å de aresta, que é o valor 'default' do programa 4D-QSAR 3.0 (HOPFINGER, 1999)(Figura 4).



Figura 4 - Representação geral da célula da caixa 3D virtual onde as conformações amostradas são colocadas.

O tamanho da célula da grade, ou da caixa 3D virtual, é considerado na análise QSAR-4D como um parâmetro metodológico e o modelo QSAR-3D pode ser otimizado como função do número de parâmetros metodológicos. Além do tamanho da célula da caixa 3D virtual, outros parâmetros metodológicos podem ser considerados em uma análise QSAR-4D (IR): temperatura da simulação de DM; molécula de referência; tamanho do conjunto amostrado (número de conformações de partida distintas na amostragem); número de passos ("frames") registrados no CEP; número de alinhamentos; número de descritores selecionados inicialmente para utilização do algoritmo genético; tipos de IPEs incluídos na análise (HOPFINGER, 1999).

#### 3.3.4. DEFINIÇÃO DOS ALINHAMENTOS TESTADOS

Esta é uma das etapas mais importantes na metodologia de QSAR-4D(IR) e corresponde à seleção dos alinhamentos que serão testados na sobreposição de todas as conformações, de todos os compostos do conjunto de treinamento, obtidas na etapa de simulação de DM para a inclusão na caixa 3D virtual. O alinhamento das estruturas considera a suposição de que não somente o sítio de ligação é idêntico, mas também que o modo de ligação é idêntico, ou similar. Portanto, a orientação geral das moléculas dentro do sítio ativo seria comum para todos os compostos, o que não necessariamente é considerado como verdadeiro (HOPFINGER, 1999).

Os alinhamentos são selecionados para explorar cada uma das principais partes da molécula, bem como as combinações possíveis das partes principais de uma molécula. O algoritmo de sobreposição usual, na metodologia QSAR-4D (IR), considera somente regras de alinhamento combinando três átomos quaisquer. Porém, não há restrição no algoritmo que impeça a inclusão de qualquer regra de alinhamento. No entanto, a ordem dos três átomos selecionados para cada alinhamento é importante. Por exemplo, espera-se que o primeiro átomo especificado na molécula A ocupe uma posição semelhante no espaço que o primeiro átomo especificado na molécula B para o alinhamento selecionado. Isto também é válido para o segundo e o terceiro átomos. Além disso, os alinhamentos devem ser selecionados para abranger toda a topologia da estrutura química comum dos compostos do conjunto de treinamento (HOPFINGER, 1999).

Neste estudo, foram testados sete alinhamentos distintos, definidos na Tabela I como alinhamentos 1, 2, 3, 4, 5, 6 e 7, respectivamente. A estrutura do composto INH1 (isonicotínico-acil:NAD) do conjunto de treinamento foi utilizada para demonstrar que cada um dos alinhamentos testados era constituído por uma seqüência diferenciada de três átomos, de modo a abranger toda a estrutura molecular do ligante (composto:NAD) (Tabela I).

Tabela I – Definição da seqüência de três átomos, designados pelas letras "**a-I**", para cada um dos sete alinhamentos testados, utilizando a estrutura do composto INH1 como exemplo



Alinhamento	1º átomo	2º átomo	3º átomo
1	a	b	c
2	ь	d	e
3	a	c	t
4	ь		9

KERLY FERNANDA MESQUITA PASQUALOTO

Tabela I (continuação) – Definição da seqüência de três átomos, designados pelas letras "**a**-I", para cada um dos sete alinhamentos testados, utilizando a estrutura do composto INH1 como exemplo

1º átomo	2º átomo	3º átomo
a	g	h
T	J	k
ь	g	1
	l° átomo a i b	l°átomo 2°átomo a g i j b g

#### 3.3.5. DEFINIÇÃO DOS TIPOS DE OCUPAÇÃO

A freqüência de ocupação das células da caixa virtual é computada para as conformações registradas na trajetória de DM para todos os compostos em estudo.

A freqüência de ocupação das células da caixa virtual pode ser calculada por três tipos de medida de ocupação para cada IPE (**3.3.2.**). O primeiro tipo é denominado ocupação absoluta (A<sub>o</sub>, absolute-occupancy) e corresponde à soma de todos os átomos IPE que ocupam as células em que estão as conformações do ligante em estudo, registradas na simulação de DM, independente de um composto de referência. A ocupação associada (J<sub>o</sub>, *joint-occupancy*), segundo tipo de medida de ocupação, é definida como o número de átomos IPE que ocupam as células da grade comuns para as conformações do ligante em estudo e um composto de referência. Ou seja, a ocupação das células referente a um composto de referência é somada. O terceiro tipo é denominado auto-ocupação (S<sub>o</sub>, *self-occupancy*), onde o algoritmo QSAR-4D calcula a diferença entre a ocupação absoluta das conformações do ligante em estudo e a ocupação associada para as conformações do ligante em estudo e o composto de referência (HOPFINGER, 1999).

Neste estudo, a freqüência de ocupação absoluta das células foi utilizada por ser o único tipo que independe de um composto de referência. Além disso,

fornece informações mais específicas sobre o tipo de átomo (IPE) que ocupa preferencialmente a célula da caixa virtual em relação ao ligante em estudo.

# 3.3.6. OBTENÇÃO DOS DESCRITORES DE OCUPAÇÃO DAS CÉLULAS (GCODS)

Nesta etapa, cada conformação do CEP de cada composto é colocada na caixa 3D virtual de acordo com o alinhamento selecionado. Uma grade de referência, que consiste em um espaço cartesiano fixo, é gerada automaticamente para cada molécula. Esta grade possui eixos x, y e z, que excedem as coordenadas atômicas máxima e mínima de cada molécula do conjunto de treinamento. Os perfis de ocupação das células da caixa virtual para cada um dos IPEs selecionados (3.3.2.) são computados e utilizados como base para formar o conjunto de descritores na construção dos modelos de QSAR-4D (IR) (HOPFINGER, 1999).

Os valores de ocupação da caixa virtual referentes a cada IPE são denominados GCODs ("grid cell occupancy descriptors"). Estes valores constituem as variáveis independentes (descritores) na metodologia de QSAR-4D (IR) e são calculados conforme o tipo de medida de ocupação selecionada (HOPFINGER, 1999). Neste estudo, os GCODs foram computados de acordo com os valores de ocupação absoluta dos IPEs, para cada um dos sete alinhamentos. Estes valores podem ser normalizados de acordo com o número de conformações selecionadas, ou seja, podem ser transformados em valores percentuais.

# 3.3.7. PRÉ-TRATAMENTO DOS DESCRITORES (GCODS) - REDUÇÃO DO NÚMERO TOTAL DE GCODS

A análise de QSAR-4D, como a análise comparativa de campos moleculares (CoMFA, "Comparative Molecular Field Analysis") (CRAMER, PATTERSON, BUNCE, 1988), gera um número enorme de variáveis independentes (descritores, GCODs) devido ao grande número de células da caixa virtual, aos tipos de IPEs selecionados e suas correspondentes freqüências de ocupação (A<sub>o</sub>, J<sub>o</sub> e S<sub>o</sub>). O

40

conjunto resultante de propriedades das células da caixa virtual constitui o conjunto de descritores (GCODs), para cada um dos alinhamentos selecionados. Entretanto, os esquemas de QSAR-4D diferem da metodologia de QSAR-3D, CoMFA, porque definem, intrinsicamente, o conjunto total de células da caixa virtual que será incluído na análise. Cortes (*cuttoffs*) de potencial de avaliação, utilizados no método CoMFA, não são necessários na metodologia de QSAR-4D (HOPFINGER, 1999).

No formalismo de QSAR-4D, a análise de regressão de mínimos quadrados parciais (PLS, "Partial Least Squares") (GLEN, DUNN, SCOTT, 1989) é utilizada para promover o ajuste de redução de dados entre as variáveis dependentes (atividade biológica observada ou experimental) e os correspondentes valores de GCODs (HOPFINGER, 1999).

Antes de submeter os dados à análise de regressão PLS, dois critérios podem ser empregados para aumentar a velocidade de processamento e melhorar a qualidade da análise de PLS. O usuário pode 'filtrar' os GCODs utilizando o critério de variância ou o de valor mínimo para o descritor. No critério de variância, qualquer coluna de descritor que apresentar variância menor que o valor especificado pelo usuário pode ser omitida da análise de PLS. No critério de valor mínimo para o GCOD, as colunas de GCODs que apresentarem valores de ocupação menores que o valor especificado pelo usuário, considerando todo o conjunto de moléculas, podem ser omitidas da análise de PLS. Somente um dos critérios de pré-tratamento de dados pode ser selecionado. Os descritores com maior peso resultantes da análise de regressão PLS serão utilizados como conjunto de partida para gerar os modelos finais (otimizados) de QSAR-4D (HOPFINGER, 1999).

Neste estudo, a análise de regressão PLS foi empregada para promover uma redução inicial dos dados entre as medidas da variável dependente experimental,

-

pMIC, e os correspondentes descritores, GCODs, para cada um dos sete alinhamentos. O critério de escolha para o pré-tratamento dos GCODs foi o 'filtro' de variância. As variáveis independentes com valores de variância menor que 2,0 (valor 'default' do programa 4D-QSAR 3.0) foram eliminadas. A redução automática dos GCODs pela análise de PLS proporciona a seleção dos GCODs que apresentam maior correlação com a atividade biológica. Os 200 GCODs que tiveram melhor peso na análise de regressão PLS foram utilizados para constituir a série selecionada de descritores para a análise de aproximação da função genética (GFA, "genetic function approximation") (ROGERS, 1991; ROGERS, HOPFINGER, 1994).

### 3.3.8. OBTENÇÃO DOS MODELOS DE QSAR-4D (IR)

Nesta etapa são gerados os modelos de QSAR, partindo da construção, otimização, comparação e avaliação dos modelos através da utilização de um algoritmo genético, GA (HOLLAND, 1975). Os modelos ou equações de QSAR-4D são obtidos pelo método que combina algoritmo genético e mínimos quadrados parciais (GA-PLS) existente no programa 4D-QSAR 3.0 (HOPFINGER, 1999). O algoritmo genético específico implementado no programa 4D-QSAR é a análise de aproximação da função genética, GFA (ROGERS, 1991; ROGERS, HOPFINGER, 1994). Os 200 GCODs que tiveram melhor peso na análise de PLS, para cada um dos sete alinhamentos, foram selecionados como conjunto inicial de dados para a análise de GA. Outros descritores, não provenientes da análise de QSAR-4D, como por ex., logP (logaritmo do coeficiente de partição n-octanol-água), refratividade molar, descritores de mecânica quântica, podem ser adicionados, pelo usuário, ao conjunto inicial de dados no começo desta etapa (HOPFINGER, 1999). Considerando-se que a lipofilicidade é fator determinante na farmacocinética e farmacodinâmica dos fármacos (MANNHOLD, van de WATERBEEMD, 2001), neste estudo, os valores de logP foram calculados (ClogP) para os compostos do

42

conjunto de treinamento, utilizando o programa HyperChem 6.0 (Hypercube, Inc. 1996), e incluídos na etapa de análise de GA.

A vantagem do método GFA é a construção e análise de múltiplos modelos, ao invés de apenas um único modelo (ROGERS, HOPFINGER, 1994). Modelos de uma população inicial são criados a partir da escolha aleatória de variáveis do conjunto formado pelos GCODs de maior peso provenientes da análise de PLS. Os modelos são avaliados pela medida estatística conhecida como LOF ("lack-of-fit score") ou fator de desajuste de Friedman (FRIEDMAN, 1988; ROGERS, 1991; ROGERS, HOPFINGER, 1994), que é a medida dos mínimos quadrados penalizada. Portanto, quando dois modelos apresentarem o mesmo LSE ("least square error", erro de mínimos quadrados), o que tiver o menor número de termos (descritores) terá o melhor (menor) valor de LOF (ROGERS, HOPFINGER, 1994). Novos modelos são criados pelo processo de recombinação ("crossover"). Este processo consiste na combinação de duas ou mais partes dos modelos 'pais' para gerar um novo modelo 'filho'. Os modelos 'pais' têm uma parte de seus termos (descritores) copiados. Estes termos são escolhidos de forma aleatória. Estes 'pedaços de modelo' (parte do modelo 'pai' e parte do modelo 'mãe') são conectados para criar um novo modelo 'filho'. O tamanho dos modelos é parcialmente determinado pela medida LOF. Esta medida pode ser controlada pelo fator de ajuste ("smoothing factor"). O fator de ajuste altera o equilíbrio entre o número de termos no modelo (descritores, GCODs) e a redução do erro de mínimos quadrados (LSE), além de controlar o 'overfitting' (super ajuste na reta de regressão). Não há maneira de saber antecipadamente quantos descritores (GCODs) deverão fazer parte de um modelo. Isto dependerá do sistema ligante-receptor. Geralmente, quanto maior e/ou mais flexível for o ligante maior será o número de descritores (GCODs). Entretanto, as regras da estatística impõem um limite de 4 a 5 observações (compostos) por descritor (HOPFINGER, 1999).

Neste estudo, foram utilizadas 10.000 (valor 'default') recombinações na análise dos GCODs gerados com células de 1 Å. A freqüência de mutação durante o ciclo de otimização de recombinação ("crossover") foi selecionada em 10% para cada nova geração de modelos ou equações de QSAR-4D e foram testados valores de fator de ajuste de 1 a 2,5. Foi estabelecido que os 10 melhores modelos seriam registrados para a análise de cada um dos sete alinhamentos.

Muitas 'medidas de diagnóstico' para analisar os modelos resultantes são determinadas como parte do processo GA de otimização. Estas medidas incluem: a) os descritores empregados na análise como função da operação de recombinação ("crossover"); b) correlação cruzada entre descritores e/ou medidas de atividade biológica; c) número de modelos significantes; d) coeficiente de correlação linear (r<sup>2</sup>); e) coeficiente de correlação quadrático da validação cruzada "leave-one-out" (q<sup>2</sup>) e d) fator de desajuste, LOF (ROGERS, 1991; ROGERS, HOPFINGER, 1994).

#### 3.3.9. IDENTIFICAÇÃO DOS MELHORES MODELOS (EQUAÇÕES) DE QSAR-3D

Esta etapa corresponde à inspeção e avaliação da população de modelos obtidos para cada um dos alinhamentos selecionados. O principal objetivo é identificar os melhores modelos de QSAR-3D considerando o respectivo alinhamento, proveniente da análise de QSAR-4D. Entretanto, este objetivo pode ser generalizado para permitir explorar e otimizar os modelos de QSAR-3D não somente quanto ao alinhamento, mas também quanto à amostragem conformacional, aos IPEs e outros parâmetros metodológicos (**3.3.3**). Além disso, o melhor modelo de QSAR-3D pode ser identificado utilizando-se uma ou mais medidas estatísticas de significância adicionadas de informações provenientes da comparação da família de melhores modelos (HOPFINGER, 1999).

Neste estudo, os 10 melhores modelos para cada alinhamento foram registrados. Os modelos foram classificados de acordo com seus valores de coeficiente de correlação da validação cruzada (xv-r<sup>2</sup> ou q<sup>2</sup>) (KUBINYI, 1993).

No programa 4D-QSAR 3.0 (HOPFINGER, 1999), a técnica de validação cruzada utilizada é denominada "leave-one-out". Esta técnica consiste na eliminação dos compostos do conjunto de treinamento, um de cada\_vez, aleatoriamente ou de forma sistemática, enquanto os coeficientes da equação são recalculados até que todos os compostos tenham sido eliminados pelo menos uma vez. A validação cruzada "leave-one-out" foi empregada para obter a medida da capacidade de predição interna dos modelos no conjunto de treinamento.

Para cada um dos 10 modelos selecionados, considerando-se o alinhamento testado, foram registradas as seguintes medidas estatísticas de significância: coeficiente de correlação linear (r<sup>2</sup>), coeficiente de correlação da validação cruzada (q<sup>2</sup>), erro de mínimos quadrados (LSE) e fator de desajuste (LOF).

O programa de 4D-QSAR 3.0 (HOPFINGER, 1999) transforma os valores de q<sup>2</sup> em valores ajustados de q<sup>2</sup> para que a comparação entre equações (modelos) com diferentes números de termos seja possível.

Além das medidas estatísticas mencionadas anteriormente, o programa de 4D-QSAR 3.0 (HOPFINGER, 1999) fornece a matriz de correlação linear dos resíduos (ou erros) de ajuste dos 10 melhores modelos e a matriz de correlação linear dos descritores (GCODs) para cada um dos 10 melhores modelos, de cada alinhamento testado.

A comparação dos coeficientes de correlação linear dos resíduos de ajuste entre pares de modelos, gerados nas análises de GFA, é ferramenta importante para organizar e explorar informações provenientes de múltiplos bons modelos, permitindo, por exemplo, a determinação de um subconjunto de modelos distintos em um conjunto de bons modelos. Os valores residuais correspondem à diferenca

44

entre os valores de atividade biológica observada ou experimental do conjunto de treinamento e os respectivos valores de atividade biológica calculada. Quando um par de modelos apresentar valores de resíduo de ajuste semelhantes ( $R \approx 1$ ), estes modelos serão considerados equivalentes. Um par de modelos é considerado distinto, ou seja, os modelos não fornecem as mesmas informações, quando os valores de resíduo de ajuste forem diferentes (R < 0.5) (ROGERS, HOPFINGER, 1994; HOPFINGER, 1999).

A análise da matriz de correlação linear dos GCODs de cada um dos 10 melhores modelos permite verificar se as variáveis independentes são ou não correlacionadas entre si, em um mesmo modelo e entre modelos distintos. Ou seja, permite constatar se cada variável (GCOD) fornece contribuição única para cada modelo selecionado (princípio da ortogonalidade das variáveis) (KUBINYI, 1993), ou se está fornecendo informação redundante ("overfitting", super ajuste na reta de regressão) (TOKARSKI, HOPFINGER, 1997b; HOPFINGER, 1999).

Neste estudo foram considerados como "outliers" (compostos desviantes) de um modelo os compostos cujas diferenças entre as atividades predita e experimental excederam em duas vezes o valor de desvio padrão (DP) calculado dos resíduos.

## 3.3.10. CONJUNTO DE AVALIAÇÃO - VALIDAÇÃO EXTERNA E PREDIÇÃO DOS MODELOS

Conforme já mencionado, os compostos do conjunto de avaliação não participaram da etapa de geração dos modelos, na análise de QSAR-4D (IR). Porém, foram utilizados para validar o(s) modelo(s) obtido(s) com o conjunto de treinamento e avaliar sua capacidade de predição (HOPFINGER, 1999).

Os valores de atividade biológica prevista (pMIC) foram calculados para cada composto do conjunto de avaliação por meio da substituição dos valores dos índices de coordenadas cartesianas gerados pela ocupação dos descritores nas células da grade virtual 3D (GCODs) para cada um dos compostos do conjunto de avaliação na equação, ou modelo, e seu respectivo alinhamento, resultantes da análise do conjunto de treinamento.

### 3.3.11. SELEÇÃO E VISUALIZAÇÃO DA CONFORMAÇÃO 'BIOATIVA'

Nesta etapa, a conformação 'bioativa' para cada composto do conjunto de treinamento é idealizada pelo programa de 4D-QSAR 3.0 (HOPFINGER 1999). A seleção da conformação 'bioativa' foi realizada empregando como critério a identificação de todos os estados conformacionais, para cada composto do conjunto treinamento, que estivessem na faixa de energia,  $\Delta E$ , da conformação de energia mínima global do CEP, obtido nas simulações de DM. O valor de  $\Delta E$  utilizado neste estudo foi de 5 kcal/mol (valor 'default'do programa de 4D-QSAR 3.0) (HOPFINGER, 1999).

As conformações de menor energia selecionadas foram avaliadas individualmente, utilizando o alinhamento e o respectivo modelo de QSAR-3D escolhidos, atribuindo para a ocupação das células da caixa virtual o valor zero, se o IPE em questão não ocupa a célula da caixa no modelo, ou o valor calculado de A<sub>o</sub>, J<sub>o</sub>, S<sub>o</sub>, se o IPE correto ocupa a célula da caixa virtual (HOPFINGER, 1999).

A conformação de menor energia, dentro da faixa de ΔE (5 kcal/mol), que apresentou a maior atividade prevista (calculada), considerando o modelo e o alinhamento escolhidos, foi selecionada como a conformação 'ativa' de cada composto do conjunto de treinamento, pelo programa de 4D-QSAR 3.0. As conformações postuladas como 'ativas' poderão ser utilizadas como protótipo em outras abordagens de CAMD ("Computer-Assisted Molecular Design", Planejamento Molecular Auxiliado por Computador) (HOPFINGER, 1999).

