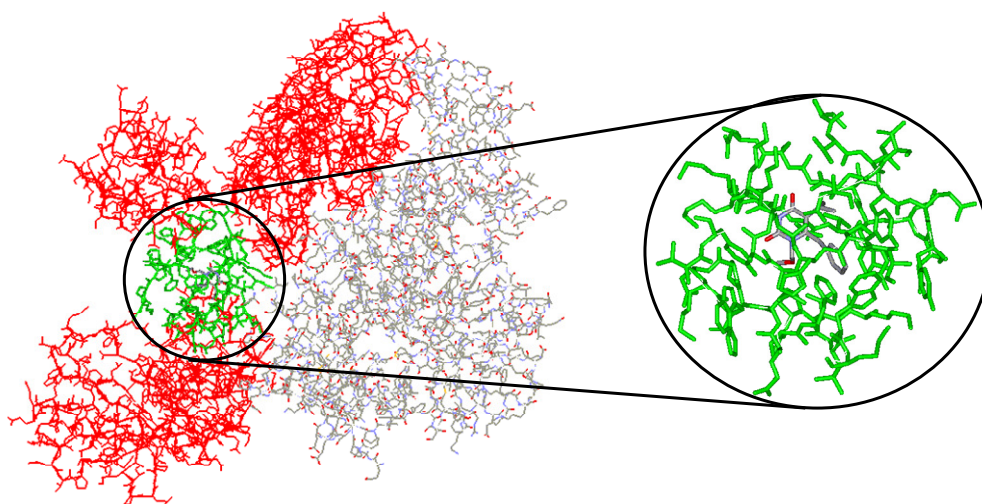




Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Química
Programa de Pós-Graduação em Química

QSAR-3D de Inibidores Não-Nucleosídeos
da Transcriptase Reversa do HIV-1:
Estudos Independente e Dependente da Enzima

Monique Araújo de Brito



Rio de Janeiro - Brasil

Setembro de 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**QSAR-3D de Inibidores Não-Nucleosídeos da Transcriptase
Reversa do HIV-1: Estudos Independente e Dependente da Enzima**

Monique Araújo de Brito

**Tese de Doutorado apresentada ao Programa de Pós-Graduação em Química
(Química Orgânica), Instituto de Química, Universidade Federal do Rio de
Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor
em Ciências.**

Orientadores:

Magaly Girão Albuquerque (Instituto de Química, UFRJ)

e

Carlos Rangel Rodrigues (Faculdade de Farmácia, UFRJ)

Rio de Janeiro – Brasil

Setembro de 2008

FICHA CATALOGRÁFICA

BRITO, Monique Araújo de

QSAR-3D de Inibidores Não-Nucleosídeos da Transcriptase Reversa do HIV-1:
Estudos Independente e Dependente da Enzima

Rio de Janeiro, UFRJ, Instituto de Química, 2008, 231 fls.

Tese: Doutor de Ciências (Química Orgânica)

- | | |
|------------------|-----------------------------------|
| 1. QSAR-3D | 3. Transcriptase Reversa do HIV-1 |
| 2. S- e NH-DABOs | 4. Tese |

I. Universidade Federal do Rio de Janeiro – IQ

II. Título

**QSAR-3D de Inibidores Não-Nucleosídeos da Transcriptase
Reversa do HIV-1: Estudos Independente e Dependente da Enzima**

Monique Araújo de Brito

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Química (Química Orgânica), Instituto de Química, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Ciências.

Rio de Janeiro, 22 de setembro de 2008.

Dra. Magaly Girão Albuquerque (IQ-UFRJ) (Presidente)

Dr. Carlos Rangel Rodrigues (FF-UFRJ) (Orientador)

Dr. José Daniel Figueroa Villar (IME)

Dr. Carlos Maurício Rabello de Sant'Anna (ICE-UFRRJ)

Dra. Alice Maria Rolim Bernardino (IQ-UFF)

Dr. Joaquim Fernando Mendes da Silva (IQ-UFRJ)

Dra. Nanci Câmara de Lucas Garden (IQ-UFRJ) (Suplente)

Dra. Helena Carla Castro (IB-UFF) (Suplente)

Esta Tese é dedicada aos Professores **Carlos Frederico de Souza Castro** (CEFET-GO) e **Carlos Maurício Rabello de Sant'Anna** (UFRRJ-RJ), aos quais não pude agradecer de forma adequada na oportunidade da dissertação de Mestrado.

AGRADECIMENTOS

Ao curso de Pós-Graduação em Química (Química Orgânica) da UFRJ.

Aos órgãos de apoio à pesquisa, CAPES, CNPq e FAPERJ.

À Profa. Dra. **Magaly Girão Albuquerque**, pela orientação sempre presente, pelos esclarecimentos na redação da tese e dos trabalhos científicos, pelo exemplo de correção de teses, pelas discussões teóricas, pela calma nos momentos difíceis e pela amizade.

Ao Prof. Dr. **Carlos Rangel Rodrigues**, pelo exemplo de força e determinação, pela amizade e por me dar oportunidades de crescer cientificamente com a colaboração em seus trabalhos.

Ao Prof. Dr. **Ricardo Bicca de Alencastro**, pelo exemplo de dedicação, pela disponibilização dos inúmeros livros de modelagem molecular e estrutura de proteínas, pela ajuda financeira nos congressos e encontros científicos, e pelo casaco que ajudou a tornar suportável o frio do laboratório.

À Profa. Dra. **Helena Carla Castro**, pelas oportunidades de desenvolvimento científico com a colaboração em seus trabalhos.

Aos Profs. das disciplinas que freqüentei no doutorado (em ordem alfabética), **Cristian Follmer** (DFQ-UFRJ), **Gilberto Domont** (DB-UFRJ), **Joaquim Fernando Mendes da Silva** (DQO-UFRJ), **Pedro Geraldo Pascutti** (IB-UFRJ) e **Pierre Mothé Esteves** (DQO-UFRJ). Agradeço pela oportunidade de aprender com vocês.

Ao Dr. **José Jair Vianna Cirino**, pela ajuda com o Gromacs, pelo incentivo, amizade e discussões teóricas.

Ao meu marido **Diego Evaristo de Lacerda**, por todo o incentivo, por me dar força para começar essa jornada, pelas dicas de Linux e de Gromacs, pela ajuda na formatação final da tese e pela paciência nos momentos difíceis. Te amo!

Aos meus queridos pais, **José Cunegundes Neto** e **Terezinha de Jesus Araújo de Brito**, e irmão, **José Cunegundes Araújo de Brito**, pelo estímulo e apoio durante esses quatro anos. Mãe, você foi, é e será minha maior incentivadora. Obrigada por acreditar sempre em mim. Mano, essa Tese é mais um motivo de orgulho para você!

Aos meus sogros **José Rodrigues de Lacerda** e **Francisca Evaristo de Lacerda**, que me deram a maior força nesse período.

Aos amigos do LabMMol pelo companheirismo, convivência harmoniosa e discussões produtivas no decorrer desses quatro anos (em ordem alfabética), **Ana Carolina Sodero**, **Bruno Horta**, **César Oliveira**, **Felipe Fleming**, **Isabella Guedes**, **Jocley Araújo**, **Lucas Hoelz**, **Rafael Silva** e **Samuel Pitta**.

Aos amigos do ModMolQSAR, **Ilídio Afonso** e **Uiaran Magalhães**. É um prazer trabalhar com vocês, espero que nossos caminhos estejam próximos.

Às amigas do LaBioMol da UFF pelos bons momentos nos trabalhos em colaborações e pela amizade e companheirismo, **Alessandra Mendonça Teles** e **Paula Alvarez Abreu**.

Aos amigos **André Borsato** e **Louise Quintino**, por compreenderem as inúmeras vezes que eu não pude me reunir.

À amiga **Ilana T. Balassiano**, que me incentivou tantas vezes.

Aos amigos da Pós-Graduação, pela convivência construtiva.

À banca examinadora, por aceitar o convite e pelas contribuições.

“Man’s mind once stretched by a new
idea, never regains its original dimension”

Oliver Wendell Holmes

(Médico, escritor e poeta americano, 1809-1894)

RESUMO

QSAR-3D DE INIBIDORES NÃO-NUCLEOSÍDEOS DA TRANSCRIPTASE REVERSA DO HIV-1: ESTUDOS INDEPENDENTE E DEPENDENTE DA ENZIMA

Monique Araújo de Brito

Orientadores: Magaly Girão Albuquerque (Instituto de Química - UFRJ)
Carlos Rangel Rodrigues (Faculdade de Farmácia - UFRJ)