O modelo resultante da análise de QSAR-4D IR deve proporcionar ajuste estérico satisfatório e apresentar complementaridade na interação receptorligante. Esta verificação deve ser efetuada por meio da exploração das informações estruturais do modelo selecionado, considerando-se o alinhamento dos descritores (GCODs) e suas respectivas conformações 'ativas' previstas dentro do sítio de interação (HOPFINGER *et al.*, 1997; ALBUQUERQUE *et al.*, 1998; VENTAKATARANGAN, HOPFINGER, 1999).

Neste estudo, a análise dos descritores (GCODs) no sítio ativo da enzima cristalizada **1zid** foi realizada por meio da comparação das informações estruturais entre as conformações 'ativas' previstas pelo modelo e alinhamento selecionados com o ligante co-cristalizado (INH). As conformações 'ativas' previstas de cada um dos compostos do conjunto de treinamento e seus respectivos descritores foram ancorados no sítio ativo da enzima **1zid** utilizando-se como referência o alinhamento e modelo selecionados na análise QSAR-4D IR. Distâncias entre resíduos de aminoácido do sítio ativo, importantes no processo de interação, e as estruturas das conformações 'ativas' previstas + GCODs e do ligante cristalizado foram utilizadas como objeto desta comparação.

Na tabela II estão apresentadas as etapas de uma análise de QSAR-4D e no fluxograma 1 está representado o método de análise de todos os dados no formalismo QSAR-4D, originando modelos QSAR-3D finais (HOPFINGER, 1999).

Etapa	Descrição
1	Seleção e construção dos modelos 3D dos compostos que serão analisados;
2	Amostragem do conjunto conformacional de cada composto para gerar o CEP (simulações
	de DM);
3	Seleção das classes de elementos de interação farmacofórica, IPEs, para a grade ou caixa
	3D virtual;
4	Seleção dos alinhamentos a serem testados;
5	Formação da grade de referência e registro do perfil de ocupação das células da grade
	para cada IPE, de acordo com a medida de ocupação escolhida e com o alinhamento
	selecionado;
6	Análise de PLS - redução de dados, considerando o conjunto total de GCODs (descritores);

Tabela II – Seqüência de etapas de uma análise de QSAR-4D

Tabela II (continuação) - Seqüência de etapas de uma análise de QSAR-4D

Etapa	Descrição		
7	Utilização dos GCODs, que foram melhor classificados na análise de PLS, e quaisquer outros		
	descritores selecionados pelo usuário para o conjunto inicial de dados da análise de GA;		
8	Retornar à etapa 4 e repetir as etapas 4-7, a menos que todos os alinhamentos selecionados		
	já tenham sido incluídos na análise;		
9	Identificação dos melhores modelos de QSAR-3D de acordo com os alinhamentos testados;		
10	Visualização da conformação 'ativa' de cada composto do conjunto de treinamento, de		
	acordo com o modelo e o alinhamento escolhidos.		

Fluxograma 1 - Representação do método de análise de todos os dados no formalismo

QSAR-4D, originando modelos QSAR-3D finais	
--	--

Conjunto de descritores (GCODs)					
IPE	x y z = 1 1 1	112	113		
1	GCOD(1111)	GCOD(1 1 2 1)	GCOD(1 1 3 1)		
2	GCOD(1112)	GCOD(1 1 2 2)	GCOD(1132)		
3	GCOD(1113)	GCOD(1123)	GCOD(1133)		
				•••	

Redução de dados pela variância do GCOD e/ou pela relação individual do GCOD com a atividade biológica

GCOD(x1, y1, Z1, IPE1), GCOD(x2, y2, Z2, IPE2),..., GCOD(xn, yn, Zn, IPEn)

Redução de dados por PLS

GCOD(x1', y1', Z1', IPE1'), GCOD(X2', Y2', Z2', IPE2'),..., GCOD(Xn', Yn', Zn', IPEn')

Construção dos modelos por GA



KERLY FERNANDA MESQUITA PASQUALOTO
3.4. METODOLOGIA DE FEFF ("FREE ENERGY FORCE FIELD", campo de força de energia livre) QSAR-3D (DR, Dependente do Receptor)

#### 3.4.1. OBTENÇÃO E OTIMIZAÇÃO DO MODELO DA ENZIMA (MODELO DO RECEPTOR)

A estrutura da enzima **1zid** (ROZWARSKI *et al.*, 1998) apresenta 4095 átomos, incluindo os átomos de hidrogênio, que foram adicionados, em uma primeira etapa, utilizando-se a opção existente no programa HyperChem 6.0 (Hypercube Inc., 1996). Então, aos resíduos Arg e Lys foram atribuídas cargas +1 e aos resíduos Glu e Asp, cargas -1. Os "Ione pairs" (pares de elétrons desemparelhados) foram excluídos da estrutura da enzima, por meio de inspeção visual. As moléculas de água localizadas na estrutura do complexo cristalizado não foram incluídas no modelo. Cargas parciais AMBER (WEINER, KOLLMAN, NGUYEN, 1986; CORNELL *et al.*, 1995) foram atribuídas a todos os átomos da estrutura da enzima (TOKARSKI, HOPFINGER, 1997a). A enzima apresentou carga total igual a 1 (HyperChem 6.0 – Hypercube Inc., 1996).

O tratamento teórico de uma estrutura protéica deste porte, considerandose todos os átomos e interações, envolveria alto custo em termos de tempo computacional, o que justificaria estudos com a utilização de modelos de tamanho reduzido de biomacromoléculas (TOKARSKI, HOPFINGER, 1997a).

Neste estudo, foi utilizada a técnica de redução do tamanho ou de corte ("pruning") da estrutura da enzima, descrita por Tokarski e Hopfinger (1997a, b), com a finalidade de torná-la acessível em termos de tempo computacional para as simulações de DM. A análise foi restrita aos resíduos de aminoácidos da enzima próximos à região do sítio ativo.

Após assinalar as cargas parciais da enzima, utilizando o campo de força AMBER (WEINER, KOLLMAN, NGUYEN, 1986), foram construídos modelos com esferas de raios de 12 e 14 Å, utilizando-se como referência o centro de massa do complexo formado pelo cofator (NAD) e o fármaco (INH), da estrutura original obtida por cristalografia de raios-X (**1zid**), no programa HyperChem 6.0 (Hypercube Inc., 1996). Somente os átomos contidos em cada esfera foram incluídos nos modelos e o restante da enzima foi negligenciado. Os resíduos da enzima que possuíam pelo menos um átomo de hidrogênio desconectado dentro das esferas de corte foram incluídos nos correspondentes modelos (TOKARSKI, HOPFINGER, 1997a).

O processo de redução de tamanho da enzima resulta em modelos formados por conjuntos de 'pedaços de resíduos peptídicos' desconectados. Os 'pedaços de resíduos peptídicos' separados por menos de cinco resíduos de aminoácidos foram 'conectados' pela inclusão dos resíduos de aminoácidos que estavam faltando na seqüência da cadeia principal (esqueleto peptídico) dos modelos. A terminação dos 'pedaços de resíduos peptídicos' foi efetuada pela adição de grupos metila (carga = zero). Este procedimento tem como finalidade manter a integridade geométrica local do modelo do sítio ativo da enzima para o processo de corte (TOKARSKI, HOPFINGER, 1997a, b). A esfera de 12 Å não incluiu todos os resíduos de aminoácidos do sítio ativo da enzima, então, o modelo de 12 Å foi descartado.

Neste estudo, foi reduzido o tamanho do receptor, InhA, de 4095 para 1912 átomos, desconsiderando-se todos os resíduos fora de uma esfera de 14 Å. A otimização do modelo reduzido da enzima foi realizada utilizando os métodos "steepest descent" e "conjugate gradients", do programa MOLSIM (DOHERTY, 1994), com um critério de convergência de energia de 5 kcal/mol, para relaxar tensões intra e intermoleculares provenientes do empacotamento cristalográfico. Utilizou-se a mesma T do ensaio biológico (310 K) e constante dielétrica de 3,5 (do vácuo), considerada realística para a representação de interações de curto alcance (entre átomos ou grupos de átomos próximos espacialmente, mas não conectados). Foi utilizado um limite de até 10 Å para o cálculo das interações não ligadas.

O modelo reduzido da enzima (14 Å) contém grupos peptídicos desconectados, que poderiam 'vagar' pelo espaço durante os processos de minimização e de simulações de DM. A atribuição de massas fictícias elevadas aos átomos do modelo reduzido da enzima produziria um 'momentum reservoirs' (estado de congelamento), que manteria os átomos do receptor próximos de suas posições iniciais referentes à estrutura completa original, mas permitiria alguma flexibilidade de posicionamento para acomodar um ligante, por exemplo. O uso de massas fictícias é virtualmente o mesmo que usar restrições cartesianas, particularmente quando as massas selecionadas são muito grandes. Além disso, a variação dos valores de massa fictícia constitui um modo conveniente de ajustar os movimentos do modelo reduzido, durante as simulações de DM, permitindo a reprodução aproximada dos movimentos da enzima completa original (TOKARSKI, HOPFINGER, 1997b). Assim, aos átomos da cadeia principal peptídica do modelo reduzido foram conferidas massas fictícias de 2500 a 5000 u.m.a., para impedir afastamentos significativos da geometria do modelo devido à exclusão do resto da enzima.

A estabilidade estrutural do modelo reduzido da enzima em relação à estrutura cristalográfica original (completa) foi avaliada quanto ao desvio da raiz dos mínimos quadrados ("root mean square deviation", r.m.s.), no programa HyperChem 6.0 (Hypercube Inc., 1996), pela sobreposição das duas estruturas.

### 3.4.2. DINÂMICA MOLECULAR DO MODELO REDUZIDO DA ENZIMA

Na metodologia de FEFF ("free energy force field", campo de força de energia livre), simulações de DM em várias temperaturas são processadas com a finalidade de se obter uma amostragem do espaço energético dos estados ligado e não-ligado do sistema ligante-receptor (TOKARSKI, HOPFINGER, 1997b). Tendo como finalidade, neste estudo, obter-se um complexo aceitável do ponto de vista estérico, a estrutura minimizada do modelo reduzido da enzima foi aquecida lentamente ('warming up'). Partiu-se de uma simulação de DM de 10 ps (10.000 passos, cada passo de 1 fs, com a trajetória dos estados conformacionais registrada a cada 20 passos) à T de 25 K e a conformação de menor energia desta simulação foi selecionada. A conformação de menor energia do modelo reduzido à T de 25 K foi utilizada para desenvolver uma simulação do DM de 10 ps à T de 50 K e a conformação de menor energia de 50 K e a conformação de DM de 10 ps à T de 50 K e a conformação de conformação de menor energia de 50 K foi utilizada para desenvolver uma simulação foi selecionada. Este procedimento foi continuado até 300 K, com incrementos de 50 K. A conformação de menor energia do modelo reduzido selecionada na simulação de DM à T de 300 K foi utilizada para desenvolver uma simulação de DM de 10 ps à T de 310 K, temperatura do ensaio biológico do conjunto de treinamento.

Os procedimentos de DM foram realizados utilizando o programa MOLSIM (DOHERTY, 1994) e as conformações de menor energia foram selecionadas pelo programa OSFRAME (The Chem21 Group, Inc.). A seleção da conformação de menor energia, representativa do equilíbrio, considera o valor da energia potencial total do complexo.

Nas simulações de DM realizadas utilizou-se um banho térmico (BERENDSEN et al., 1984) com tempo de relaxação de 0,1 ps, constante dielétrica de 3,5 e um limite de até 10 Å para o cálculo das interações não ligadas.

O modelo da camada de hidratação ("hydration shell") proposto por Hopfinger (1973) (FORSYTHE, HOPFINGER, 1973) foi incluído na representação do campo de força na análise de FEFF para estimar a energia de solvatação de cada estrutura de energia mínima resultante do processo de aquecimento ("warming up"). A inclusão deste modelo é importante porque considera a competição das interações receptor-ligante com as moléculas de água. As simulações de DM de sistemas solvatados representam elevado custo computacional. Portanto, neste estudo, as energias de solvatação foram calculadas somente para as conformações de energia mínima, utilizando o programa MOLSIM (DOHERTY, 1994).

## 3.4.3. DOCKING DO LIGANTE MAIS VOLUMOSO

A conformação de menor energia do modelo da enzima reduzida selecionada na simulação de DM à T de 310 K foi utilizada para o ancoramento ("docking") do ligante (composto:NAD) mais volumoso do conjunto de treinamento, INHd19, proveniente da análise de QSAR-4D (conformação "bioativa"), considerando-se o alinhamento e o modelo escolhidos.

O ligante INHd19 (composto:NAD), resultante da análise de QSAR-4D, foi ancorado no sítio ativo da enzima (modelo reduzido) por sobreposição com a estrutura do isonicotínico-acil:NAD complexada no modelo reduzido da enzima. Foi utilizada a seqüência de três átomos ordenados do alinhamento escolhido na análise de QSAR-4D para o ancoramento do ligante no sítio ativo. Após a sobreposição, a estrutura do isonicotínico-acil:NAD foi eliminada do sítio ativo do modelo reduzido da enzima. Este procedimento foi efetuado no programa HyperChem 6.0 (Hypercube Inc., 1996).

O processo de ancoramento foi realizado com o ligante mais volumoso para assegurar a capacidade do modelo reduzido da enzima de conter os demais ligantes do conjunto de treinamento.

A estrutura do modelo reduzido da enzima contendo o ligante INHd19 foi otimizada no programa MOLSIM (DOHERTY, 1994), utilizando as mesmas atribuições descritas para a minimização do modelo reduzido (**3.4.1**.).

# 3.4.4. DINÂMICA MOLECULAR DO COMPLEXO ENZIMA-LIGANTE MAIS VOLUMOSO

A estrutura otimizada do modelo reduzido da enzima contendo o ligante INHd19 foi submetida ao processo de aquecimento lento ("warming up") (item **3.4.2.**). As conformações de menor energia de cada simulação de DM em cada T foram selecionadas e suas energias de solvatação foram obtidas, conforme descrito no item **3.4.2.** 

O processo de aquecimento lento permite a obtenção de parâmetros termodinâmicos de interação para cada T de simulação e, por conseqüência, modelos de QSAR para cada T (TOKARSKI, HOPFINGER, 1997b). Neste estudo, os modelos de QSAR foram obtidos com os parâmetros termodinâmicos resultantes à T de 310 K.

O processo de "warming up" apresenta como vantagem a possibilidade de se realizar uma varredura do espaço conformacional do complexo enzima-ligante, da enzima e do ligante. Em temperaturas mais elevadas, o ligante complexado possui liberdade suficiente para procurar seu alinhamento intermolecular (encaixe) ideal dentro do sítio ativo (IOKARSKI, HOPFINGER, 1997b).

A amostragem em diferentes temperaturas permite que as energias cinética e potencial dos sistemas envolvidos no processo de interação sejam equilibradas uma em relação à outra. Além disso, a análise estatística destas energias permite a avaliação da temperatura de simulação em que os parâmetros termodinâmicos de ligação estejam mais bem correlacionados com os valores experimentais (TOKARSKI, HOPFINGER, 1997b).

3.4.5. DOCKING DOS LIGANTES DO CONJUNTO DE TREINAMENTO E DM DOS COMPLEXOS ENZIMA-LIGANTE

A conformação de menor energia do modelo reduzido da enzima contendo o ligante INHd19 selecionada na simulação de DM à T de 310 K foi utilizada para o ancoramento ("docking") dos demais ligantes (composto:NAD) do conjunto de treinamento, provenientes da análise de QSAR-4D (conformação "ativa" prevista), considerando-se o alinhamento e o modelo escolhidos.

O ancoramento dos ligantes no sítio ativo da enzima foi realizado no programa HyperChem 6.0 (Hypercube Inc., 1996) pela sobreposição das estruturas

dos ligantes do conjunto de treinamento, provenientes da análise de QSAR-4D, com o ligante INHd19, presente no sítio ativo do modelo reduzido da enzima, utilizando a seqüência de três átomos ordenados correspondente ao alinhamento escolhido na análise de QSAR-4D. Após a sobreposição de cada ligante, o ligante INHd19 foi eliminado do sítio ativo do modelo da enzima. Este procedimento proporcionou a obtenção dos complexos enzima-ligante.

Os complexos enzima-ligante foram otimizados, no programa MOLSIM (DOHERTY, 1994), utilizando-se as mesmas atribuições descritas para a minimização do modelo reduzido da enzima (3.4.2.). As estruturas otimizadas dos complexos enzima-ligante foram submetidas a simulações de DM de 10 ps à T de 310 K e as energias de solvatação das conformações de menor energia foram calculadas. Neste estudo, o procedimento de "warming up" efetuado com o complexo enzima-ligante mais volumoso (3.4.4.) consistiu em uma alternativa de economia de tempo computacional para os demais ligantes do conjunto de treinamento, permitindo que as simulações de DM fossem realizadas somente na T do ensaio biológico.

Com a finalidade de incluir um descritor no campo de amostragem energético que permitisse mais explicações sobre o mecanismo de ação dos compostos deste estudo, a ligação covalente entre o carbono carbonílico dos ligantes e o C4 do anel nicotinamida do cofator (composto:NAD) foi cindida (HyperChem 6.0). Os complexos enzima-ligante com ligação covalente cindida foram otimizados no programa MOLSIM (DOHERTY, 1994), utilizando-se as mesmas atribuições descritas para a minimização do modelo reduzido da enzima (**3.4.2.**). As estruturas otimizadas dos complexos enzima-ligante com ligação covalente cindida foram submetidas a simulações de DM de 10 ps à T de 310 K e as energias de solvatação das conformações de menor energia foram calculadas. Neste caso, a conformação de menor energia selecionada na simulação de DM à T de 310 K de cada complexo foi utilizada para medir a distância entre os átomos envolvidos na ligação covalente (HyperChem 6.0) e esta distância (d<sub>Lcov</sub>) foi incluída como um descritor a mais no processo de obtenção dos modelos de FEFF QSAR-3D.

### 3.4.6. DINÂMICA MOLECULAR DOS LIGANTES E DA ENZIMA NÃO-LIGADOS

As estruturas otimizadas dos ligantes provenientes da análise de QSAR-4D IR (modelo e alinhamento escolhidos) foram submetidas a simulações de DM de 50 ps (50.000 passos, cada passo de 1 fs e trajetória dos estados conformacionais registrada a cada 20 passos) à T de 310 K. Utilizou-se um banho térmico (BERENDSEN *et al.*, 1984) com tempo de relaxação de 0,1 ps, constante dielétrica de 3,5 e um limite de até 10 Å para o cálculo das interações não ligadas. As conformações de menor energia foram selecionadas e suas energias de solvatação foram calculadas.

A conformação de menor energia selecionada à T de 310 K no processo de aquecimento lento do modelo reduzido da enzima foi utilizada para amostrar as energias da enzima no estado não-ligado. O complexo isonicotínico-acil:NAD foi excluído do sítio ativo e, após otimização, a estrutura do modelo reduzido sem o inibidor foi empregada como estrutura inicial de uma simulação de DM de 10 ps à T de 310 K. Utilizou-se um banho térmico (BERENDSEN *et al.*, 1984) com tempo de relaxação de 0,1 ps, constante dielétrica de 3,5 e um limite de até 10 Å para o cálculo das interações não-ligadas. Foram atribuídas massas fictícias de 5000 u.m.a aos átomos da cadeia peptídica principal do modelo da enzima. A conformação de menor energia foi selecionada e sua energia de solvatação foi calculada.

Neste estudo, os procedimentos de DM foram realizados utilizando o programa MOLSIM (DOHERTY, 1994) e as conformações de menor energia foram selecionadas pelo programa OSFRAME (The Chem21 Group, Inc.). O programa INTERCALC (The Chem21 Group, Inc.) foi utilizado para obter as energias intra e intermoleculares dos complexos enzima-ligante.

### 3.4.7. CONSTRUÇÃO DOS MODELOS DE FEFF QSAR-3D

Na metodologia de FEFF QSAR-3D (DR), os termos de energia intra e intermoleculares, obtidos a partir das simulações de DM, são considerados os descritores (variáveis independentes) para a construção dos modelos de QSAR-3D e predição das propriedades termodinâmicas de interação (TOKARSKI, HOPFINGER, 1997b). Neste estudo, além dos descritores de termos de energia, definidos na Tabela III, outros descritores, não determinados diretamente dos termos do campo energético, foram incluídos como possíveis características dos modelos FEFF QSAR-3D. Estes descritores consideram a distância entre os átomos envolvidos na ligação covalente do ligante (3.4.5.) e as seguintes propriedades do ligante não complexado à enzima: energia de HOMO ("highest occupied molecular orbital"; energia do orbital molecular ocupado de maior energia - método hamiltoniano AM1), energia de LUMO ("lowest unoccupied molecular orbital"; energia do orbital molecular desocupado de menor energia – método hamiltoniano AM1), momento de dipolo, ClogP, volume e área superficial (final da Tabela III). As variáveis dependentes são as propriedades de interação experimentais (TOKARSKI, HOPFINGER, 1997b). Neste estudo, as variáveis dependentes constituem os valores de pMIC do conjunto de treinamento.