Estudos de correlação quantitativa entre estrutura química tridimensional e atividade biológica (QSAR-3D) independente e dependente do receptor foram aplicados a uma série de 74 derivados diidro-alcoxi-benzil-4-oxopirimidínicos (DABOs) sintetizados e avaliados farmacologicamente como inibidores não-nucleosídeos da transcriptase reversa do HIV-1 (Ragno *et al.*, *J. Med. Chem.*, 47, 928-934, 2004; Mai *et al.*, *J. Med. Chem.*, 42, 619-627, 1999; Mai *et al.*, *J. Med. Chem.*, 40, 1447-1454, 1997; Mai *et al.*, *J. Med. Chem.*, 38, 3258-3263, 1995). A primeira abordagem de QSAR-3D, independente do receptor, aplicou a metodologia de Análise Comparativa do Campo Molecular (CoMFA), que permite a construção de modelos farmacofóricos utilizando como descritores as energias de interação estéricas e eletrostáticas dos inibidores com um átomo de prova que mimetiza átomos do receptor. O melhor modelo de CoMFA apresentou boa capacidade preditiva ($q^2=0,691$). A segunda abordagem, dependente do receptor, utilizou a estrutura da enzima selvagem e foi realizada através de uma metodologia de algoritmos genéticos, tendo como descritores as energias de interação estéricas e eletrostáticas dos inibidores com resíduos selecionados da enzima, obtidas por simulação por dinâmica molecular. O melhor modelo de QSAR-3D dependente do receptor também apresentou boa capacidade preditiva ($q^2=0,660$). Os modelos de QSAR-3D obtidos foram os primeiros da literatura com os derivados DABOs e possibilitaram o planejamento e a proposição de um novo derivado desta classe, como potencial inibidor da transcriptase reversa do HIV-1 no tratamento da AIDS.

Palavras-Chave: CoMFA, QSAR-3D Dependente do Receptor, Inibidores Não-Nucleosídeos, Transcriptase Reversa, HIV-1, AIDS.

Rio de Janeiro - Brasil

Setembro de 2008

ABSTRACT

3D-QSAR RECEPTOR-INDEPENDENT AND RECEPTOR-BASED STUDIES OF HIV-1 REVERSE TRANSCRIPTASE NON-NUCLEOSIDE INHIBITORS

Monique Araújo de Brito

**Advisers: Magaly Girão Albuquerque (Instituto de Química - UFRJ)
Carlos Rangel Rodrigues (Faculdade de Farmácia - UFRJ)**

Receptor-independent and receptor-based studies of three-dimensional quantitative structure-activity relationship (3D-QSAR) was applied to a series of 74 dihydro-alcoxi-benzyl-4-oxypyrimidines (DABOs) derivatives synthesized and evaluated pharmacologically as non-nucleoside Reverse Transcriptase inhibitors of the HIV-1 (Ragno *et al.*, *J. Med. Chem.*, 47, 928-934, 2004; Mai *et al.*, *J. Med. Chem.*, 42, 619-627, 1999; Mai *et al.*, *J. Med. Chem.*, 40, 1447-1454, 1997; Mai *et al.*, *J. Med. Chem.*, 38, 3258-3263, 1995). The first approach of 3D-QSAR was receptor-independent through the methodology Comparative Molecular Field Analysis (CoMFA), which allows the construction of pharmacoforic models using as descriptors the steric and electrostatic interaction energies of inhibitors with a probe atom that mimics the receptor atoms. The best CoMFA model presented good predictive ability ($q^2 = 0.691$). The second approach was receptor-dependent, using the wild-type enzyme through the Genetic Algorithm methodology, and using as descriptors the steric and electrostatic interaction energies of inhibitors with selected residues of the enzyme, obtained by molecular dynamic simulation. The best receptor-based 3D-QSAR model also presented good predictive ability ($q^2 = 0.660$). The resulting 3D-QSAR models were the first ones in the literature using only DABO derivatives and allowed the planning and proposals of new derivatives of the DABOs class as reverse transcriptase inhibitors in the AIDS treatment.

Key-words: CoMFA, Receptor-Based 3D-QSAR, Non-Nucleoside Inhibitors, Reverse Transcriptase, HIV-1, AIDS.

Rio de Janeiro - Brazil

September, 2008

LISTA DE FIGURAS

Figura 1. Mapa-múndi dividido por regiões indicando o número estimado de pessoas (crianças e adultos) com o HIV. (A) Total de infectados até 2007. (B) Total de infectados no ano de 2007 (Reproduzido de AIDS Epidemic Update, 2007).

Figura 2. Organização estrutural esquemática do vírus HIV (Adaptada de www.mclld.co.uk/hiv).

Figura 3. Esquema do ciclo replicativo do HIV (Adaptada de <http://www.tthivclinic.com/lifecycle4.htm>).

Figura 4. Estrutura 3D da transcriptase reversa do HIV-1. (A) Representação da estrutura secundária em modelo de fita (hélices-alfa em vermelho e folhas-beta em azul claro). (B) Representação das subunidades p66 (em vermelho) e p51 (colorido por elemento).