A entropia vibracional, associada a ligações e ângulos de ligação, não foi calculada neste estudo.

A maioria dos sistemas receptor-ligante apresenta dois problemas inerentes relacionados ao desenvolvimento da abordagem estatística na avaliação de uma análise FEFF QSAR-3D. O primeiro refere-se ao número de análogos (componentes do conjunto treinamento), *m* (observações), ser usualmente pequeno quando comparado ao número de possíveis variáveis independentes, *n* (termos de energia FEFF que podem ser classificados). O segundo trata-se do fato de que muitos dos termos de energia (variáveis independentes) são inter-relacionados. Para minimizar

estes problemas os métodos de PLS ("partial least squares"), para a redução do número de descritores, e de GFA ("genetic function approximation"), para a otimização dos modelos, foram incluídos na avaliação estatística de uma análise de FEFF QSAR-3D (TOKARSKI, HOPFINGER, 1997b).

Neste estudo, os descritores, ou termos de energia do campo de força FEFF, foram submetidos à análise estatística pela metodologia de GFA/PLS, implementada no programa WOLF 5.5 (ROGERS, 1994). A otimização dos modelos foi realizada com a análise de GFA (ROGERS, HOPFINGER, 1994). A combinação da análise de mínimos quadrados parciais, PLS (GLEN, DUNN, SCOIT, 1989), e regressão linear multidimensional foi utilizada para ajustar o número de valores das medidas experimentais (pMIC) aos respectivos descritores nos modelos de FEFF QSAR-3D envolvidos na otimização por GFA. Na etapa de filtragem dos dados, PLS, o número de componentes na análise foi variado de 4 a 6. Na etapa de construção e atimização dos modelos por algoritmo genético, os números de recombinações ("crossovers") testados foram de 25.000 a 70.000 e as taxas de probabilidade de mutação testadas foram de 5 % e 10 %. O critério para convergência foi a inspeção dos valores estatísticos e do número de descritores dos modelos gerados. Foram testados valores de fator de ajuste ("smoothing factor"), filtro do PLS, de 0,1 a 1,5.

Tabela III - Definição dos termos do campo de força de energia livre (FEFF) utilizados na construção dos modelos de FEFF QSAR-3D

FEFF	Definição
$\Delta E_{est} = E_{est RL} - E_{est R} - E_{est L}$	Variação na energia de deformação axial das ligações na complexação
∆Edef = EdefRL - EdefR - EdefL	Variação na energia de deformação angular na complexação
$\Delta E_{tors} = E_{tors  RL} - E_{tors  R} - E_{tors  L}$	Variação na energia torsional na complexação
AEvaw = Evaw RL - Evaw R - Evaw L	Variação na energia de van der Waals na complexação
AEet = Eet LR - Eet R - Eet L	Variação na energia eletrostática (Coulomb) na complexação
AEELA = EELALR - EELAR - EELAL	Variação na energia de interação do tipo 1,4 na complexação
ΔEHD = EHD LR - EHD R - EHD L	Variação na energia de ligação de hidrogênio na complexação

# KERLY FERNANDA MESQUITA PASQUALOTO

Tabela III (continuação) - Definição dos termos do campo de força de energia livre (FEFF)

FEFF	Definição
$\Delta E_{solv} = E_{solv LR} - E_{solv R} - E_{solv L}$	Variação na energia de solvatação na complexação
∆Eest+det = Eest+det LR - Eesi+det R -	Variação na soma das energias de deformação axial e angular na
Eest+def L	complexação
$\Delta E_{est+def+1ars} = E_{est+def+1ars LR} - E_{est+def+1ars R}$	Variação na soma das energias de deformação axial, angular e
- Eest+def+tors L	torsional na complexação
ΔEel+Hb = Eel+Hb LR - Eel+Hb R - Eel+Hb L	Variação na soma das energias eletrostática e de ligação de
	hidrogênio na complexação
$\Delta E_{\text{el+Hb+E}_{14}} = E_{\text{el+Hb+E}_{14} \text{ LR}} - E_{\text{el+Hb+E}_{14} \text{ R}} -$	Varlação na soma das energias eletrostática, de ligação de
Eel+Ho+Eu+L	hidrogênio e de interação do tipo 1,4 na complexação
ELR(LL,RR,LR)	Energia do complexo ligante-receptor
E <sub>LR</sub> (LR)	Energia intermolecular total do complexo ligante-receptor
E <sub>LR, vdw</sub>	Energia de van der Waals intermolecular do complexo ligante- receptor
Funct	Energia eletrostática intermolecular do complexo ligante-receptor
Figure .	Eneraia de ligação de hidrogênio intermoleculares do complexo
elx,nu	ligante-receptor
$E_{LR,el+Hb} = E_{LR,el} + E_{LR,Hb}$	Soma das energias eletrostática e de ligação de hidrogênio
	intermoleculares do complexo ligante-receptor
$E_{LR,el+Hb+vdW} = E_{LR,el} + E_{LR,Hb} + E_{LR,vdW}$	Soma das energias intermoleculares eletrostática, de ligação de
	hidrogênio e de van der Waals do complexo ligante-receptor
$\Delta E_{L}(LL) = E_{LR}(LL) - E_{L}(LL)$	Variação na energia intramolecular do ligante na complexação
Eur(LL)	Energia intramolecular do ligante no complexo
EL(LL)	Energia intramolecular do ligante não-complexado
$\Delta E_{R}(RR) = E_{LR}(RR) - E_{R}(RR)$	Variação na energia intramolecular do receptor na complexação
E <sub>LR</sub> (RR)	Energia intramolecular do receptor no complexa
E <sub>R</sub> (RR)	Energia intramolecular do receptor não-complexado
ELR(LRM)	Energia de solvatação dom complexo ligante-receptor
$\Delta E_{L}(LM) = E_{LR}(LM) - E_{L}(LM)$	Variação na energia de solvatação do ligante na complexação
ELR(LM)	Energia de solvatação do ligante no complexo
EL(LM)	Energia de solvatação do ligante não-complexado
$\Delta E_{R}(RM) \cong E_{LR}(RM) - E_{R}(RM)$	Variação na energia de solvatação do receptor na complexação
ELR(RM)	Energia de solvatação do receptor no complexo
E <sub>R</sub> (RM)	Energia de solvatação do receptor não-complexado
dree	Distância entre os átomos que formam a ligação covalente (ligante:
	cofator) no complexo
Energia de HOMO	Energia do orbital molecular ocupado de maior energia do ligante
	não-complexado
Energia de LUMO	Energia do orbital molecular desocupado de menor energia do
	ligante não-complexado
Volume	Volume do ligante não-complexado

.

Tabela III (continuação) - Definição dos termos do campo de força de energia livre (FEFF) utilizados na construção dos modelos de FEFF QSAR-3D

FEFF	Definição
ClogP	logP calculado do ligante não-complexado
Momento de dipolo	Momento dipolar do ligante não-complexado
Superfície_A	Área superficial do ligante não-complexado

LR = complexo ligante-receptor; L = ligante; R = receptor; M = meio de solvatação (H2O)

## 3.4.8. IDENTIFICAÇÃO DOS MELHORES MODELOS DE FEFF QSAR-3D

Como na análise de QSAR-4D (IR), os modelos de FEFF QSAR-3D devem ser avaliados pelas medidas estatísticas de ajuste, que incluem o coeficiente de correlação linear (r<sup>2</sup>), o coeficiente de correlação da validação cruzada "leaveone-out" (xv-r<sup>2</sup> ou q<sup>2</sup>), erro dos mínimos quadrados (LSE), matriz de correlação linear das variáveis independentes (descritores) e fator de desajuste de Friedman (LOF) (ROGERS, 1991; ROGERS, HOPFINGER, 1994). Estas variáveis são calculadas para testar a robustez dos modelos. A matriz de correlação linear dos descritores deve ser examinada para eliminar modelos em que pares de termos de energia tenham coeficientes de correlação linear maiores que 0,5 ("overfitting") (TOKARSKI, HOPFINGER, 1997b).

Neste estudo, os 5 melhores modelos de FEFF QSAR-3D foram selecionados, considerando-se as medidas estatísticas mencionadas.

Com a finalidade de verificar se os 5 melhores modelos selecionados de FEFF QSAR-3D apresentavam informações comuns ou distintas, a matriz de correlação linear dos resíduos de ajuste (diferença entre as atividades predita e experimental) entre todos os pares dos 5 melhores modelos foi computada.

Modelos serão considerados equivalentes quando apresentarem distribuição muito semelhante dos resíduos de ajuste e com valores de coeficientes de correlação elevados, ao passo que modelos serão considerados diferentes quando apresentarem padrões pobremente correlacionados de seus resíduos de

61

ajuste e com baixos valores de coeficientes de correlação. Este é o diagnóstico para determinar subconjuntos de modelos distintos entre um conjunto de bons modelos gerados na análise de GFA (ROGERS, HOPFINGER, 1994; ROGERS, 1997; HOPFINGER, 1999).

Como na análise dos modelos de QSAR-4D, foram considerados como "outliers" (compostos desviantes) de um modelo de FEFF QSAR-3D os compostos cujas diferenças entre as atividades prevista (calculada) e experimental excederam em duas vezes o valor de desvio padrão (DP) calculado dos resíduos.

### 3.4.9. CONJUNTO DE AVALIAÇÃO - VALIDAÇÃO E PREDIÇÃO DO MODELO

Da mesma forma que na análise de QSAR-4D (IR), a análise de FEFF QSAR-3D utiliza os compostos do conjunto de avaliação para validar o modelo selecionado e avaliar sua capacidade de predição.

Os procedimentos de DM relatados para os ligantes do conjunto de treinamento também foram realizados com os ligantes do conjunto de avaliação, considerando os estados ligado e não-ligado do complexo enzima-ligante, para a obtenção dos descritores do campo de força de energia livre considerados na análise de FEFF QSAR-3D.

Os fluxogramas 2 e 3 apresentados a seguir contêm os passos para obtenção dos descritores do campo de força de energia livre na metodologia de FEFF QSAR-3D.

	62
METODOLOGIA	

Estrutura otimizada do modelo reduzido da enzima (14 Å) DM 10 ps a 25 K Conformação de E<sub>min</sub> 310 K Conformação de energia mínima (E<sub>min</sub>) 25 K DM 10 ps a 50 K DM 10 ps a 310 K Conformação de Emin 50 K Conformação de Emin 300 K DM 10 ps a 300 K DM 10 ps a 100 K Conformação de E<sub>min</sub> 100 K Conformação de Emin 250 K DM 10 ps a 150 K DM 10 ps a 250 K Conformação de E<sub>min</sub> 150 K Conformação de Emin 200 K DM 10 ps a 200 K Cálculo das energias de solvatação Conformações de energia mínima 

Fluxograma 2 – Processo de "warming up" realizado com o modelo reduzido da enzima

Fluxograma 3 – Processo de "warming up" realizado com o complexo enzima-ligante INHd19



(ligante mais volumoso do conjunto de treinamento)

KERLY FERNANDA MESQUITA PASQUALOTO

Resultados e Discussão

.

.

.

.

.

· ·

.

65

### 4. RESULTADOS E DISCUSSÃO

### 4.1. ANÁLISE DE QSAR-4D

A abordagem QSAR-4D pode ser aplicada nas duas situações: dependente do receptor (DR) e independente do receptor (IR). Na primeira, a geometria do receptor (alvo molecular, geralmente uma enzima) está disponível. Em contraste, na segunda, a geometria do receptor não é parte dos dados disponíveis para desenvolver a análise (HOPFINGER *et al.*, 1997). Neste estudo, embora a estrutura cristalizada da enzima alvo (**1zid**) estivesse disponível, a metodologia de QSAR-4D independente do receptor (IR) foi aplicada porque o modo de interação dos ligantes com a biomacromolécula alvo não está elucidado. Assim, a geometria da enzima alvo foi negligenciada.

A análise de QSAR-4D tem sido extensivamente validada e aplicada com sucesso a um número de diferentes conjuntos de treinamento: inibidores benzilpirimidínicos de diidrofolato redutase (HOPFINGER *et al.*, 1997), análogos abortivos de prostaglandina PGF<sub>2</sub>α (HOPFINGER *et al.*, 1997), inibidores de transcriptase reversa (RT) HIV-1 (HOPFINGER *et al.*, 1997), antagonistas do receptor A<sub>2</sub> de tromboxana (ALBUQUERQUE *et al.*, 1998), inibidores de glicogênio fosforilase (VENKATARANGAN, HOPFINGER, 1999), inibidores de diidrofolato redutase do *Plasmodium falciparum* (SANTOS-FILHO, HOPFINGER, 2001), análogos do propofol – receptor GABA (A) (KRASOWSKI *et al.*, 2002), inibidores não-peptídicos de HIV protease (SANTOS-FILHO, HOPFINGER, 2002). Todas estas aplicações são exemplos de estudos de QSAR-4D independente do receptor (IR).

Neste estudo, o formalismo QSAR-4D IR foi aplicado para um conjunto de hidrazidas, análogas à isoniazida. Foi considerada a hipótese de que os compostos estudados agiriam da mesma forma que o protótipo, conforme descrito por Rozwarski e colaboradores (1998). Estes pesquisadores relataram que o mecanismo de ação da isoniazida envolve a formação de ligação covalente entre a forma E de Li O TECA ativa do fármaco (supostamente um radical isonicotínico-acil ou ânion) e o anel nicotinamida do cofator NAD (carbono 4 do anel) no sítio ativo da enzima enoilacp redutase do *M. tuberculosis* (InhA), criando uma conexão com o cofator NAD (aduto), que funciona como um inibidor fortemente ligado (Quadro 4).

A análise de QSAR-4D considera o farmacóforo, a liberdade conformacional e de alinhamento no desenvolvimento de modelos de QSAR-3D para os dados de estrutura química e atividade biológica do conjunto de treinamento avaliando a média deste conjunto em relação ao tempo, a quarta "dimensão" (HOPFINGER et al., 1997).

Os dez melhores modelos para cada um dos sete alinhamentos, resultantes da análise de QSAR-4D, foram avaliados e os resultados estão apresentados na Tabela IV.

A Tabela IV inclui o número de descritores (GCODs) e as medidas estatísticas, como valores de r<sup>2</sup> (coeficiente de regressão linear) e q<sup>2</sup> (coeficiente de correlação de validação cruzada), para os dez melhores modelos de 4D-QSAR, considerando cada alinhamento selecionado e um fator de ajuste ("smoothing factor") igual a 2,5. O número de outliers (compostos desviantes) em cada modelo também está relacionado nesta tabela.

•• <b>•</b> ••				· ·		
Alinhamento	№ de GCODs	r²	q²	outlier(s)	LSE	LOF
1	3 - 5	0,54 - 0,65	0,41 - 0,50	0 – 2	0,47 0,61	1,10 - 1,58
2	3 – 7	0,62 - 0,71	0,50 - 0,58	1-3	0,39 - 0,51	0,90 2,56
З	3-6	0,50 - 0,70	0,35 - 0,42	0-3	0,56 - 0,67	1,11 - 1,75
4	4 - 7	0,80 - 0,88	0,71 – 0,80	0 - 2	0,16 - 0,27	0,53 - 0,98
5	5 - 6	0,76 - 0,81	0,65 - 0,70	0-4	0,25 - 0,32	0,80 - 1,21
6	3 – 5	0,46 - 0,55	0,29 - 0,40	0-2	0,61 - 0,74	1,16 - 1,89
7	3 - 6	0,54 - 0,71	0,38 - 0,55	0-2	0,39 - 0,62	1,02 - 2,27

Tabela IV – Medidas estatísticas e número de descritores encontrados para os dez melhores modelos de cada um dos alinhamentos testados

De acordo com a Tabela IV, os modelos obtidos com o fator de ajuste de 2,5 apresentaram valores elevados de q<sup>2</sup> e r<sup>2</sup>, mas nenhum super ajuste da reta de regressão ("overfitting"), comprovado pelas matrizes de correlação linear das variáveis independentes de cada modelo de cada um dos alinhamentos testados.

O alinhamento 4, que abrange as porções acil e NAD da estrutura dos ligantes investigados (3.3.4.), foi considerado o mais indicado à obtenção de melhores modelos, de acordo com os valores do coeficiente de correlação de validação cruzada, q<sup>2</sup>, ou seja, os modelos do alinhamento 4 apresentaram maiores valores de q<sup>2</sup>. Além disso, os modelos do alinhamento 4 apresentaram maiores valores de r<sup>2</sup> e menores valores de LSE e LOF.

Na Tabela V estão apresentados os resultados dos dez melhores modelos de QSAR-4D construídos para o alinhamento 4.

Tabela V – Medidas estatísticas, número de descritores e número de outliers para os dez melhores modelos do alinhamento 4

Modelos	Nº de GCODs	r²		LSE	LOF	outlier(s)
1	6	0,88	0,80	0,16	0,64	0
2	5	0,87	0,79	0,18	0,53	1
3	6	0,87	0,78	0,17	0,68	1
4	6	0,87	0,77	0,18	0,72	1
5	7	0,87	0,76	0,17	0,98	1
6	5	0,84	0,76	0,22	0,63	0
7	5	0,83	0,75	0,23	0,67	0
8	5	0,82	0,74	0,24	0,69	2
9	4	0,80	0,72	0,27	0,61	1
10	4	0,80	0,71	0,27	0,61	1

Os valores elevados dos coeficientes de correlação linear e de correlação de validação cruzada, r<sup>2</sup> e q<sup>2</sup>, para todos os modelos do alinhamento 4 são notáveis. Entretanto, os modelos 2, 3, 4, 5, 8, 9 e 10 apresentaram pelo menos um outlier, e, conseqüentemente, foram desconsiderados.

O modelo 1, do alinhamento 4, apresentou o maior valor de q<sup>2</sup>, nenhum outlier e o número de descritores está de acordo com as 'regras' da estatística, que impõem um limite de 4 a 5 observações (compostos) por descritor ou parâmetro. Apesar de não apresentarem outliers, os modelos 6 e 7 possuíam menores valores de r<sup>2</sup> e q<sup>2</sup> e maiores valores de LSE, em relação ao modelo 1.

A matriz de correlação-cruzada linear dos resíduos de ajuste entre pares de modelos do alinhamento 4 foi computada e está apresentada na Tabela VI. A base para determinar a correlação linear entre pares destes resíduos de ajuste consiste no fato de que um par de modelos equivalente deverá apresentar resíduos de ajuste quase idênticos ( $R \approx 1$ ), enquanto que um par distinto de modelos deverá ter resíduos não correlacionados (R<0.5), conforme relatado por Rogers e Hopfinger (1994) e Hopfinger (1999).

Modelos	1	2	3	4	5	6	7	8	9	10
1	1,00									
2	0,95	1,00								
3	0,92	0,98	1,00							
4	0,94	0,99	0,97	1,00						
5	0,92	0,97	0,99	0,97	1,00					
6	0,87	0,76	0,74	0,76	0,75	1,00				
7	0,82	0,85	0,82	0,85	0,81	0,86	1,00			
8	0,82	0,76	0,73	0,75	0,74	0,83	0,69	1,00		
9	0,77	0,82	0,79	0,82	0,79	0,89	0,92	0,76	1,00	
10	0,77	0,82	0,79	0,81	0,79	0,59	0,66	0,76	0,62	1,00

Tabela VI – Matriz de correlação linear dos resíduos de ajuste para os dez melhores modelos do alinhamento 4

Considerando os dados da Tabela VI, todos os modelos, do alinhamento 4, são equivalentes, pois possuem valores de resíduos de ajuste elevados ( $R \approx 1$ ) e

68

correlacionados. Isto significa que existe somente um modelo e, neste caso, é o modelo 1 (valores de r<sup>2</sup> e de q<sup>2</sup> mais elevados, nenhum outlier e número de GCODs condizente com as regras estatísticas).

O melhor modelo da análise de QSAR-4D, modelo 1 do alinhamento 4, é definido pela seguinte equação:

pMIC = -11,94 GC1(np) + 11,37 GC2 (a) + 23,07 GC3 (a) + 1,57 GC4 (a) + 6,43 GC5 (a) + 30,76 GC6 (np) - 0,23

N = 30 r<sup>2</sup> = 0,88 q<sup>2</sup> = 0,80 LSE = 0,16 LOF = 0,64 (1)

A equação ou modelo 1 (1) apresenta apenas 2 tipos de elementos farmacofóricos de interação (IPEs): qualquer tipo de átomo (a ou any) e átomo não polar (np).