Figura 5. Estrutura da RT do HIV-1 destacando os sítios dos NRTIs e NNRTIs. Em azul os resíduos catalíticos (**Asp110**, **Asp185** e **Asp186**) e em vermelho alguns resíduos do sítio alostérico (**Lys101**, **Lys103**, **Val106**, **Tyr181** e **Tyr188**), todos em modelo CPK.

Figura 6. Etravirina (em bastão amarelo) no NNBS (Código PDB 1SV5). Os resíduos nomeados (em bastão) são os mais suscetíveis a mutações. Todos estão coloridos por elemento, exceto **Asn103** (em verde) que é uma mutação de **Lys103**.

Figura 7. Estruturas de um HEPT e de um DABO.

Figura 8. Estruturas de O-, S- e NH-DABOs.

Figura 9. Esquema da rota sintética dos derivados DABOs.

Figura 10. Representação da caixa 3D na qual os compostos alinhados são inseridos no processo do CoMFA (Adaptado de Kubinyi, 1997).

Figura 11. Esquema do processo de geração das energias no CoMFA e tabelamento dos resultados (Reproduzido de Kubinyi, 1997).

Figura 12. Esquema do processo de validação cruzada pelo método *Leave-One-Out*. (Reproduzido de Kubinyi, 1997).

Figura 13. Esquema da operação de cruzamento dentro da técnica de algoritmos genéticos. **Var** representa um descritor no estudo de QSAR.

Figura 14. Esquema da operação de mutação dentro da técnica de algoritmos genéticos. **Var** representa um descritor no estudo de QSAR.

Figura 15. Perfil estrutural dos S- e NH-DABOs utilizados nos estudos de QSAR-3D.

Figura 16. Distribuição dos valores de atividade biológica (pIC_{50}) dos compostos do conjunto de treinamento (A) e do conjunto de teste (B).

Figura 17. Sobreposição entre as estruturas 3D do MKC-442 (átomos de carbono coloridos em verde) e do NH-DABO **59** (átomos de carbono coloridos em cinza). Para melhor visualização, os átomos de hidrogênio foram omitidos.

Figura 18. Composto de referência NH-DABO **59** marcado com asteriscos nos átomos usados para os três Alinhamentos testados.

Figura 19. Representação esquemática do recorte do complexo do inibidor MKC-442 com a enzima RT do HIV-1 (código no PDB 1RT1) usado para no cálculo das energias de interação ligante-enzima. À esquerda, é mostrada a estrutura da RT com as subunidades p66 (colorida em vermelho) e p51 (colorida por elemento) e o círculo delimita a região do raio de corte de 10 Å a partir do inibidor. No detalhe, à direita, encontram-se os resíduos da enzima compreendidos no recorte (coloridos em verde) e o inibidor (colorido por elemento).

Figura 20. Sobreposição simultânea entre os 59 compostos do conjunto de treinamento da série dos S- e NH-DABOs de acordo com o Alinhamento 1.

Figura 21. Gráficos dos valores de pIC_{50} observados *versus* preditos para os conjuntos de (A) treinamento (**1-59**) e de (B) teste (**60-74**) do melhor modelo de CoMFA obtido usando Alinhamento 1, cargas PM3 e valores padrões de corte de energia (30 kcal.mol^{-1}), átomo de prova (C_{sp}^3 , carga +1) e espaçamento da grade ($2,0 \text{ \AA}$).

Figura 22. Mapas de contorno do melhor modelo de CoMFA exemplificado para o composto **59** (em modelo bastão e colorido por elemento). (A) O mapa estérico mostra as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a potência. (B) O mapa eletrostático mostra as áreas onde grupos com alta densidade eletrônica ou carga negativa aumentam (vermelho) ou diminuem (azul) a potência.

Figura 23. Modo de ligação do NH-DABO **59** (em modelo bastão e colorido por elemento) no sítio não-nucleosídeo (NNBS) da RT do HIV-1. Todos os resíduos estão representados em linhas e coloridos por elemento, exceto **Tyr181** e **Tyr188** (próximo ao grupo benzila de **59**), que está representado em modelo bastão e colorido em amarelo, e **Asp110**, **Asp185** e **Asp186** (resíduos catalíticos), que estão representados em modelo bastão e coloridos em azul. Os átomos de hidrogênio foram omitidos para melhor visualização.

Figura 24. Mapa de contorno estérico do melhor modelo de CoMFA para o composto **1** (em modelo bastão e colorido por elemento) mostrando as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a potência.

Figura 25. Mapa de contorno eletrostático do melhor modelo de CoMFA para o composto **50** (em modelo bastão e colorido por elemento), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a potência.