A matriz de correlação linear dos GCODs (variáveis independentes) do modelo 1, do alinhamento 4, foi computada e está apresentada na Tabela VII. A finalidade destes dados é constatar se os descritores estão correlacionados entre si, em um mesmo modelo (KUBINYI, 1993; TOKARSKI, HOPFINGER, 1997 b; HOPFINGER, 1999).

Descritores	GCI	GC2	GC3	GC4	GC5	GC6
GC1	1,00					
GC2	-0,01	1,00				
GC3	-0,13	-0.17	1,00			
GC4	-0,15	0,14	-0,09	1,00		
GC5	-0,11	-0,16	-0,01	0,33	1,00	
GC6	-0,00	-0,21	-0,23	0,05	0,35	1,00

Tabela VII – Matriz de correlação-cruzada linear dos GCODs para o modelo 1 do alinhamento 4 Como demonstrado na Tabela VII, nenhum dos descritores do modelo 1 está consideravelmente correlacionado com o outro. Isto significa que cada um dos descritores fornece informação independente para este modelo. Portanto, não existe overfitting.

Visando verificar o poder de predição do modelo 1, do alinhamento 4, os valores de atividade biológica, traduzidos como pMIC, de cada um dos compostos do conjunto de avaliação foram calculados utilizando a equação ou modelo 1 (1).

A Tabela VIII apresenta os índices de coordenadas dos descritores para cada um dos compostos do conjunto de avaliação, onde x, y e z são as coordenadas cartesianas referentes à posição espacial dos descritores na célula da grade ou caixa 3D virtual.

GC1(-1,-3,-1)	GC2(-3,2,-3)	GC3(-1,8,4)	GC4(1,-2,-3)	GC5(4.0.4)	GC6(4,4,0)	Test_Set
0	0,102	0	0,064	0,021	0,0146	Idv90
0,0394	0,162	0,0022	0,1346	0,0112	0,0078	ldv128
0,0066	0,014	0,0004	0,0152	0,005	0,0264	ldv124
0,03	0,0412	0	0,0304	0,0006	0,0038	ldv131
0	0,1358	0,0028	0,0888	0,0007	0,0174	ldv132
0,0008	0,1634	0	ο	0,0094	0,012	ldv136
0	0,127	0	0	0,0002	0,0086	INHd51

Tabela VIII – Índices de coordenadas dos descritores para cada um dos compostos do conjunto de avaliação (Test\_Set)

Os valores da Tabela VIII, para cada composto do conjunto de avaliação, foram substituídos na equação ou modelo 1 (1), do alinhamento 4, para obter os valores calculados de atividade biológica (pMIC) e verificar a capacidade de predição do modelo.

A Tabela IX apresenta os valores de atividade biológica experimental (ATIV\_EXP) e prevista (ATIV\_PREV) para cada composto do conjunto de avaliação e os seus respectivos resíduos (diferença entre valores de atividade biológica calculada e experimental), considerando o modelo 1 do alinhamento 4.

ATIV_EXP	ATIV_PREV	RESÍDUOS	Test_Set
4,22	1,62	-2,60	Idv90
2,82	1,72	-1,10	ldv128
2,70	0,73	-1,97	Idv124
2,00	0,05	-1,95	ldv131
1,82	2,10	0,28	ldv132
0,52	2,05	1,53	ldv136
0,22	1,48	1,26	INHd51

Tabela IX – Predição de atividade biológica para os compostos do conjunto de avaliação

(Test\_Set)

As predições de atividade biológica do conjunto de avaliação estão apresentadas na Tabela IX e pode-se constatar que o modelo 1 exibiu um poder de predição médio pobre, 42,86%, considerando-se o valor de desvio padrão da média dos resíduos, 1,67. Três dos sete compostos (Idv90, Idv124, Idv131) apresentaram o valor de resíduo maior que o desvio médio.

Para descartar a possibilidade de que o conjunto de avaliação não era representativo do conjunto de treinamento, os conjuntos de treinamento e de avaliação formaram um só conjunto e o programa de 4D-QSAR foi rodado novamente para o alinhamento 4.

Na Tabela X estão apresentados os resultados dos dez melhores modelos construídos para o alinhamento 4 com o conjunto [treinamento+avaliação], utilizando o valor de fator de ajuste de 2,5.

Modelos	N° de GCODs	q²	r <sup>2</sup>	outlier(s)	LSE	LOF
1	5	0,60	0,68	1	0,45	1,02
2	6	0,59	0,69	1	0,43	1,23
3	5	0,57	0,68	2	0,45	1,04
4	5	0,57	0,66	1	0,47	1,08
5	5	0,56 '	0,65	ĩ	0,49	1,11
6	6	0,56	0,66	1	0,47	1,34
7	5	0,55	0,64	1	0,50	1,14
8	4	0,55	0,63	1	0,52	0,97
9	5	0,55	0,66	2	0,47	1,08
10	5	0,54	0,68	2	0,45	1,02

Tabela X – Medidas estatísticas, número de descritores e número de outliers para os dez melhores modelos do alinhamento 4 com o conjunto [treinamento+avaliação]

De acordo com a Tabela X, os valores de q<sup>2</sup> e r<sup>2</sup> diminuíram. O modelo 1 apresentou o maior valor de q<sup>2</sup> e um composto desviante (outlier). O composto desviante deste modelo era o ligante Idv90, que faz parte do conjunto de avaliação.

A matriz de correlação linear dos resíduos de ajuste entre pares de modelos do alinhamento 4 para o conjunto [treinamento+avaliação] foi computada e está apresentada na Tabela XI, com a finalidade de verificar se os modelos estão ou não correlacionados entre si.

.

Modelos	1	2	3	4	5	6	7	8	9	10
]	1,00									
2	0,99	1,00								
3	0,88	0,86	1,00							
4	0,93	0,92	0,89	1, <b>0</b> 0						
5	0,87	0,85	0,89	0,91	1,00					
6	0,76	0,74	0,74	0,88 <sup>-</sup>	0,86	1,00				
7	0,91	0,89	0,92	0,95	0,92	0,79	1,00			
8	0,93	0,92	0,94	0,96	0,97	0,82	0,98	1,00		
9	0,82	0,81	0,96	0,85	0,87	0,78	0,88	0,90	1,00	
10	0,91	0,91	0,91	0,87	0,90	0,69	0,91	0,93	0,85	1,00

Tabela XI – Matriz de correlação linear dos resíduos de ajuste para os dez melhores modelos

do alinhamento 4, conjunto (treinamento+avaliação)

Considerando-se os dados da Tabela XI, todos os modelos do alinhamento 4 para o conjunto [treinamento+avaliação] apresentaram resíduos de ajuste correlacionados entre si ( $R \approx 1$ ), indicando que os modelos são equivalentes. Isto significa que existe apenas um modelo que, de acordo com as medidas estatísticas demonstradas na Tabela X, é o modelo 1 (maior valor de q<sup>2</sup>).

O composto desviante do modelo 1 (Idv90) foi eliminado e o modelo foi reconstruído. Os valores de q<sup>2</sup> e r<sup>2</sup> aumentaram novamente e o número de outliers dos modelos variou de zero a três. O modelo 1 apresentou o maior valor de q<sup>2</sup>, conforme demonstrado na Tabela XII.

Modelos	Nº de GCODs	q²	r <sup>2</sup>	LSE	LOF	outlier(s)
1	7	0,77	0,85	0,20	0,75	1
2	6	0,75	0,83	0,22	0,65	2
3	7	0,74	0,83	0,22	0,84	2
4	7	0,74	0,83	0,22	0,82	2
5	7	0,73	0,83	0,22	0,83	2
6	6	0,72	0,81	0,24	0,70	0
7	6	0,71	0,80	0,26	0,75	1
8	6	0,71	0,82	0,23	0,67	2
9	7	0,71	0,83	0,22	0,83	1
10	6	0,69	0,79	0,27	0,79	3

Tabela XII – Medidas estatísticas, número de descritores e número de outliers para o modelo 1

do alinhamento 4 (conjunto [treinamento+teste])

A matriz de correlação linear dos resíduos de ajuste entre pares de modelos foi computada a fim de constatar se os modelos resultantes eram ou não equivalentes (Tabela XIII).

Tabela XIII – Matriz de correlação linear dos resíduos de ajuste para os modelos do alinhamento 4, sem o outlier, Idv90

Modelos	1	2	3	4	5	6	7	8	9	10
1	1,00									
2	0,94	1,00								
3	0,94	1,00	1,00							
4	0,95	0,99	0,99	1,00						
5	0,96	0,99	0,99	Ó,99	1,00					
6	0,81	0,89	0,89	0,88	0,87	1,00				
7	0,90	0,91	0,91	0,92	0,92	0,79	1,00			
8	0,89	0,94	0,94	0, <b>93</b>	0,95	0,82	0,86	1,00		
9	0,92	0,99	0,99	0,98	0,98	0,89	0,89	0,94	1,00	
10	0,81	0,85	0,85	0,87	0,85	0,79	0,91	0,7 <b>8</b>	0,84	1,00

De acordo com a Tabela XIII, os valores dos resíduos de ajuste dos modelos do alinhamento 4 estão correlacionados ( $R \approx 1$ ) uns com os outros, com valores

elevados de coeficientes de correlação linear, indicando que todos os modelos são equivalentes. Então, a hipótese de que o conjunto de avaliação não era representativo do conjunto de treinamento foi descartada.

O ligante Idv90 do conjunto de avaliação perturbou o poder de predição do modelo 1, resultante para o conjunto de treinamento com o alinhamento 4, provavelmente porque ele é muito mais ativo que o restante dos compostos do conjunto de avaliação. O ligante Idv90 é o segundo composto mais ativo, considerando os conjuntos de treinamento e de avaliação. Por esta razão, poderia estar causando algum desvio na capacidade de predição do modelo escolhido como o melhor modelo do alinhamento 4.

Conforme relatado por Golbraikh e Tropsha (2002), nem sempre um elevado valor da medida estatística q<sup>2</sup> (q<sup>2</sup> > 0,5) é considerado como prova de que um modelo apresentará elevada capacidade de predição. É condição necessária, mas não suficiente. Estes pesquisadores também sugeriram que o conjunto de avaliação deve incluir pelo menos cinco compostos, cujas atividades e características estruturais abranjam o conjunto de atividades e características estruturais dos compostos do conjunto de treinamento. Este critério é necessário para que se obtenham medidas estatísticas confiáveis para a comparação entre os valores das atividades experimental e calculada dos compostos envolvidos na análise.

Após ser descartada a hipótese de que o conjunto de avaliação não era representativo do conjunto de treinamento, partiu-se para a etapa seguinte da análise de QSAR-4D IR: a visualização das conformações 'ativas' previstas pelo modelo 1 (1), alinhamento 4, para cada um dos ligantes do conjunto de treinamento (programa 4D-QSAR 3.0) (HOPFINGER, 1999). As conformações 'ativas' previstas foram selecionadas considerando-se a identificação de todos os estados conformacionais amostrados para cada um dos ligantes dentro de uma variação

de energia de 5 kcal/mol (ΔΕ) da conformação de energia mínima global do CEP. Os GCODs de cada um dos grupos de conformações de menor energia resultantes foram empregados para a predição das atividades de cada um dos ligantes utilizando a equação ou modelo 1 (1), alinhamento 4, e, o confôrmero com maior atividade prevista foi selecionado como a conformação 'ativa'de cada ligante.

As conformações 'ativas' previstas apresentam os grupos fosfato protonados, já que todos os átomos de hidrogênio foram adicionados às estruturas dos ligantes (HyperChem 6.0) antes da otimização e, posterior "*input*" ao programa de 4D-QSAR 3.0. As conformações 'ativas' previstas para dois compostos (o mais ativo e o menos ativo) são demonstradas na Figura 5. Os confôrmeros foram desenvolvidos utilizando uma grade ou caixa 3D virtual com células de 1 Å de resolução. Esta medida corresponde ao diâmetro das esferas modeladas para representar os GCODs da equação (1) resultante da análise de QSAR-4D.



Figura 5 - Conformações ativas previstas pelo modelo 1, alinhamento 4, da análise de QSAR-4D. Os átomos de nitrogênio estão apresentados em azul; os hidrogênios estão em branco; oxigênios em vermelho; fósforos em cor de laranja e carbonos em cinza. Descritores GC1 e GC6 apresentam IPE = np; descritores GC2, GC3, GC4 e GC5 apresentam IPE = a ou any (equação (1)). Os descritores que aumentam a atividade biológica são apresentados como esferas amarelas (GC2, GC3, GC4, GC5 e GC6) e o descritor que diminui a atividade biológica (GC1) é apresentado como esfera verde. A intensidade da cor amarela dos descritores que aumentam a atividade está relacionada com seus respectivos coeficientes na equação (1). Quanto maior o valor do coeficiente, mais clara a coloração.

Na Figura 6 estão representados os dois compostos selecionados anteriormente (mais ativo e menos ativo) considerando a caixa 3D virtual originada pelo programa 4D-QSAR 3.0.



Figura 6 - Visualização da conformação 'ativa' prevista e seus respectivos descritores na caixa 3D virtual do programa 4D-QSAR 3.0: (a) Ligante INH1, mais ativo; (b) ligante INHd49, menos ativo do conjunto de treinamento.

77

78

Para verificar se o modelo selecionado se ajusta de forma satisfatória aos ligantes do conjunto de treinamento, as conformações 'ativas' previstas de cada um dos ligantes e seus respectivos descritores foram ancoradas (*docking*) no sítio ativo da enzima cristalizada **1zid**, utilizando o alinhamento 4 - modelo 1 como referência. Nas figuras 7, 8 e 9 este procedimento pode ser visualizado, por exemplo, para três ligantes do conjunto de treinamento, considerando-se os grupos de atividade biológica: INH1 (mais ativo), INHd18 (atividade intermediária) e INHd49 (menos ativo), respectivamente.



Figura 7 - Ligante INH1 da análise de QSAR-4D, com seus respectivos GCODs, ancorado no sítio ativo da InhA (1zid).

## RESULTADOS E DISCUSSÃO



Figura 8 - Ligante INHd18 da análise de QSAR-4D, com seus respectivos GCODs, ancorado no sítio ativo da InhA (1zid).



Figura 9 - Ligante INHd49 da análise de QSAR-4D, com seus respectivos GCODs, ancorado no sítio ativo da InhA (1zid).

79

Considerando-se os resíduos do sítio ativo, algumas distâncias foram medidas para comparar as possíveis interações entre as conformações 'ativas' previstas e seus respectivos descritores, provenientes da análise de QSAR-4D, e os resíduos de aminoácidos do sítio ativo com as interações entre o fármaco cocristalizado (INH) e os resíduos de aminoácidos da enzima InhA. Nas Tabelas XIV, XV e XVI estão apresentadas as distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os respectivos descritores dos ligantes INH1, INHd18 e INHd49.

Tabela XIV – Distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os descritores do ligante INH1 (mais ativo)

GC4 (a)	GC1(np)	GC2(a)	GC3(a)	GC5(a)	GC6(np)
Met155 S = 4,4 Å C (CH3) = 4,4 Å Phe149 {cadeia lateral) 3,6 Å	Phe 149 2,3 Å Ala 191 O (O=C) = 2,2 Å C (O=C) = 3,4 Å	<i>lle</i> 194 N (NH) = 2,1 Å <i>Pro</i> 193 O (O=C) = 3Å C (O=C) = 2Å N = 2,5 Å	Thr196 O (OH) = 1,9 Å Leu197 N(NH) = 3,7 Å Thr17 O (OH) = 5,1 Å Ser20 O (OH) = 4,9 Å	<i>ll</i> e95 O {O=C} = 5,9 Å <i>Ly</i> s165 N(NH) = 2,7 Å <i>Met161</i> C (CH <sub>3</sub> ) = 1,8 Å S = 3,4 Å	Tyr 158 O (OH) = 4,4 Å anei = 3,4 Å Met 103 C (CH <sub>3</sub> ) = 3,5 Å S = 4,3 Å

Tabela XV – Distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os descritores do ligante INHd18 (atividade intermediária)

GC4 (a)	GC1(np)	GC2(a)	GC3(a)	GC5(a)	GC6(np)
Met155 S = 6,4 Å C (CH3) = 6,6 Å Phe149 (cadeia lateral) 5,8 Å	Phe 149 4 Å Ala 191 O {O=C} = 3,4 Å C {O=C} = 4,3 Å	lle 194 - Pro 193 -	Thr 196 O (OH) = 6,5 Å Leu 197 N(NH) = 7,7 Å Thr 17 O (OH) = 5,8 Å Ser20 O (OH) = 3,2 Å	<i>lle95</i> O (O=C) = 5,2 Å <i>Lys165</i> N(NH) = 1,8 Å <i>Met161</i> C (CH <sub>3</sub> ) = 1,9 Å S = 3,7 Å	Tyr 158 O (OH) = 7,5 Å anel = 7,5 Å Met 103 C (CH <sub>3</sub> ) = 5,2 Å S = 5 Å

				the second s	
GC4 (a)	GC1(np)	GC2(a)	GC3(a)	GC5(a)	GC6(np)
Met155	Phe149	lle 194	Thr 196	lle95	Tyr158
S = 5,4 Å	4 Å	-	O (OH) = 4,2 Å	O (O=C) = 5,6 Å	O (OH) = 6 Å
C (CH3) = 5,6 Å	Ala 191	Pro 193	Leu197	Lys165	anel = 6,1 Å
Phe 149	O (O=C) = 3 Å	-	N(NH) = 5,3 Å	N(NH) = 2,6 Å	Met103
(cadeia lateral)	C (O=C) = 3.9 Å	Met199	Thr17	Met161	C (CH3) = 4.2 Å
5.6 Å		S = 4,1 Å	O (OH) = 4,5 Å	C (CH3) = 1.8 Å	S = 4,3 Å
		C(CG) = 4,1 Å	Ser20	S = 3,5 Å	Met199
		H(2HG) = 2,6 Å	O (OH) = 3 Å		C(CH3) = 2,8 Å
					S = 4,6 Å

Tabela XVI – Distâncias entre resíduos de aminoácidos do sítio ativo da enzima InhA e os descritores do ligante INHd49 (menos ativo)

De acordo com os dados apresentados na Tabela XIV, referentes ao ancoramento da conformação 'ativa' prevista do ligante mais ativo do conjunto de treinamento, INH1 (isoniazida ativa:NAD), na análise de QSAR-4 (Figura 7), pôdese obter as seguintes informações em relação ao complexo InhA-NAD-INH cristalizado (**1zid**) (ROZWASRKI *et al.*, 1998):

O descritor GC4, equação (1), é responsável por um pequeno aumento na atividade biológica (coeficiente = 1,57). Este descritor está localizado no átomo de nitrogênio do anel piridínico (0,6 Å). Os resíduos de aminoácidos do sítio ativo envolvidos em possíveis interações com este descritor são a metionina 155 (Met155) e a fenilalanina 149 (Phe149). O descritor GC4 apresenta como IPE qualquer tipo de átomo (*a*), provavelmente, porque está localizado numa região que permite ligação de hidrogênio com uma molécula de água próxima à cadeia lateral do resíduo Met155, na estrutura cristalizada, e, também, permite interações hidrofóbicas com outros resíduos do sítio ativo. Na estrutura cristalizada do complexo InhA-NAD-INH, o anel piridínico da porção isonicotínico-acil está cercado por resíduos hidrofóbicos, que incluem fenilalanina 149 (Phe149), glicina 192 (Gly192), prolina 193 (Pro193), leucina 218 (Leu218), tirosina 158 (Tyr158) e triptofano 222 (Trp22) (ROZWARSKI et al., 1998). A cadeia lateral do resíduo *Phe149* está

orientada paralelamente ao anel piridínico, permitindo sua participação em uma interação do tipo *m*-staking. A distância entre os dois anéis aromáticos é a mesma, tanto na estrutura cristalizada **1zid** como na conformação 'ativa' prevista do ligante INH1 ancorada no sítio ativo da enzima, proveniente da análise de QSAR-4D.

O descritor GC1, equação (1), é responsável pela diminuição da atividade biológica (coeficiente = -11,94) e apresenta IPE apolar (*np*). A distância entre GC1 e o átomo de carbono C49 (Figura 10) do anel piridínico é 1,6 Å e a distância entre este descritor e a cadeia lateral do resíduo fenilalanina 149 (*Phe149*) é 2,3 Å. Grupos substituintes apolares nesta posição (GC1) poderiam causar impedimento estérico impossibilitando a rotação da cadeia lateral do resíduo *Phe149* e, desta maneira, a interação *m*-stacking entre o anel piridínico e o anel aromático da cadeia lateral do resíduo *Phe149* não aconteceria. O GC1 também se encontra na região do sítio ativo cercada pelos resíduos hidrofóbicos *Phe149*, *Gly192*, *Pro193*, *Leu218*, *Tyr158* e *Trp222*.



Figura 10 – Conformação 'ativa' prevista do ligante INH1 e seus respectivos descritores: demonstração do carbono 49 do anel piridínico. O descritor GC2, equação (1), é responsável por aumento considerável na atividade biológica (coeficiente = 11,37) e está localizado próximo ao átomo de nitrogênio da porção nicotinamida do cofator (2,8 Å). Os resíduos que interagem com este descritor são prolina 193 (*Pro193*) e isoleucina 194 (*Ile194*). Então, o GC2 pode participar de interações como aceptor e doador de ligações de hidrogênio, e também de interações hidrofóbicas, já que está localizado na região do sítio ativo formada pelos resíduos hidrofóbicos *Phe149*, *Gly192*, *Pro193*, *Leu218*, *Tyr158* e *Trp222*. Estes tipos variados de interações explicariam a razão do IPE deste descritor ser qualquer tipo de átomo (a).

O descritor GC3, equação (1), é responsável por aumento ainda maior que o GC2 na atividade biológica (coeficiente = 23,07) e está localizado entre os grupos fosfato do NAD. A distância entre o GC3 e o átomo de oxigênio entre os dois grupos fosfato é de 2,6 Å. Os resíduos que, possivelmente, interagem com este descritor são treonina 196 (*Thr196*), leucina 197 (*Leu197*), treonina 17 (*Thr17*) e serina 20 (*Ser20*). Na estrutura cristalizada existe uma molécula de água participando de interações entre os resíduos *Thr17*, *Leu197* e o segundo grupo fosfato do NAD. O descritor GC3 apresenta IPE para qualquer tipo de átomo (a), o que poderia ser explicado pela sua participação em interações como aceptor e/ou doador de ligações de hidrogênio.

O descritor GC5, equação (1), também aumenta, de forma bem moderada, a atividade biológica (coeficiente = 6,43) e está localizado próximo ao átomo de hidrogênio do grupo hidroxila da primeira ribose do cofator (1,1 Å). No complexo cristalizado, este grupo hidroxila interage com os resíduos de aminoácidos lisina 165 (Lys165) e isoleucina 95 (Ile95). Há também uma molécula de água interagindo com estes dois resíduos. O descritor GC5 apresenta como IPE qualquer tipo de átomo (a), provavelmente, porque pode participar de interações

84

como aceptor e/ou doador de ligações de hidrogênio e também de interações hidrofóbicas com o resíduo de aminoácido metionina 161 (Met161).

O descritor GC6, equação (1), é responsável por aumento considerável da atividade biológica (coeficiente = 30,76) e está localizado próximo ao grupo hidroxila da segunda ribose do cofator (1,5 Å). Este descritor apresenta IPE para grupos apolares (*np*). No complexo cristalizado, a segunda ribose está próxima aos resíduos de aminoácido glicina 14 (*Gly14*) e alanina 22 (*Ala22*) e a estrutura INH-NAD está estirada. A conformação 'ativa' prevista do ligante INH1 (grupo isonicotínico-acil:NAD) na análise de QSAR-4D é 'dobrada' (forma de "U") e participa de interações hidrofóbicas com os resíduos de aminoácido tirosina 158 (*Tyr158*) e metionina 103 (*Met103*).

De modo geral, as conformações 'ativas' previstas da análise de QSAR-4D IR apresentaram-se mais 'dobradas' que a estrutura do fármaco co-cristalizado. Isto poderia ser atribuído, em parte, ao fato de que o perfil conformacional (CEP), gerado pelas simulações de DM, foi efetuado com os ligantes isolados, sem considerar as possíveis interações com os resíduos de aminoácido do sítio de interação.

Conforme os dados apresentados para o ligante INHd18, de atividade intermediária, na Figura 8 e Tabela XV, pode-se verificar que a conformação 'ativa' prevista do ligante INHd18 (4-nitrobenzóico-acil:NAD) na análise de QSAR-4D apresenta-se mais 'dobrada' ("U" mais fechado) que a do ligante INH1, resultando em certos 'colapsos' com os resíduos de aminoácido do sítio ativo da enzima cristalizada (1zid), como por exemplo, a segunda ribose do NAD com o resíduo metionina 199 (Met199), a porção adenina do NAD com os resíduos prolina 193 (Pro193) e isoleucina 194 (Ile194) e o impedimento estérico entre o grupo nitro do ligante INHd18 e o anel aromático do resíduo de aminoácido tirosina 158 (Tyr158). Os descritores GC1 e GC2 não estavam localizados na região do sítio ativo cercada
pelos resíduos hidrofóbicos *Phe149*, *Gly192*, *Pro193*, *Leu218*, *Tyr158* e *Trp222*. Este dois descritores são importantes na determinação da atividade biológica (equação (1)) e suas mudanças de posição (em relação ao ligante mais ativo) são indicativos consideráveis para explicar a atividade intermediária do ligante INHd18. Cabe ressaltar, que os descritores GC1 e GC2 participam de interações intramoleculares do ligante INHd18, o que pode ser compreendido pela forma mais 'dobrada' da conformação 'ativa' prevista. Além destes dois descritores, o GC3 também sofre alteração de posição distanciando-se dos resíduos Thr196 e Leu197. Isto prejudicaria de forma significativa a atividade biológica em relação ao composto mais ativo, uma vez que a contribuição deste descritor é a segunda mais importante (coeficiente = 23,07).

No caso do ligante inativo, INHd49, pode-se verificar pela análise dos dados apresentados na Figura 9 e na Tabela XVI, que além da mudança de posição dos descritores GC1, GC2 e GC3, em relação ao composto mais ativo INH1, os descritores GC2 e GC6 apresentaram interações com um resíduo de aminoácido a mais no sítio ativo, a metionina 199 (Met199). Provavelmente, este tipo de interação não é desejável, já que estes dois descritores, quando em outras posições, são responsáveis pelo aumento considerável da atividade biológica, de acordo com o modelo e alinhamento escolhidos neste estudo.

Como descrito por Hopfinger et al. (1997), Albuquerque et al. (1998) e Venkatarangan e Hopfinger (1999), a exploração das informações estruturais no modelo de QSAR-4D IR é realizada pelo alinhamento dos descritores (GCODs) do modelo e suas respectivas conformações 'ativas' previstas dentro do sítio de interação para verificar se o modelo de QSAR-4D se ajusta estericamente e apresenta complementaridade na interação ligante-receptor.

Neste estudo, a análise dos descritores (GCODs) no sítio ativo da enzima cristalizada utilizando as conformações 'ativas' previstas pela análise de QSAR-4D

foi realizada para todos os ligantes do conjunto de treinamento e o critério de ajuste, conforme já discutido, foi avaliado em função da comparação das medidas de distância dos resíduos de aminoácido do sítio de interação entre o fármaco cocristalizado, INH, e as conformações 'ativas' previstas dos ligantes e seus respectivos descritores, resultantes do modelo 1, alinhamento 4. Considerando-se os 10 compostos mais ativos, o modelo de QSAR-4D IR escolhido (modelo 1, alinhamento 4) ajustou-se bem para sete compostos.

Os três compostos que não se ajustaram ao modelo 1, alinhamento 4, foram: INHd46, INHd20 e INHd23. Apesar de considerados ativos, as possíveis interações entre os descritores das conformações 'ativas' previstas do modelo selecionado e os resíduos de aminoácido do sítio de interação (medidas de distância) não estavam dispostas nos locais considerados adequados, de acordo com a estrutura do complexo cristalizado (**1zid**). Assim, para os compostos INHd46, INHd20 e INHd23 o modelo selecionado na análise de QSAR-4D IR (modelo 1, alinhamento 4) não forneceu informações que possibilitassem a explicação de suas atividades biológicas.

Apesar de incluído na análise de QSAR-4D IR, na etapa de construção e otimização dos modelos, o descritor ClogP não se fez presente entre os 10 melhores modelos (equações) selecionados para o alinhamento 4 (programa 4D-QSAR 3.0), o melhor alinhamento. Portanto, nesta análise, este descritor não foi considerado significativo.

Klopmann e colaboradores (1996) consideraram, entre outros fatores à contribuição da atividade de derivados de INH, o efeito do valor de pKa de algumas hidrazidas. Os valores de pKa das hidrazidas dos ácidos isonicotínico, nicotínico, benzólco e cianoacético, por exemplo, estão muito distantes do pH fisiológico para serem considerados como tendo qualquer efeito em suas reatividades. Todas estas hidrazidas têm seus grupos ácido/básico não-ionizados no

pH celular e suas diferenças de reatividade não podem ser atribuídas à diferença de seus valores de pKa. Sugeriram como elementos característicos da reatividade da INH o grupo hidrazida, com seu nitrogênio nucleofílico, e a distância entre os nitrogênios piridínico e hidrazídico, que está associada ao processo químico de formação de base de Schiff da vitamina B<sub>6</sub>. Considerando-se o mecanismo de inibição de síntese de ácidos micólicos, estes pesquisadores relataram que a INH interage de modo covalente, após ligar-se ao cofator da proteína carregadora, com um α-cetoácido formado em estágio anterior no processo biossintético dos ácidos micólicos. Esta interação interrompe a construção dos ácidos micólicos e a parede celular micobacteriana torna-se suscetível ao sistema imunológico do hospedeiro.

### 4.2. ANÁLISE DE FEFF QSAR-3D

A análise de FEFF QSAR-3D, proposta por Tokarski e Hopfinger (1997a, 1997b) é considerada metodologia de QSAR-3D DR (dependente do receptor) em que (1) todas as contribuições de entalpia e entropia da interação receptor-ligante em meio solvente são levadas em consideração; (2) as contribuições de entalpia e entropia são tratadas como variáveis independentes no desenvolvimento da análise de QSAR para a interação receptor-ligante e (3) os modelos de FEFF QSAR-3D 'ideais' são construídos utilizando um algoritmo genético.

Segundo Venkatarangan e Hopfinger (1999), as informações da termodinâmica de interação inerentes ao modelo de QSAR-4D IR podem ser avaliadas pela comparação dos termos de energia do modelo de FEFF QSAR-3D aos descritores (GCODs) do modelo de QSAR-4D IR.

Neste estudo, para confirmar a validade do modelo 1 e utilizá-lo como alinhamento para a análise de FEFF QSAR-3D, simulações de DM de cada um dos compostos empregando o alinhamento da análise de QSAR-4D como ponto de partida foram desenvolvidas para verificar se as menores energias de interação das simulações de DM estavam relacionadas com os valores de atividade biológica. Isto significa, por exemplo, constatar se o composto mais ativo, INH1, apresenta a menor (melhor) energia de interação.

A estrutura da enzima **1zid**, cristalizada por Rozwarski e colaboradores (1998), com resolução de 2,7 Å, teve seu tamanho reduzido pela técnica de Tokarski e Hopfinger (1997 a), para torná-la acessível em termos de tempo computacional das simulações de DM. O tamanho da enzima foi reduzido de 4095 para 1912 átomos (Figura 11). A análise foi restrita aos resíduos de aminoácidos da enzima próximos à região do sítio ativo. A sobreposição da estrutura cristalográfica original e do modelo da enzima de tamanho reduzido resultou em um desvio da raiz dos mínimos quadrados (r.m.s.) de 0,20 Å. Este valor indica que, em termos de posicionamento dos átomos, a estabilidade estrutural do modelo reduzido da enzima em relação à estrutura cristalográfica foi mantida.



Figura 11 – Visualização das estruturas da enzima enoil-àcp redutase complexada com o inibidor isoniazida e o cofator NAD. a) enzima original (**1zid**); b) modelo reduzido da enzima (14 Å). Os átomos de hidrogênio não estão demonstrados. No quadro 8 estão apresentados os valores dos termos de energia para as conformações de energia mínima (Conf\_Emin) selecionadas durante o processo de "warming up" do modelo reduzido da enzima **1zid** otimizado. Os valores de energia potencial total (Etor) correspondem à soma dos valores de energia de deformação axial das ligações (Eest), energia de deformação angular (Edef), energia de torsão (Etors), energia de interação do tipo 1,4 (E1,4), energia de van der Waals (Evorw), energia eletrostática (Eel), energia intermolecular de van der Waals e eletrostática (Einterv+el), energia de solvatação (Esotv) e energia de ligação de hidrogênio (Ehb).

Quadro 8 – Termos de energia obtidos no processo de "warming up" do modelo reduzido da enzima **1zid** cristalizada com o inibidor (isonicotínico-acil:NAD)

Conf_Emin	Eest	Eder	E <sub>lors</sub>	•E1,4	Evaw	Ee	Entervtel	Esolv	Бњ	Ετοτ
25 K	124,51	564,74	202,79	1828.18	-976,19	-1920,28	-583,91	-93.18	-16560,30	-17413.64
50 K	161,38	641,76	223,61	1824.68	-987,69	-1890,40	-665,33	-100,12	-16589,37	-17381,48
100 K	280,30	751,07	226,71	1859,86	-960,10	-2022,61	-685,13	-111.76	-16467,44	-17129,10
150 K	363,49	873.72	260,24	1871,49	-940,94	-1986,66	-674,95	-91,92	-16469.92	-16795,45
200 K	511,15	988,84	264,77	1838,92	-897,71	-1957,47	-707,97	-95.15	-16417.26	-16471,88
250 K	563,72	1101,50	311,64	1915.42	-864,71	-1995.23	-608,16	-77,52	-16337,04	-15990,38
300 K	689,88	1187,07	327,43	1918,56	-860.02	-1976.82	-647.07	-68,21	-16549,77	-15978,95
310 K	722.56	1169,96	343,74	1930.08	-827.24	-2056,62	-700,03	-75,65	-16519,40	-16012,60

\* As unidades dos valores de energia são kcal/mol. E<sub>est</sub> = energia de deformação axial das ligações; E<sub>det</sub> = energia de deformação angular; E<sub>tors</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>vdw</sub> = energia de van der Waals;
 E<sub>et</sub> = energia eletrostática; E<sub>ntervel</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>xdw</sub> = energia de solvatação; E<sub>te</sub> = energia de ligação de hidrogênio; Eror = energia potencial total.

De acordo com o quadro 8, a conformação de energia mínima do modelo reduzido da enzima **1zid** selecionada à T de 310 K apresentou energia potencial total de -16.012,60 kcal/mol. Esta conformação foi utilizada pard<sup>4</sup> o ancoramento do ligante mais volumoso do conjunto de treinamento resultante da análise de QSAR-4D (modelo 1), considerando o alinhamento 4. Na Tabela XVII estão apresentados os volumes das conformações 'ativas' previstas dos compostos do conjunto de treinamento obtidos com o programa HyperChem 6.0 (Hypercube Inc., 1996).

Ligantes	Volume (ų)
інні	1626,81
INHd2	1609,59
INHd31	1711,35
INHd43	1704,89
INHd14	1679,34
INHd20	1685,34
INHd46	1695,25
INHd37	1618,21
INHd15	1631,39
INHd23	1694,55
iNHd29	1765,01
INHd16	1719,93
INHd18	1631,75
INHd25	1638,56
INHd30	1702,94
INHd44	1626,72
ldv130	1554, 19
INHd22	1692,30
INHd27	1738,45
INHd34	1676,73
INHd42	1759,05
INHd47	1723,97
INHd45	1712,63
ldv107	1731,74
ldv126	1671,79
INHd19	1800,91
ldv125	1749,43

Tabela XVII – Volume das conformações 'ativas' previstas dos compostos do conjunto de treinamento, resultantes da análise de QSAR-4D IR

## 90

Tabela XVII (continuação) – Volume das conformações 'ativas' previstas dos compostos do conjunto de treinamento, resultantes da análise de QSAR-4D IR

Ligantes	Volume (ų)	
INHd41	1718,30	
INHd48	1680,99	
INHd49	1646,46	

Conforme resultados da Tabela XVII, o ligante de maior volume do conjunto de treinamento foi o INHd19. A conformação 'ativa' prevista deste ligante, resultante da análise de QSAR-4D (modelo1), foi ancorada na conformação de energia mínima a 310 K do modelo reduzido da enzima **1zid**, utilizando o alinhamento 4. O complexo enzima-ligante INHd19 otimizado foi submetido ao procedimento de "warming up" e os termos de energia das conformações de energia mínima (Conf\_Emin) selecionadas em diferentes temperaturas estão apresentados no quadro 9, bem como seus valores de energia potencial total (Eror) resultantes deste processo.

Quadro 9 – Termos de energia obtidos no processo de "warming up" do modelo reduzido da enzima InhA com o ligante INHd19 (conformação 'ativa' prevista da análise de QSAR-4D IR)

Conf_E <sub>min</sub>	Eest	Edet	Elon	E1,4	Evaw	Eel	Enterret	Esolv	Бњ	Етот
25 K	129,49	566,08	212,16	1804,17	-975,35	-1951,28	-660,82	-52,84	-17102,70	-18031,09
50 K	180,23	629,80	219,46	1809,86	-976,91	-2016,27	-668,51	-58,27	-17022.94	-17903,55
100 K	286,53	760,50	243,04	1812,50	-954,92	-2072.22	-710,58	-56,28	-16970,35	-17661.78
150 K	402,68	885.34	271,58	1833,92	-925,34	-2068,39	-677,58	-61.69	-16968,45	-17307,93
200 K	492,00	1013.64	306,88	1809,38	-909,65	-2161,91	-729,56	-61,35	-17031.29	-17271.86
250 K	604,14	1082,01	318,36	1849,05	-869,19	-2135,37	-788,28	-61,74	-16757,11	-16758,13
300 K	706,37	1186,02	350,97	1888,86	-832,08	-2038,88	-614.24	-78,55	-16370,17	-15801,70
310 K	733,31	1252,56	336,86	. 1884,31	-832,90	-2049,93	-625,37	-69,43	-16399,89	-15770,48

\* As unidades dos valores de energia são kcal/mol. E<sub>sti</sub> = energia de deformação axial das ligações; E<sub>det</sub> = energia de formação angular; E<sub>let</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>valv</sub> = energia de van der Waals; E<sub>d</sub> = energia eletrostática; E<sub>nervel</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>salv</sub> = energia de solvatação; E<sub>tb</sub> = energia de ligação de hidrogênio; E<sub>tor</sub> = energia potencial total.

KERLY FERNANDA MESQUITA PASQUALOTO

91

Conforme o quadro 9, a conformação de energia mínima do modelo reduzido da enzima com o ligante INHd19 selecionada à T de 310 K apresentou energia potencial total de -15.770,48 kcal/mol. Esta conformação foi utilizada para o ancoramento dos demais ligantes do conjunto de treinamento resultantes da análise de QSAR-4D (modelo 1), considerando o alinhamento 4. Os termos de energia resultantes da simulações de DM de 10 ps a 310 K dos complexos enzimaligante otimizados, incluindo a energia potencial total, estão apresentados no quadro 10.

# Quadro 10 – Valores dos termos de energia resultantes das simulações de DM a 310 K dos complexos enzima-ligante (E-L)

Complexos E-L	E.st	Edel	Etors	E1.4	Evaw	Eat	Eintervtal	Esolv	Ень	Eror
INH1	712.59	1191,81	371,39	1868,95	-773.24	-2034,75	-470,96	-68,72	-16119,46	-15322,39
INHd2	640,13	1252,10	380,70	1910,08	-792,67	-2044,52	-573,05	-60,95	-15481.94	-14770,12
INHd31	652,54	1241.17	399,69	1840,60	-759.12	-2010,40	-576,64	-66,92	-16058.60	-15337,68
INHd43	705,61	1190,13	364.21	1919.07	-846,56	-2111,53	-751.67	-55,47	-17087.25	-16673.46
INHd14	675.25	1207.71	356,79	1907,56	-792,45	-1952,28	-537,72	-62,61	-16030,70	-15228,45
INHd20	618,17	1230,50	380.24	1907,22	-783,44	-2024,69 -	-624,06	-67,48	-16381,13	-15744.67
INHd46	649,59	1263,41	364,43	1848,10	-835,64	-2018;87	-563,49	-67,78	-16189,82	-15550,07
INHd37	729,89	1212,33	358,24	1845,40	-865,83	-1990,11	-603,81	-54,49	-16299,42	-15667,80
INHd15	668,87	1200.84	391,65	1900,28	-778,32	-2107,79	-609,36	-73,19	-15648,59	-15055,61
INHd23	1429,40	1280.09	391.09	1909,71	-425.98	-1927,16	-129,74	-60.80	-16345,96	-13879,35
INHd29	633,02	1301.20	352,78	1869,81	-813,73	-1957,57	-537,56	-77.03	-15751.65	-14980,73
INHd16	686,07	1200,03	386,16	1885,13	-796,48	-1995,41	-607,02	-82,76	-16056,29	-15380.57
INHd18	652,29	1219,91	372,11	1929,51	-776,70	-1973,94	-553,31	-64,94	-16223,62	-15418.69
INHd25	688,97	1202,25	360,42	1893,59	-803,40	-1986,77	-511,79	-71,70	-16241.77	-15470.20
INHd30	690,98	1248,01	369,63	1845,41	-787.82	-2070,66	-600,72	-78,23	-15927,62	-15311,02
INHd44	656,04	1242,04	400,18	1858,91	-767,35	-1977,33	-513,27	-69,13	=15700,27	-14870.18
ldv130	703.23	1183,79	355,99	1875,47	-775,44	-1995,02	-492,16	-74,95	-15979.22	-15199.31
INHd22	667,72	1204.79	382,74	1868,75	-767,07	-1989,58	-526,65	-67,91	-16306.23	-15533.44
INHd27	1575,24	1292.42	355,75	1874,80	-283,38	-1993,89	-35,27	-79,35	-15327,75	-12621,43
INHd34	668,48	1196,74	402,24	1875,30	-751,83	-2001.56	-507.03	-75.81	-15533,83	-14727:30

KERLY FERNANDA MESQUITA PASQUALOTO

Quadro 10 (continuação) – Valores dos termos de energia resultantes das simulações de DM a 310 K dos complexos enzima-ligante (E-L)

Complexos E-L	Eest	Edet	Eton	E1,4	Evdw	Ee	Eintervtel	Esotv	Екь	Etor
INHd42	701,38	1218,52	385,75	1883,63	-773,77	-1985,92	-568,52	-66,25	-16029,62	-15234,80
INHd47	571,49	1173,16	331,45	1826,45	-811,82	-1850,17	-469,42	-70,67	-15710,10	-15009,63
INHd45	681,43	1235,43	389,10	1873,81	-790,74	-2009,24	-599,72	-58,59	-16099,30	-15377,82
ldv107	574,91	1079,49	344,21	1882,08	-779,68	-1850,76	-449,22	-63,63	-16243,73	-15506,33
ldv126	650,35	1239,10	363,79	1872,11	-774,31	-2001,87	-566,79	-76,05	-16242,13	-15535,80
INHd19	703,04	1241,53	345,76	1894,93	-828,06	-2038,50	-615,40	-67,17	-16328,53	-15692,40
ldv125	654,49	1235,51	378.32	1864,93	-764,29	-1983,39	-437.96	-74,95	-15351,43	-14478,77
INHd41	701,20	1255,82	364,63	1817,40	-770,16	-2042,26	-594,87	-61,23	-15792,18	-15121.65
INHd48	646,45	1255,82	381,00	1886,78	-781,77	-1915,72	-491,99	-61,20	-16071,43	-15152,06
INHd49	673,08	1253,59	374.22	1894,05	-747,61	-2029,55	-463,91	-55,64	-15483,64	-14585,41

As unidades dos valores de energia são kcal/mol. E<sub>st1</sub> = energia de deformação axial das ligações; E<sub>de1</sub> = energia de deformação angular; E<sub>lors</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>votw</sub> = energia de van der Waals;
 E<sub>ot</sub> = energia eletrostática; E<sub>intervent</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>sotv</sub> = energia de solvatação; E<sub>1t0</sub> = energia de ligação de hidrogênio; E<sub>tot</sub> = energia potencial total.

Comparando-se os valores de Eror (Conf\_Emin a 310 K) para os complexos E-L com o inibidor co-cristalizado (INH) (Quadro 8) e com o ligante INH1, proveniente da análise de QSAR-4D, pode-se observar que o complexo com o ligante INH1 apresentou maior energia potencial total e sua estrutura está mais 'dobrada' do que a do inibidor cristalizado, no sítio de interação. A porção acil das duas estruturas (INH1 e INH) ocupa, praticamente, o mesmo espaço na enzima, mas a porção NAD do ligante INH1 apresenta outra disposição no sítio de interação em relação à porção NAD do inibidor co-cristalizado (Figura 12).

94



Figura 12 – Visualização das conformações de energia mínima a 310 K dos complexos E-L com o ligante INH1 (a) e com o inibidor co-cristalizado, INH (b), respectivamente. Os hidrogênios do modelo reduzido da enzima não estão demonstrados. O ligante INH1 e o inibidor co-cristalizado INH estão apresentados em modelo "CPK" e os resíduos de aminoácidos em "stick" (WebLab ViewerLite 4.0). Os átomos de hidrogênio do ligante INH1 e do inibidor INH estão demonstrados em esferas brancas; os demais átomos estão em esferas azuis.

Os termos de energia resultantes das simulações de DM a 310 K dos complexos E-L otimizados, com a ligação covalente entre o carbono carbonílico dos compostos e o C4 do anel nicotinamida do NAD (cofator) cindida, estão apresentados no quadro 11, incluindo a energia potencial total. Após a cisão da ligação covalente, o grupo hidrazida da porção acil de cada um dos ligantes foi reconstruído. A porção NAD de cada ligante também foi reconstruída por meio da adição de um átomo de hidrogênio ao C4 do anel nicotinamida e as estruturas minimizadas dos complexos modelo reduzido da enzima com ligante reconstruído (porção acil e porção NAD separadas, no sítio ativo) foram submetidas a simulações de DM a 310 K.

Quadro 11 - Valores dos termos de energia resultantes das simulações de DM a 310 K dos complexos E-L com a ligação covalente entre a porção acil e o NAD dos ligantes cindida

Complexos E-L	Eest	Eder	Etors	E1,4	Evaw	Ee	Einterviel	Esoty	Ень	Etor
INH1	686,54	1192,28	372,1	1892,16	-815,86	-1946,78	-541,94	-68,56	-15798,06	-15028,12
INHd2	710,06	1246,12	350,48	1859,48	-799,57	-1957,47	-545,71	-69,75	-15496,41	-14702,77
INHd31	682,88	1264,92	362,29	1890,55	-783,39	-2041,40	-610,05	-79,93	-15244,8	-14558,93
INHd43	652,53	1261,13	372,32	1835,96	-856,44	-1960,18	-638,90	-52,49	-16832,42	-16218,49
INHd14	677,89	1211.17	351,93	1891,65	-789,93	-1968,67	-562,72	-58,90	-15806,67	-15054,25
INHd20	699,26	1219,07	403,62	1855,12	-803,10	-2037,96	-570,63	-67,18	-16355,15	-15656,95
INHd46	644,75	1228,42	371,39	1912,91	-806,90	-1962,60	-526,66	-61,56	-16065,44	-15265,69
INHd37	609,27	1231,94	335,39	1847,34	-858,24	-2062,89	-732,49	-66,44	-16152,28	-15848,40
INHd15	667,03	1239,78	359,67	1807,98	-785,20	-1948,11	-478,09	-65,78	-15751,79	-14954,51
INHd23	1462.22	1248,72	378,89	1919,04	-451,99	-2037,61	-218,95	-66,29	-16223,12	-13989.09
INHd29	686,37	1239,76	362,86	1867,91	-812,27	-1963,43	-537,68	-63,10	-15821,66	-15041,24
INHd16	664,14	1214,47	385,14	1837,75	-788,09	-1929,74	-498,48	-61,30	-15695,92	-14872,03
INHd18	704,13	1205,61	352,35	1889,21	-771.85	-2062,18	-504,88	-51,15	-15441.56	-14680,32
INHd25	654,76	1233,81	372,36	1860,64	-769,27	-2036,04	-599,54	-68,20	-15627,48	-14978,96
INHd30	706,06	1223,34	391,58	1841,07	-807,86	-2037,90	-554,58	-94,04	-15796,23	-15128,56
INHd44	692,80	1184,27	381.95	1919,97	-769,43	-2090,83	-614,14	-59,99	-16113,40	-15468,8
Idv130	679,08	1199,99	366,8	1846,94	-809,34	-1982.86	-553,74	-77,42	-15559,78	-14890,33
INHd22	673,20	1189,38	390,32	1900,90	-783,35	-2030,28	-557,49	-68,25	-15657,30	-14942,87
INHd27	1600,55	1247,09	370,86	1862,16	-339,55	-1955,22	-163,01	-69,25	-15918,14	-13364,51
INHd34	686,06	1231,62	331,45	1883,65	-779,94	-1910.80	-492.54	-84,81	-15489.08	-14624,39
INHd42	562,68	1038,82	301,08	1896,92	-836,78	-1949,07	-511,53	-68,01	-16353,70	-15919,59
INHd47	641,75	1157,8	369,49	1850,79	-793,82	-1836,22	-475,26	-81,05	-15441.97	-14608,49
INHd45	683,21	1243,04	412,47	1813,84	-819,53	-1996,05	-553,38	-68,65	-15947,25	-152 <b>32,30</b>
ldv107	681,80	1234,35	395,09	1828,38	-801.02	-2003,56	-525,56	-72.44	-15620,99	-14883,95
ldv126	691,13	1175,7	373,06	1884,8	-771,86	-1922,66	-473.40	-77,58	-15687,06	-14807,87
INHd19	718.05	1200.3	377,02	1830,8	-802,54	-2006,57	-544,42	-75,040	-16154.99	-15457,39
ldv125	641,23	1211,09	370,01	1818,95	-775,43	-1837,00	-462,73	-67,040	-15586.20	-14687,12
INHd41	678,65	1237,81	381,12	1850,92	-798,83	-1984,07	-518,44	-61,75	-15422,45	-14637.04
INHd48	695,84	1208,82	396,36	1892,10	-816,61	-2010.28	-596,60	-57.77	-16591,21	-15879,35
INHda9	702.05	1243.10	354.78	1868,37	-790.04	-1988.47	-587.43	-71.63	-16022 57	-15291 84

As unidades dos valores de energia são kcal/mol. E<sub>est</sub> = energia de deformação axial das ligações; E<sub>det</sub> = energia de deformação angular; E<sub>tors</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>vot</sub> = energia de van der Waals;
 E<sub>et</sub> = energia eletrostática; E<sub>ntervel</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>sot</sub> = energia de solvatação; E<sub>to</sub> = energia de ligação de hidrogênio; Eror = energia potencial total.

KERLY FERNANDA MESQUITA PASQUALOTO

95

As conformações de menor energia dos complexos E-L apresentadas no quadro 11 foram utilizadas para medir a distância entre os átomos de cada ligante (composto:NAD) que participam da ligação covalente entre a porção acil e o NAD e este dado foi utilizado como mais um descritor na construção dos modelos de FEFF QSAR-3D. Apesar da limitação dos parâmetros de energia resultantes das simulações de DM, devido à condição assumida na reconstrução dos ligantes após a cisão da ligação covalente, este descritor teria como propósito introduzir nos modelos hipóteses a respeito do mecanismo de ação dos compostos em estudo, considerando-se a aproximação dos compostos do conjunto de treinamento ao cofator NAD, no sítio de interação. Os valores da distância entre os átomos envolvidos na ligação covalente (d<sub>Lcov</sub>), entre a porção acil e NAD, dos compostos do conjunto de treinamento, nas conformações de menores energias resultantes das simulações de DM 10 ps a 310 K, estão apresentados no quadro 12.

Quadro 12 – Valores de distância entre os átomos envolvidos na ligação covalente (d<sub>Lcov</sub>) entre a porção acil e o NAD dos ligantes nos complexos E-L (E<sub>mín</sub> a 310 K) do conjunto de treinamento

Complexos E-L	dLoov (Å)
INH1	4,43
INHd2	7,94
INHd31	8,10
INHd43	3,74
INHd14	6,69
INHd20	5,26
INHd46	8,04
INHd37	4,12
INHd15	11,64
INHd23	4,94
INHd29	5,17
INHd16	4,87
INHd18	5,50
INHd25	4,39

KERLY FERNANDA MESQUITA PASQUALOTO

Quadro 12 (continuação)- Valores de distância entre os átomos envolvidos na ligação covalente (d<sub>Lcov</sub>) entre a porção acil e o NAD dos ligantes nos complexos E-L (E<sub>mín</sub> a 310 K) do conjunto de treinamento

Complexos E-L	dicov (Å)
INHd30	4.24
INHd44	3,79
ldv130	3,77
INHd22	9,33
INHd27	5,21
INHd34	5,68
INHd42	5,60
INHd47	5,81
INHd45	. 4.33
ldv107	6,10
ldv126	4,51
INHd19	4,14
ldv125	7,36
INHd41	10,40
INHd48	5,40
INHd49	6,42

Os valores dos termos de energia para os estados não-ligados do ligante e da enzima também foram calculados para se obter os descritores de energia que serão utilizados na construção dos modelos de FEFF QSAR-3D.

Os termos de energia do estado não-ligado da enzima foram obtidos a partir da simulação de DM de 10 ps a 310 K da estrutura otimizada correspondente à conformação de energia mínima do modelo reduzido da enzima (310 K) com o inibidor co-cristalizado excluído do sítio ativo. Os valores dos termos de energia, incluindo a energia potencial total estão apresentados no quadro 13.

Quadro 13 – Valores dos termos de energia resultantes da simulação de DM de 10 ps a 310 K do estado não-ligado do modelo da enzima InhA (sem o inibidor co-cristalizado)

Enzima	Een Eder	Eton	E1.4	Evdw	Eel	Einterv+ei	Esotv	Ень	Еют
310 K	680,15 1160,38	341,96	2252,97	-757,64	-2059,10	-486,54	-62.09	-14352,84	-13.282,75

As unidades dos valores de energía são kcal/mol. E<sub>01</sub> = energía de deformação axial das ligações; E<sub>001</sub> = energía de deformação angular; E<sub>1015</sub> = energía de torsão; E<sub>1.4</sub> = energía de interação do tipo 1,4; E<sub>volw</sub> = energía de van der Waals;
 E<sub>01</sub> = energía eletrostática; E<sub>ntervol</sub> = energía intermolecular de van der Waals e eletrostática; E<sub>volv</sub> = energía de solvatação; E<sub>105</sub> = energía de ligação de hidrogênio; E<sub>107</sub> = energía potencial total.

Os termos de energia do estado não-ligado dos ligantes do conjunto de treinamento foram obtidos a partir de simulações de DM de 50 ps a 310 K das estruturas otimizadas provenientes da análise de QSAR-4D (conformações 'ativas' previstas pelo modelo 1, do alinhamento 4). Os valores dos termos de energia, incluindo a energia potencial total estão apresentados no quadro 14.

Ligantes	Eest	Eder	Elors	E1,4	Evaw	En	Esotv	Ень	Eror
INH1	28,69	57,86	20,76	-393,79	-21,94	117,41	-12,66	-345,36	-549,02
INHd2	32,22	65,16	26.63	-407.24	-19,58	105,21	-12,95	-344,86	-555.41
INHd31	35,69	64,23	42.09	-417,46	-14,84	111,33	-10,92	-343,80	-533,68
INHd43	25,11	70,52	23,48	-405.29	-14,27	106,38	-14.23	-271,52	-479.82
INHd14	27,52	64,36	24,82	-404,42	-14,62	115,10	-16,42	-288,44	-492,10
INHd20	32,17	70,87	25,68	-395,41	-12,67	85,90	-22,23	-272,99	-488,68
INHd46	31,26	67,01	25,84	-408.32	-16,76	110,54	-14.90	-353,94	-559,27
INHd37	20,07	67,26	21,02	-402,52	-11.05	108,45	-18,31	-236,91	-451.99
INHd15	32.09	58,42	22,28	-398,83	-16,21	97,22	-14.06	-323,20	-542,29
INHd23	27,50	69.26	28,26	-419.10	-19,27	120,15	-11.39	-418.55	-623.14
INHd29	24,75	63,27	22,46	-405,55	-24,68	109,87	-15,19	-267.65	-492.72
INHd16	35,16	61,42	24,82	-409,46	-18,69	106,10	-13,51	-333,07	-547.23
INHd18	34,79	62.03	27,08	-397,89	-18,99	97,53	-15,28	-330,39	-541,12
INHd25	27,84	69,76	.25.69	-421,84	-16,19	113,39	-16,97	-309,04	-527, 36
INHd30	26,60	64,87	23,31	-415,50	-23,13	120,70	-14.58	-303,52	-521-25

Quadro 14 – Valores dos termos de energia resultantes das simulações de DM de 50 ps a 310 K do estado não-ligado dos compostos do conjunto de treinamento

KERLY FERNANDA MESQUITA PASQUALOTO

Quadro 14 (continuação)- Valores dos termos de energia resultantes das simulações de DM de 50 ps a 310 K do estado não-ligado dos compostos do

Ligantes	E.,	Eder	Elons	E1.4	Evaw	Ea	Eacty	Ень	Etor
INHd44	30,45	-71.71	24,11	-417.73	-15,78	84,71	-16,84	-277,31	-516,68
ldv130	25,76	61,31	21,60	-410,51	-19,03	112,46	-14.96	-288,63	-512,00
INHd22	38,13	64,46	20,75	-426,72	-20,08	113,41	-13,99	-381,73	-605,77
INHd27	37,58	69.70	27,27	-406,40	-21,34	108,24	-16.69	-345,96	-547,59
INHd34	25,99	78.24	23,26	-415,56	-19,64	108,39	-16.00	-300.09	-515,41
INHd42	33,66	62.90	21,67	-396.59	-16,81	125,76	-9.91	-417,82	-597.14
INHd47	29,18	56,70	28,18	-409,81	-20,87	96,78	-15,38	-301,24	-536.46
INHd45	21,80	65,76	25,45	-389,69	-18,92	114.59	-9.53	-340,10	-530,65
Idv107	29,75	64.11	27.72	-411.65	-15.50	96,04	-14,01	-269,14	-492,68
Idv126	23,07	70,78	30,60	-418,99	-17,11	114,02	-15,43	-274,36	-487,41
INHd19	33,31	71,65	23,64	-409,56	-18,90	109,33	-16,37	-369,08	-575,98
Idv125	24,69	87.58	27,82	-419,02	-16,61	113,36	-12,17	-302,01	-496,36
INHd41	28,54	75,92	22,77	-419,43	-17,90	107,77	-14,15	-355,03	-571,51
INHd48	29,58	78,12	33,57	-427,71	-23,58	121,98	-11,95	-330,48	-530,47
INHd49	32,28	68,54	27,51	-407,40	-18,23	120,15	-10,31	-338,98	-526,44

conjunto de treinamento

As unidades dos valores de energia são kcal/mol. E<sub>est</sub> = energia de deformação axial das ligações; E<sub>det</sub> = energia de deformação angular; E<sub>ters</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>vol</sub>w = energia de van der Waals;
 E<sub>et</sub> = energia eletrostática; E<sub>intervet</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>intervet</sub> = energia de solvatação; E<sub>40</sub> = energia de ligação de hidrogênio; E<sub>tor</sub> = energia potencial total.

Como mencionado na metodologia (item **3.4.7.**), outros descritores não determinados diretamente dos termos de campo energético foram incluídos na construção dos modelos de FEFF QSAR-3D. Além da distância dos átomos envolvidos na ligação covalente entre a porção acil e o NAD dos ligantes (d<sub>Lcov</sub>), outras propriedades do ligante no estado não-ligado foram determinadas. Estas propriedades foram calculadas para as conformações de menor energia resultantes das simulações de DM de 50 ps a 310 K e incluem: energia do orbital molecular ocupado de maior energia (energia de HOMO), energia do orbital molecular desocupado de menor energia (energia de LUMO), momento de dipolo, coeficiente de partição *n*-octanol/água calculado (ClogP), volume e área

99

superficial. As propriedades ClogP, volume e área superficial foram obtidas com o programa HyperChem 6.0 (Hypercube, Inc., 1996) e as energias de HOMO e LUMO e o momento de dipolo foram obtidos com o programa MOPAC (método hamiltoniano AM1, DEWAR et al., 1985). Estas propriedades estão apresentadas no quadro 15. O programa HyperChem 6.0 baseia-se no método de fragmentação de átomos, desenvolvido por Ghose, Pritchett e Crippen (1988), para o cálculo do coeficiente de partição. Para moléculas orgânicas, o método gera um coeficiente de correlação (r) com os valores experimentais de 0,92 e um erro padrão de 0,36.

Quadro 15 - Propriedades dos ligantes no estado não-ligado incluídas na construção dos

Ligantes	Volume (Å <sup>3</sup> )	Superfície (Å <sup>2</sup> )	ClogP	HOMO (eV)	LUMO (eV)	Dipolo (debye)
INH1	1619,66	806.99	-3,44	-8,94	-1,13	7.71
INHd2	1655,73	866,38	-1.94	-9,23	-1.92	12.07
INHd31	1724,12	907,13	-1,56	-8,75	-0.99	4,30
INHd43	1708.01	926.55	-2.85	-9.19	-1.92	9,36
INHd14	1713,93	906,71	-1,97	-8,55	-1,37	12,28
INHd20	1855,63	982,99	-5.63	-8.91	-1,76	10.46
INHd46	1697.78	882,19	-5.04	-9,31	-1,58	8,98
INHd37	1743,22	963.01	-3,44	-8.73	-1,24	11,07
INHd)5	1687,69	861,45	-2,19	-9,10	-1,17	4,64
INHd23	1610,88	792,13	-3,69	-9,02	-1,19	:8,86
INHd29	1688,36	841,70	-2,26	-8,99	-1,65	8,18
INHd16	1742,08	906,62	-1,92	-8,82	-1,35	6.42
INHd18	1707,06	856,54	-5,63	-9,38	-2.49	9,03
INHd25	1703,81	862,61	-2.99	-8,83	-1,88	8,02
INHd30	1656,46	815,69	-2,04	-8,67	-1,62	6,49
INHd44	1758,63	937,94	-2,25	-9,37	-1,63	3.94
ldv130	1607.08	812,03	-3.92	-9,37	-1,68	8.74
INHd22	1680.07	842,08	-3.69	-8,91	-1.05	12.34
INHd27	1693,02	836,36	-3,99	-8,67	-1.39	9,67
INHd34	1786,11	934.54	-2.47	-9,37	-1.70	7,83
INHd42	1702,96	851,83	-3,28	-8,99	-1.55	7.78
INHd47	1698,50	866,03	-3.51	-9,28	-1,78	2,39
INHd45	1632,38	813,88	-3,16	-9,05	-1.13	4.76

modelos de FEFF QSAR-3D

Faculdade de Ciências Farmacéuticas	KERLY FERNANDA MESQUITA PASQUALOTO
VJFLOITRIR	Facultade de Clancias Pachacécteras

Ligantes	Volume (أبه)	Superfície (Ų)	ClogP	HOMO (●V)	LUMO (•V)	Dipolo (debye)
ldv107	1771,18	920,14	-2,94	-8,75	-1,47	9,54
ldv126	1684,79	881.67	-3,41	-9,29	-1,75	13,50
INHd19	1697,58	846.96	-5,63	-8,66	-1,25	13,16
Idv125	1653,94	854.98	-4,20	-9,13	-1,58	11,02
INHd41	1697,89	875,57	-3,27	-8,68	-1,14	10,96
INHd48	1645,99	814,11	-3,10	-9,08	-1,19	8.51
INHd49	1770,04	921,52	-2,74	-8,96	-0,52	12,97

Quadro 15 (continuação) – Propriedades dos ligantes no estado não-ligado incluídas na construção dos modelos de FEFF QSAR-3D

Superfície = área superficial; ClogP = coeficiente de partição n-octanol/água calculado por Ghose, Pritchett e Crippen (1998); HOMO = energia do orbital molecular ocupado de maior energia; LUMO = energia do orbital molecular desocupado de menor energia; aipolo = momento de dipolo.

A partir dos valores dos termos de energia obtidos com as simulações de DM à T 310 K para os ligantes e o receptor em seus estados ligado e não-ligado, os descritores do campo de força de energia livre (FEFF) (Tabela III, item **3.4.7.**) foram determinados e utilizados na construção dos modelos de FEFF QSAR-3D. A Tabela XVIII apresenta as medidas estatísticas, número de descritores e de compostos desviantes (*outliers*) para os cinco melhores modelos obtidos com 50.000 recombinações ("crossovers") e taxa de probabilidade de mutação de 10%. O valor do fator de ajuste ("smoothing factor") utilizado foi 0,56.

Tabela XVIII - Medidas estatísticas, número de descritores e número de outliers para os cinco melhores modelos de FEFF QSAR-3D

Modelos	Nº descritores	<b>I</b> 2	Q <sup>2</sup>	LSE	LOF	outlier(s)
1	6	0,79	0,61	0,28	0.60	4
2	5	0,74	0,55	0,35	0,63	1
3	5	0,74	0,55	0,35	0,64	0
4	5	0,74	0,55	0,35	0,64	0
5	5	0,74	0,55	0,35	0,65	2

Os valores de r<sup>2</sup> e de q<sup>2</sup> são menores do que os valores obtidos para os modelos da análise de QSAR-4D IR.

102

A abordagem de FEFF ("Free Energy Force Field") requer informações termodinâmicas de interação explícitas. Infelizmente, apenas um número pequeno de sistemas receptor-ligante apresenta suas energias livres de interação ( $\Delta$ Gs) disponíveis. Entretanto, a energia livre de interação pode ser estimada, de forma relativa, pela utilização de constantes de ligação e/ou constantes de inibição *in vitro*, como IC<sub>50</sub> (TOKARSKI, HOPFINGER, 1997b). Cabe lembrar que, neste estudo, as medidas de atividade biológica constituem dados qualitativos ou semiquantitativos (MIC) e isto, provavelmente, contribuiu significativamente para a diminuição de qualidade dos modelos resultantes.

O modelo 1, apesar de apresentar valores mais elevados de q<sup>2</sup> e r<sup>2</sup>, possui 4 compostos outliers e, quando estes desviantes foram eliminados e o programa WOLF 5.5 (ROGERS, 1994) foi novamente processado, os novos modelos resultantes também possuíam outliers. Por este motivo, o modelo 1 foi descartado.

Os modelos 2, 3, 4 e 5 apresentaram valores similares de medidas estatísticas e a matriz de correlação linear dos resíduos de ajuste entre pares de modelos foi computada para verificar se estes modelos eram ou não equivalentes (ROGERS, HOPFINGER, 1994; HOPFINGER, 199). Na Tabela XIX estão apresentados os dados da matriz de correlação linear dos resíduos de ajuste dos modelos de FEFF QSAR-3D resultantes.

Tabela XIX – Matriz de correlação linear dos resíduos de ajuste entre pares de modelos da análise de FEFF QSAR-3D

Modelos	<u> </u>	2	3	4	5
1	1,00				
2	0,40	1,00			
3	0,41	0,98	1,00		
4	0,39	0,98	0,99	1,00	
5	0,37	0,97	0,99	0,99	1,00

103

Conforme os dados da Tabela XIX, os modelos 2, 3, 4 e 5 estão intercorrelacionados (R = 1), apresentando coeficientes lineares de resíduos de ajuste elevados. Isto indica que estes modelos são equivalentes ou que existe apenas um modelo. Para escolher entre os quatro modelos equivalentes, que apresentaram medidas estatísticas similares (Tabela XVIII), optou-se pelo número de outliers. Os modelos 3 e 4 não apresentaram compostos desviantes e suas respectivas equações, (3) e (4), estão descritas abaixo:

$$pMIC = 2.33 + 0.001 (superficie_A - 342.28)^2 - 0.06 (E_L(LM) + 8.86)^2 + 0.10 (ClogP + 5.63)^2 - 0.003 (E_LR(RM) + 55.78)^2 - 0.0004 (\Delta E_{kors} - 48.62)^2$$

$$N = 30 \qquad q^2 = 0.55 \qquad r^2 = 0.74 \qquad LSE = 0.35 \qquad LOF = 0.64 \qquad (3)$$

$$pMIC = 2.36 + 0.001 (superficie_A - 343.46)^2 - 0.06 (E_L(LM) + 8.99)^2 + 0.10 (ClogP + 5.63)^2 - 0.003 (E_LR(RM) + 55.78)^2 - 0.0004 (\Delta E_{kors} - 48.62)^2$$

$$N = 30 \qquad q^2 = 0.55 \qquad r^2 = 0.74 \qquad LSE = 0.35 \qquad LOF = 0.64 \qquad (4)$$

As equações ou modelos (3) e (4) são praticamente iguais, reforçando o resultado indicado pela matriz de correlação linear dos resíduos de ajuste, com os mesmos descritores de energia e propriedades do ligante no estado não-ligado. Então, o modelo (3) foi utilizado para a predição dos valores de pMIC para o conjunto de avaliação.

O modelo (3) apresenta dois descritores não relacionados com o campo de força de energia livre (área de superfície e ClogP) e três descritores do campo de energia (energia de solvatação do ligante não-complexado ,  $E_L(LM)$ , energia de solvatação do receptor no complexo,  $E_{LR}(RM)$ , e, variação da energia torsional na complexação,  $\Delta E_{loss}$ ).

Cabe destacar que nos dois modelos, (3) e (4), não há dependência de átomos polares e, de acordo com os coeficientes, os descritores que contribuem na atividade biológica são a área de superfície e o coeficiente de partição calculado dos ligantes. Os compostos do conjunto de avaliação também foram submetidos às simulações de DM a 310 K como descrito anteriormente. Os valores dos termos de energia para os estados ligado e não-ligado estão apresentados nos quadros 16 e 17, respectivamente.

Quadro 16 - Valores dos termos de energia resultantes das simulações de DM de 10 ps a 310 K das conformações de energia mínima dos complexos enzima-ligante (E-L) para o conjunto de avaliação

Complexos E-L	Ent	Eder	Etors	E1.4	Ever	E	Enterviel	Esoty	Enb	Eror
Idv90	690,06	1242,86	371,84	1854,89	-786,06	-2009,83	-506,06	-69,15	-16066.69	-15278,14
ldv128	679,02	1216,02	367,5	1851,93	-788,89	-1988,23	-566,68	-63,93	-16122,76	-15416,02
Idv124	685.12	1248.09	380,13	1839,22	-779.57	-1988,03	-493,61	-77,54	-16203,89	-15390,08
Idv131	667.69	1231.42	380,27	1904,01	-790,85	-2053.63	-576,78	-74,55	-15991.66	-15304.08
ldv132	654,4	1198,62	387,11	1898,89	-745,54	-1926,92	-447.25	-74,89	-16013,52	-15069,1
Idv136	657,4	1296,86	403,61	1839,41	-799.84	-2017,06	-529,8	-70.04	-16233,56	-15453,02
INHd51	693,97	1262,24	375.67	1904,32	-748,36	-2109.37	-573.3	-70,83	-15639,92	-14905,58

deformação angular;  $E_{tors}$  = energia de torsão;  $E_{1,4}$  = energia de Interação do tipo 1,4;  $E_{vdw}$  = energia de van der Waals;  $E_{el}$  = energia eletrostática;  $E_{intervel}$  = energia intermolecular de van der Waals e eletrostática;  $E_{sav}$  = energia de solvatação;  $E_{tb}$  = energia de ligação de hidrogênio;  $E_{tor}$  = energia potencial total.

Quadro 17 - Valores dos termos de energia resultantes das simulações de DM de 50 ps a 310 K

Ligantes	Eest	Edel	Etors	E1.4	Evaw	Ea	Esot	Ень	Etor
ldv90	34,36	68,99	28,58	-414,40	-18,42	111,32	-15,91	-281,83	-487,31
ldv128	31,98	64,26	21,84	-407,70	-22,21	110.70	-12,96	-334,44	-548,54
ldv124	23,53	73,80	27,79	-407,21	-24,72	111,36	-13.27	-419,48	-628,20
Idv131	30,61	82,47	29,45	-415.04	-21,96	106,16	-13.93	-361,94	-564,17
ldv132	26,26	77,96	26,79	-408,13	-18,26	105,99	-14,63	-295,51	-499,53
Idv136	24.72	87,94	29,37	-409,67	-23,03	102,01	-13,17	-376.82	-578,64
INHd51	37,31	70,90	45,19	-388,24	-21.43	87,39	-11,99	-356.07	-536,94

das conformações de energia mínima dos ligantes do conjunto de avaliação

As unidades dos valores de energia são kcal/mol. E<sub>est</sub> = energia de deformação avai das ligações; E<sub>del</sub> = energia de deformação angular; E<sub>tors</sub> = energia de torsão; E<sub>1,4</sub> = energia de interação do tipo 1,4; E<sub>val</sub> = energia de van der Waals;
 E<sub>et</sub> = energia eletrostática; E<sub>ntervet</sub> = energia intermolecular de van der Waals e eletrostática; E<sub>tor</sub> = energia de solvatação; E<sub>tb</sub> = energia de ligação de hidrogênio; E<sub>tor</sub> = energia potencial total.

As propriedades dos ligantes (conformações de energia mínima resultantes das simulações de DM de 50 ps a 310 K) do conjunto de avaliação no estado nãoligado estão apresentadas no quadro 18.

Ligantes	Volume (Å)	Superfície (Å <sup>2</sup> )	ClogP	HOMO (eV)	LUMO (eV)	Dipolo (debye)
Idv90	1640,86	844,71	-3,70	-8,44	-1,51	12,93
ldv128	1599,96	806.89	-3,16	-9,10	-1,38	8,68
ldv124	1524,48	761,91	-4,45	-8,46	-1,73	17,80
ldv131	1582,81	796,07	-3,93	-8,69	-2,07	10,68
Idv132	1589.64	782,58	-3.93	-8,86	-1,30	3.48
ldv136	1648,10	821,95	-2.24	-8,93	-1,28	2,32
INHd51	1852,33	923,95	-3.22	-8.63	-0.88	8,78
uperfície = áre	a superficial; Clog	P = coeficiente de p	partição n-c	ctanol/água calc	ulado por Ghose	, Pritchett e Crippe

Quadro 18 – Propriedades dos ligantes do conjunto de avaliação no estado não-ligado

Os valores dos termos de energia do conjunto de avaliação foram utilizados para determinar os descritores do campo de força de energia livre (FEFF) (Tabela III, item **3.4.7**.) considerando o modelo de FEFF QSAR-3D selecionado, equação (3), para se obter os valores de atividade biológica (pMIC) e verificar a capacidade de predição do modelo 3.

A Tabela XX apresenta os valores de atividade biológica experimental (ATIV\_EXP) e prevista (ATIV\_PREV) para cada composto do conjunto de avaliação e os seus respectivos resíduos (diferença entre valores de atividade biológica calculada e experimental), considerando o modelo ou equação (3), da análise de FEFF QSAR-3D.

Superfície = área superficial; ClogP = coeficiente de partição n-octanol/água calculado por Ghose, Pritchett e Crippen (1998); HOMO = energia do orbital molecular ocupado de maior energia; LUMO = energia do orbital molecular desocupado de menor energia; dipolo = momento de dipolo.

ATIV_EXP	ATIV_PREV	Resíduos	Test_Set
4,22	5,78	-1,56	ldv90
2,82	4,26	-1,44	ldv128
2,70	4,53	-1,83	ldv124
2,00	3,53	-1,53	ldv131
1,82	3,35	-1,53	ldv132
0,52	1,94	-1,42	ldv136
0,22	5,63	-5,41	INHd51

Tabela XX - Predição de atividade biológica para os compostos do conjunto de avaliação

(Test\_Set)

As predições de atividade biológica para o conjunto de avaliação estão apresentadas na Tabela XX e pode-se constatar que o modelo 3 exibiu um poder de predição médio muito pobre, 28,6 %, considerando-se o valor de desvio padrão da média dos resíduos, 1,46. Cinco dos sete compostos (Idv90, Idv124, Idv131, Idv132, INHd51) apresentaram o valor de resíduo maior que o desvio da média. Este resultado não foi tão surpreendente considerando-se os valores das medidas estatísticas, q<sup>2</sup> e r<sup>2</sup>, do modelo 3.

Neste estudo, a comparação entre os modelos da análise de QSAR-4D IR e de FEFF QSAR-3D RD, sugerida por Venkatarangan e Hopfinger (1999), não foi possível porque os modelos resultantes são diferentes (coeficiente de correlação linear dos resíduos de ajuste = 0,25). Além disso, os dados disponíveis de atividade biológica (MIC) não eram os preconizados por Tokarski e Hopfinger (1997b) para a análise de FEFF. Somados aos aspectos mencionados estão, ainda, a natureza do sistema em estudo e a limitação da ferramenta computacional, visto que no tratamento de sistemas muito flexíveis (enzima) o investigador fica limitado à execução de simulações de DM muito curtas (10 ps), que, na maioria das vezes, não detectam os estados conformacionais mais aproximados àquele da interação.

#### 106

Conclusões

.

·

.

.

•

.

# 5. CONCLUSÕES

Considerando-se a hipótese de que o conjunto de hidrazidas investigado apresentaria mecanismo de ação similar ao do fármaco protótipo, isoniazida, na enzima InhA do *M. tuberculosis*, pode-se constatar com a análise de QSAR-4D IR que:

- a presença de grupos substituintes apolares no carbono (C49) do anel piridínico (ligante INH1) diminui consideravelmente a atividade biológica. Este tipo de substituinte, nesta posição, não permitiria a rotação da cadeia lateral do resíduo *Phe149* e, desta maneira, impediria a interação *m*-stacking entre o anéis piridínico e aromático do resíduo *Phe149*. Todavia, a presença de grupos apolares na hidroxila da segunda ribose da porção NAD do ligante permitiria melhor interação com os resíduos de aminoácido *Tyr158* e *Met103*, aumentando a atividade biológica de maneira significativa;

- a presença de grupos substituintes doadores e/ou aceptores de ligação de hidrogênio no átomo de nitrogênio da porção nicotinamida do ligante contribuiria para o aumento da atividade biológica porque permitiria interações com os resíduos de aminoácido *Pro193, lle194.* Da mesma forma, este tipo de substituintes no oxigênio entre os dois grupos fosfato do NAD do ligante contribuiriam para o aumento da aitvidade biológica, devido às possíveis interações com os resíduos de aminoácido *Thr196, Leu197, Thr17 e Ser20.* 

De acordo com as condições adotadas neste estudo, a análise de QSAR-4D IR mostrou-se ferramenta útil do CAMD para fornecer informações aproximadas no desenvolvimento de novos agentes potenciais de combate à tuberculose, apesar da medida dos dados de atividade biológica (MIC) não ser considerada adequada.

108

A metodologia de FEFF QSAR-3D constitui ferramenta importante à obtenção de parâmetros termodinâmicos de interação receptor-ligante, contribuindo de forma relevante ao planejamento racional de novas moléculas. Contudo, neste estudo, a análise de FEFF QSAR-3D foi prejudicada devido à natureza do sistema estudado e dos dados de atividade biológica. A técnica computacional ainda oferece certas limitações aos sistemas flexíveis (enzima) para o processamento das simulações de DM. O investigador fica limitado à execução de simulações de DM muito curtas, que, na maioria das vezes, não detectam o estado conformacional de interação mais adequado.

Considerando-se as condições assumidas neste estudo e a natureza do sistema estudado e dos dados de atividade biológica (MIC), entre as duas técnicas aplicadas, a análise de QSAR-4D IR forneceu os melhores resultados.

> BIBLIOTECA Faculdade de Ciências Farmacêuticas Universidade de São Paulo

KERLY FERNANDA MESQUITA PASQUALOTO

Perspectivas

.

.

.

.

.

. .

.

.

# 6. PERSPECTIVAS

Considerando-se os resultados da análise de FEFF QSAR-3D, como tentativa de melhorar a qualidade das informações obtidas, sugere-se que a síntese de algumas das hidrazidas do conjunto estudado seja efetuada e ensaios de inibição enzimática sejam realizados. Desta forma, serão obtidos dados de atividade biológica quantitativos e, conseqüentemente, mais indicados para a utilização em análises de campo de força de energia livre.

O conjunto de hidrazidas deste estudo também poderá ser utilizado para mapeamento do sítio ativo da enzima mutante (Ser94Ala) através da análise de QSAR-4D e, desta forma, poderão ser obtidas informações relevantes para o planejamento de inibidores de cepas MDR.

Considerando-se a homologia entre as enzimas enoil redutase do *M. tuberculosis* e da *E. coli*, estudos poderão ser realizados com a família experimental de diazoborinas, que agem na enzima enoil redutase da *E. coli*, utilizando-se a mesma metodologia adotada neste estudo para que informações estruturais adicionais sejam obtidas e aplicadas no planejamento de novos inibidores.

Visando melhorar o desempenho das análises realizadas neste estudo, ou seja, obter melhores modelos mesmo utilizando medidas de atividade biológica semi-quantitativas (MIC), sugere-se a inclusão de descritores significativos provenientes da análise de QSAR com dados de interação de membrana do *M. tuberculosis* (programa ADMET/MI-QSAR) em uma nova análise de QSAR-4D IR e de FEFF QSAR-3D RD com o mesmo conjunto de 37 hidrazidas.

# Referências Bibliográficas

.

:

.

.

.

.

.

.

.

# 7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALBUQUERQUE, M.G., HOPFINGER, A.J., BARREIRO, E.J., ALENCASTRO, R.B. Fourdimensional quantitative structure-activity relationship analysis of a series of interphenylene 5-oxabicycloheptane oxazole thromboxane A<sub>2</sub> receptor antagonists. J. Chem. Inf. Comput. Sci., v. 38, p. 925-938, 1998.
- ALLINGER N.L. Conformational analysis 130: MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J. Am. Chem. Soc., v. 99, p. 8127-8134, 1977.
- BALDOCK, C., de BOER, G.J. de, RAFFERTY, J.B., STUITJE, A.R., RICE, D.W. Mechanism of action of diazaborines. *Biochem. Pharmacol.*, v. 55, p. 1541-1549, 1998.
- BALDOCK, C., RAFFERTY, J.B., SEDELNIKOVA, S.E., BAKER, P.J., STUITJE, A. R., SLABAS, A.R., HAWKES, T.R., RICE, D.W. A mechanism of drug action revealed by structural studies of enoyl reductase. Science, v. 274, p. 2107-2110, 1996.
- BARRY, III, C.E. New horizons in the treatment of tuberculosis. *Biochem. Pharmacol.*, v. 54, p. 1165-1172, 1997.
- BARRY, III, C.E., LEE, R.E., MDLULI, K., SAMPSON, A.E., SCHROEDER, B.G., SLAYDEN, R.A., YUAN, Y. Mycolic acids: structure, biosynthesis and physiological functions. *Prog. Lipid Res.*, v. 37, p. 143-179, 1998.
- BASSO, L.A., ZHENG, R., MUSSER, J.M., Jr. JACOBS, W.R., BLANCHARD, J.S. Mechanisms of isoniazid resistance in Mycobacterium tuberculosis: enzimatic characterization of enoyl reductase mutants identified in isoniazid-resistant clinical isolates. J. Infect. Dis., v. 178, p. 769-775, 1998.
- BERENDSEN, H.J.C., POSTMAN, J.P.M., VAN GUNSTEREN, W.F., DI NOLA, A., HAAK, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys., v. 81, p. 3684-3690, 1984.
- BERGLER, H., FUCHSBICHLER, S., HÖGENAUER, G., TURNOWSKY, F. The enoyl-[acylcarrier-protein] reductase (Fabl) of *Escherichia* coli, which catalyzes a key

113

regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *Eur. J. Biochem.*, v. 242, p. 689-694, 1996.

- BERGLER, H., HÖGENAUER, G., TURNOWSKY, F. Sequences of the *envM* gene and two mutated alleles in *Escherichia coli*. J. Gen. Microbiol., v. 138, Pt. 10, p. 2093-2100, 1992.
- BERNSTEIN, J., JAMBOR, W.P., LOTT, W.A., PANSY, F., STEINBERG, B.A., YALE, H.L. Chemotherapy of experimental tuberculosis – VI. Derivatives of isoniazid. *Am. Rev. Tuberc.*, v. 67, p. 354-365, 1953a.
- BERNSTEIN, J., JAMBOR, W.P., LOTT, W.A., PANSY, F., STEINBERG, B.A., YALE, H.L. Chemotherapy of experimental tuberculosis – VII. Heterocyclic acid hydrazides and derivatives. Am. Rev. Tuberc., v. 67, p. 366-375, 1953b.
- BERNSTEIN, J., LOTT, W.A., STEINBERG, B.A., YALE, H.L. Chemotherapy of experimental tuberculosis – V. Isonicotinic acid hydrazide (nydrazid) and related compounds. *Am. Rev. Tuberc.*, v. 65, p. 357-364, 1952.
- BHARGAVA, H.N., LEONARD, P.A. Triclosan: applications and safety. Am. J. Infect. Control., v. 24, p. 209-218, 1996.
- BLOCH, K. Control mechanisms for fatty acid synthesis in Mycobacterium smegmatis. Adv. Enzymol. Relat. Areas Mol. Biol., v. 45, p. 1-84, 1977.
- BRENAN, P.J., NIKAIDO, H. The envelope of mycobacteria. Annu. Rev.Biochem., v. 64, p. 29-63, 1995.
- COHEN, N.C. Guidebook on molecular modeling in drug design. San Diego: Academic Press, 1996. 361 p.
- CORNELL, W.D., CIEPLAK, P., BAYLY, C.I., GOULD, I.R., MERZ, K.M., Jr., FERGUSON, D.M., SPELLMEYER, D.C., FOX, T., CALDWELL, J.W., KOLLMAN, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc., v. 117, p. 5179-5197, 1995.

- CRAMER, III, R.D., PATTERSON, D.E., BUNCE, J.D. Comparative molecular field analysis (COMFA) 1: Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc., v.110, p. 5959-5967, 1988.
- DE BOER, G.J., PIELAGE, G.J.A., NIJKAMP, N.I.J., SLABAS, A.R., RAFFERTY, J.B., BALDOCK, C., RICE, D.W., STUITJE, A.R. Molecular genetic analysis of enoyl-acyl carrier protein reductase inhibbition by diazaborine. Mol. Microbiol., v. 31, p. 443-450, 1999.
- DESSEN, A., QUÉMARD, A., BLANCHARD, J.S., Jr. JACOBS, W.R., SACCHETTINI, J.C. Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. Science, v.267, p. 1638-1641, 1995.
- DEWAR, M. J. S. E., ZOEBISCH, G., HEALY, E. F., STEWART, J.J.P. AM1: A new general purpose quantum mechanical molecular model. J. Am. Chem. Soc., v. 107, p. 3902-3909, 1985.
- DOHERTY, D. MOLSIM: Molecular Mechanics and Dynamics Simulation Software, version 3.0 – User's Guide. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 1994.
- FORSYTHE, K.H., HOPFINGER, A.J. The influence of solvent on the secondary structures of poly (L-alanine) and poly (L-proline). *Macromolecules*, v. 6, p. 423-437, 1973.
- FRIEDMAN, J. Multivariate adaptative regression splines. *Technical Report No. 102*, Laboratory for Computational Statistics, Departament of Statistics, Stanford University. Stanford, 1988.
- FULCO, A.J. Fatty acid metabolism in bacteria. Prog. Lipid Res., v. 22, p. 133-160, 1983.
- GEORGE, K.M., YUAN, Y., SHERMAN, D.R., BARRY, III, C.E. The biosynthesis of cyclopropanated mycolic acids in Mycobacterium tuberculosis. Identification and functional analysis of CMAS-2. J. Biol. Chem., v. 270, p. 27292-27298, 1995.

- GHOSE, A.K., PRITCHETT, A., CRIPPEN, G.M. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. J. Comput. Chem., v. 9, p. 80-90, 1988.
- GLEN, W.G., DUNN, III, W.J., SCOTT, D.R. Principal components analysis and partial least-squares regression. Tetrahedron Comput. Methods, v. 2, p. 349-354, 1989.
- GOLBRAIKH, A., TROPSHA, A. Beware of q<sup>2</sup>! J. Mol. Graph. Model., v. 20, p. 269-276, 2002.
- HEATH, R.J., YU, Y.T., SHAPIRO, M.A., OLSON, E., ROCK, C.O. Broad spectrum antimicrobial biocides target the Fabl component of fatty acid synthesis. J. Biol. Chem., v. 273, p. 30316-30320, 1998.
- HOLLAND, J. Adaptation in artificial and natural systems. Ann Arbor: University of Michigan Press, 1975.
- HOPFINGER, A.J. 4D-QSAR Software, v. 3.0 User's Manual. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 1999.
- HOPFINGER, A.J. Conformational properties of macromolecules. New York: Academic Press, 1973. p. 71.
- HOPFINGER, A.J., PEARLSTEIN, R.A. Molecular mechanics force-field parametrization procedures. J. Comput. Chem., v. 5, p. 486-499, 1984.
- HOPFINGER, A.J., WANG, S., TOKARSKI, J.S., JIN, B., ALBUQUERQUE, M., MADHAV, P.J., DURAISWAMI, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J. Am. Chem. Soc., v.119, p. 10509-10524, 1997.

HyperChem Program Release 6.0 for Windows. Hybercube, Inc., 1996.

- INTERCALC Software, The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA.
- KATER, M.M., KONINGSTEIN, G.M., NIJKAMP, H.J., STUITJE, A.R. The use of a hybrid genetic system to study the functional relationship between prokaryotic and

plant multi-enzyme fatty acid synthetase complexes. *Plant Mol. Biol.*, v. 25, p. 771-790, 1994.

- KATER, M.M., KONINGSTEIN, G.M., NIJKAMP, H.J.J., STUITJE, A.R. cDNA cloning and expressionn of Brassica napus enoyl-acyl carrier protein reductase in Escherichia coli. Plant Mol. Biol., v. 17, p. 895-909, 1991.
- KIKUCHI, S., RAINWATER, D.L., KOLATTUKUDY, P.E. Purification and characterization of an unusually large fatty acid synthase from Mycobacterium tuberculosis var. bovis BCG. Arch. Biochem. Biophys., v. 295, p. 318-326, 1992.
- KLOPMAN, G., FERCU, D., JACOB, J. Computer-aided study of the relationship between structure and antituberculosis activity of a series of isoniazid derivatives. *Chemical Physics*, v. 204, p. 181-193, 1996.
- KOLATTUKUDY, P.E., POULOSE, A.J., BUCKNER, J.S. Fatty acid synthase from the uropygial gland of goose. *Methods Enzymol.*, v. 71, Pt. C, p. 103-109, 1981.
- KOROLKOVAS, A. Essentials of medicinal chemistry, 2. ed. New York: Wiley-Interscience, 1988. 1216 p.
- KRASOWSKI, M.D., HONG, X., HOPFINGER, A.J., HARRISON, N.L. 4D-QSAR analysis of a set of propofol analogues: mapping binding sites for an anesthetic phenol on the GABA(A) receptor. J. Med. Chem., v.45, p. 3210-3221, 2002.
- KUBINYI, H. QSAR: Hansch analysis and related approaches. In: MANNHOLD, R., KROGSGAARD-LARSEN, P., TIMMERMAN, H., Eds. Methods and Principles in Medicinal Chemistry. New York: VHC, 1993. chap. 5, p. 91.
- LARSSON, L., JIMÉNEZ, J., VALERO-GUILLÉN, P., MARTIN-LUENGO, F., KUBIN, M. Establishment of 2-docosanol as a cellular marker compound in the identification of Mycobacterium xenopi. J. Clin. Microbiol. v. 27, p. 2388-2390, 1989.
- LEE, R.E., BRENNAN, P.J., BESRA, G.S. Mycobacterium tuberculosis cell envelope. Curr. Top. Microbiol. Immunol., v. 215, p. 1-27, 1996.

- LEVY, C.W., BALDOCK, C., WALLACE, A.J., SEDELNIKOVA, S., VINER, R.C., CLOUGH, J.M., STUITJE, A.R., SLABAS, A.R., RICE, D.W., RAFFERTY, J.B. A study of the structureactivity relationship for diazaborine inhibition of *Escherichia coli* enoyl-acp reductase. J. Mol. Biol., v. 309, p. 171-180, 2001.
- LIU, J., BARRY, III, C.E., BESRA, G.S., NIKAIDO, H. Mycolic acid structure determines the fluidity of the mycobacterial cell wall. *J. Biol. Chem.*, v. 271, p. 29545-29551, 1996.
- LUQUIN, M., AUSINA, V., CALAHORRA, F.L., BELDA, F., BARCELO, M.G., CELMA, C. Evaluation of practical chromatographic procedures for identification of clinical isolates of mycabacteria. J. Clin. Microbiol., v. 29, p. 120-130, 1991.
- MAGNUSON, K., JACKOWSKI, S., ROCK, C.O., CRONAN, J.E., Jr. Regulation of fatty acid biosynthesis in Escherichia coli. Microbiol. Rev., v. 57, p. 522-542, 1993.
- MANNHOLD, R., van de WATERBEEMD, H. Substructure and whole molecule approaches for calculating log P. J. Comput. Aided Mol. Des., v. 15, p. 337-354, 2001.
- MARRAKCHI, H., LANÉELLE, G., QUÉMARD, A. InhA, a target of the antituberculosis drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS II. *Microbiology*, v. 146, pt. 2, p. 289-296, 2000.
- McCARTHY, A.D., HARDIE, D.G. Fatty acid synthase an example of protein evolution by gene fusion. *Trends Biochem. Sci.*, v. 9, p. 60-63, 1984.
- MCMURRY, L.M., OETHINGER, M., LEVY, S.B. Triclosan targets lipid synthesis. Nature, v. 394, p. 531-532, 1998.
- OOMS, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.*, v. 7, p. 141-158, 2000.
- OSFRAME Software, The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA.

- PARIKH, S., MOYNIHAN, D.P., XIAO, G., TONGE, P.J. Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-acp reductase from Mycobacterium tuberculosis. Biochemistry, v. 38, p. 13623-13634, 1999.
- PARIKH, S.L., XIAO, G., TONGE, P.J. Inhibition of InhA, the enoyl redutase from Mycobacterium tuberculosis, by triclosan and isoniazid. *Biochemistry*, v. 39, p. 7645-7650, 2000.
- PASQUALOTO, K.F.M., FERREIRA, E.I. An approach for the rational design of new antituberculosis agents. *Curr. Drug Targets*, v. 2, p. 427-437, 2001.
- QUÉMARD, A., SACCHETTINI, J.C., DESSEN, A., VILCHÈZE, C., BITTMAN, R., Jr. JACOBS, W.R., BLANCHARD, J.S. Enzymatic characterization of the target for isoniazid in Mycobacterium tuberculosis. Biochemistry, v. 34, p. 8235-8241, 1995.
- QURESHI, N., SATHYAMOORTHY, N., TAKAYAMA, K. Biosynthesis of C30 to C56 fatty acids by an extract of Mycobacterium tuberculosis H37Ra. J. Bacteriol., v. 157, p. 46-52, 1984.
- RATTAN, A., KALIA, A., AHMAD, N. Multidrug-resistant Mycobacterium tuberculosis: molecular perspectives. Emerg. Infect. Dis., v. 4, p. 195-209, 1998.
- RAKE, G., JAMBOR, W., McKEE, C.M., PANSY, F., WISELOGIE, F.Y., DONOVICK, R. The use of mouse in a standardized test for antituberculous activity of compounds of natural or synthetic origin: III. The standardized test. Am. Rev. Tuberc., v. 60, p. 121, 1949.
- ROCK, C.O., CRONAN, J.E. Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. Biochim. Biophys. Acta, v.1302, p. 1-16, 1996.

ROGERS, D. Evolutionary Statistics: Using a genetic algorithm and model reduction to

<sup>5</sup> isolate alternate statistical hypotheses of experimental data. In: Proceedings of the Seventh International Conference on Genetic Algorithmos, East Lansing, MI, Morgan-Kaufmann, San Francisco, CA, EUA, 1997. ROGERS, D. G/SPLINES: A hybrid of Friedman's multivariate adaptative regression splines (MARS) algorithm with Holland's genetic algorithm. In: *The Proceedings of the Fourth International Conference on Genetic Algorithms*. San Diego. 1991. p. 38-46.

ROGERS, D. WOLF Reference Manual Version 5.5. Molecular Simulation Inc., 1994.

- ROGERS, D., HOPFINGER, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J. Chem. Inf. Comput. Sci., v. 34, p. 854-866, 1994.
- ROGNAN, D., SOKOLOFF, P., MANN, A., MARTRES, M.P., SCHWARTZ, J.C., COSTENTIN, J., WERMUTH, C.G. Optically active benzamides as predictive tools for mapping the dopamine D<sub>2</sub> receptor. *Eur. J. Pharmacol.*, v. 189, p. 59-70, 1990.
- ROUJEINIKOVA, A., SEDELNIKOVA, S., de BOER, G.J., STUITJE, A.R., SLABAS, A.R., RAFFERTY, J.B., RICE, D.W. Inhibitor binding studies on enoyl redutase reveal conformational changes related to substrate recognition. *J. Biol. Chem.*, v. 274, p. 30811-30817, 1999.
- ROZWARSKI, D.A., GRANT, G.A., BARTON, D.H.R., Jr. JACOBS, W.R., SACCHETTINI, J.C. Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. Science, v. 279, p. 98-102, 1998.
- ROZWARSKI, D.A., VILCHÈZE, C., SUGANTINO, M., BITTMAN, R., SACCHETTINI, J.C. Crystal structure of the Mycobacterium tuberculosis enoyl-acp reductase, InhA, in complex with NAD<sup>+</sup> and a C16 fatty acyl substrate. J. Biol. Chem., v. 274, p. 15582-15589, 1999.
- SANT'ANNA, C.M. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). Quim. Nova, v. 25, p. 505-512, 2002.
- SANTOS-FILHO, O.A., HOPFINGER, A.J. A search for sources of drug resistance by the 4D-QSAR analysis of a set of antimalarial dihydrofolate reductase inhibitors. J. Comput.-Aided Mol. Des., v. 15, p. 1-12, 2001.
- SANTOS-FILHO, O.A., HOPFINGER, A.J. The 4D-QSAR paradigm: Application to a novel set of non-peptidic HIV protease inhibitors. Quant. Struct.-Act. Relat., v. 21, p. 368-381, 2002.
- SENSI, P. Approaches to the development of new antituberculosis drugs. *Rev. Infect. Dis.*, v. 11, suppl. 2, p. 467-470, 1989.
- SENSI, P., GRASSI, G.G. Antimycobacterial agents. In: BURGER, A., WOLFF, M.E. Burger's Medicinal Chemistry and Drug Discovery. 5. ed. New York: John Wiley & Sons, Inc., 1996. v. 2, chap. 34, p. 575-635.
- SMITH, S. The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. FASEB J., v. 8, p. 1248-1259, 1994.
- STEWART, M.J., PARIKH, S., XIAO, G., TONGE, P.J., KISKER, C. Structural basis and mechanism of enoyl redutase inhibition by triclosan. J. Mol. Biol., v. 290, p. 859-865, 1999.
- TOKARSKI, J.S., HOPFINGER, A.J. Constructing protein models for ligand-receptor binding thermodynamic simulations: An application to a set of peptidometic rennin inhibitors. J. Chem. Inf. Comput. Sci., v. 37, p. 779-791, 1997a.
- TOKARSKI, J.S., HOPFINGER, A.J. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: Application to a set of peptidometic rennin inhibitors. J. Chem. Inf. Comput. Sci., v. 37, p. 792-811, 1997b.
- TURNOWSKY, F., FUCHS, K., JESCHEK, C., HÖGENAUER, G. envM genes of Salmonella typhimurium and Escherichia coli. J. Bacteriol., v. 171, p. 6555-6565, 1989.
- VENKATARANGAN, P., HOPFINGER, A. Prediction of ligand-receptor free energy by 4D-QSAR analysis: Application to a set of glucose analogue inhibitors of glycogen phosphorylase. J. Chem. Inf. Comput. Sci., v. 39, p. 1141-1150, 1999.
- WAKIL, S.J., STOOPS, J.K., JOSHI, V.C. Fatty acid synthesis and its regulation. Annu. Rev. Biochem., v. 52, p. 537-579, 1983.

- WEINER, S.J, KOLLMAN, P.A., NGUYEN, D.T. An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem., v. 7, p. 230-252, 1986.
- WERMUTH, C.G. Strategies in the search for new lead compounds or original working hypotheses. In: WERMUTH, C.G. Ed. The Pratice of Medicinal Chemistry. San Diego: Academic Press, 1996. part. II, p. 81-99.
- WERMUTH, C.G., GANELLIN, C.R., LINDBERG, P., MITSCHER, L.A. Glossary of terms used in Medicinal Chemistry. Ann. Rep. Med. Chem., v. 33, p. 385-395, 1998.
- WOLINSKY, E. Tuberculosis. In: WYNGAARDEN, J.B., SMITH JR., L.H., BENNETT, J.C. Cecil Textbook of Medicine. 19. ed. Philadelphia: W. B. Sounders Company, 1992. v. 2, chap. 332, p. 1733-1742.
- WORLD HEALTH ORGANIZATION. An expanded DOTS framework for effective tuberculosis control. WHO/CDS/TB/2002.297. 2002c [www.who.int/gtb/dots/2002-297.htm].
- WORLD HEALTH ORGANIZATION. Country Profiles: Brazil. Annual Report, 1997. [www.who.int/gtb/publications/tbrep\_97/countries/brazil.htm].
- WORLD HEALTH ORGANIZATION. Global tuberculosis control: surveillance, planning, financing. WHO Report 2002. WHO/CDS/TB/2002.295. 2002d [www.who.int/gtb/publications/globrep02/index.html].
- WORLD HEALTH ORGANIZATION. Report of a WHO workshop on Tuberculosis Control and Medical Schools. Italy. WHO/TB/98.236. 1998 [www.who.int/gtb/publications/tb\_control/index.htm].

WORLD HEALTH ORGANIZATION. Tuberculosis. Fact Sheet nº104 - Revised in August 2002. 2002a [www.who.int/mediacentre/factsheets/who104/en/index.htm]. WORLD HEALTH ORGANIZATION. What is DOTS? A guide to understanding the WHO-

recommended TB control strategy knows as DOTS. WHO/CDS/CPC/TB/99.270.

1999 [www.who.int/gtb/publications/whatisdots/index.htm].

WORLDHEALTHORGANIZATION.WorldTBdayhighlightsreport.WHO/CDS/STB/2002.20.2002b

[stoptb.org/world.tb.day/WTBD\_2002/Final\_Highlights\_Report\_2002.pdf].

YAUN, Y., BARRY, III, C.E. A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA*, v. 93, p. 12828-12833, 1996.