Figura 26. Gráfico da variação da energia potencial (kJ/mol) *versus* o tempo de simulação (ps), em temperatura constante ($T = 310 \text{ K}$), do complexo referente ao composto **59**.

Figura 27. A) Recorte da enzima RT do HIV-1 (modelo bastão, átomos de carbono da subunidade p66 em azul claro e da p51 em verde) mostrando os 53 resíduos compreendidos no raio de corte de 10 \AA , a partir do inibidor MKC-442 (modelo bastão-e-bola, átomos de carbono em cinza). B) Visão próxima do recorte mostrando apenas os resíduos compreendidos num raio de 5 \AA .

Figura 28. Representação gráfica 3D da **Eq.E (BD-I)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Ile94**, **Pro97**, **Lys101**, **Tyr181**, **Tyr188**, **His235** e **Asn137**) estão coloridos em verde e os que representam contribuições de Coulomb (**Gln182**, **Ser191** e **Pro226**), em azul claro. Os átomos de hidrogênio foram omitidos para melhor visualização.

Figura 29. Médias das energias de interação (kcal.mol^{-1}) dos compostos **1-74** com cada um dos termos selecionados na **Eq.E (BD-I)**.

Figura 30. Representação gráfica 3D da **Eq.E (BD-I)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento), destacando os resíduos (em modelo bastão e colorido por elemento) que podem sofrer mutação **Lys101**, **Tyr181** e **Tyr188** (Lennard-Jones). A seta indica uma possível interação por ligação hidrogênio entre **59** e **Lys101**. Em verde (Lennard-Jones) e em azul (Coulomb) estão representados os demais resíduos (em modelo bastão) da equação. Os átomos de hidrogênio foram omitidos para melhor visualização.

Figura 31. Gráfico de barras dos resíduos dos compostos dos conjuntos de (A) treinamento (**1-59**) e de (B) teste (**60-74**) de acordo com a **Eq.E (BD-I)**.

Figura 32. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($\text{pIC}_{50\text{Obs}} - \text{pIC}_{50\text{Pred}}$) dos compostos *outliers* da **Eq.E (BD-I)**.

Figura 33. Representação gráfica 3D da **Eq.J (BD-II)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) selecionados nesta equação (**Leu187**, **Asp237**, **Trp229**, **Glu224**, **Gly99**, **Phe227**, **Ser191**, **Tyr188**, **Thr139** e **Tyr183**) representam as interações de Lennard-Jones e Coulomb

Figura 34. Médias das energias de interação (kcal.mol^{-1}) dos compostos **1-74** para cada um dos termos selecionados na **Eq.J (BD-II)**. O gráfico foi truncado em $-10 \text{ kcal.mol}^{-1}$. A média da energia referente ao termo **Phe227** = $-29,037 \text{ kcal.mol}^{-1}$.

Figura 35. Gráfico de barras dos resíduos ($\text{pIC}_{50\text{Obs}} - \text{pIC}_{50\text{Pred}}$) dos compostos dos conjuntos de (A) treinamento (**1-59**) e de (B) teste (**60-74**) de acordo com a **Eq.J (BD-II)**.

Figura 36. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.J (BD-II)**.

Figura 37. Representação gráfica 3D da **Eq.L (BD-III)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento. Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Ile94**, **Tyr181**, **His235** e **Asn137**) estão coloridos em verde e os que representam contribuições de Coulomb (**Glu138** e **Pro225**), em azul claro. O resíduo **Glu224**, colorido em azul escuro, representa o somatório das contribuições de Lennard-Jones e Coulomb. Os átomos de hidrogênio foram omitidos para melhor visualização.

Figura 38. Médias das energias de interação ($kcal.mol^{-1}$) dos compostos **1-74** para cada um dos termos selecionados na **Eq.L (BD-III)**. O gráfico foi truncado em $10 kcal.mol^{-1}$. A média da energia referente ao termo **Glu138C**= $47,676 kcal.mol^{-1}$.

Figura 39. Gráfico de barras dos resíduos ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos dos conjuntos de (A) treinamento (**1-59**) e de (B) teste (**60-74**) de acordo com a **Eq.L (BD-III)**.

Figura 40. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.L (BD-III)**.

Figura 41. Representação gráfica 3D da **Eq.Q (BD-IV)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Lys103**, **Val179**, **Phe227** e **Asn136**) estão coloridos em verde e os que representam contribuições de Coulomb (**Gly99**, **Leu187**, **Tyr188**, **Pro225** e **Pro226**), em azul claro. Os átomos de hidrogênio foram omitidos para melhor visualização.

Figura 42. Médias das energias de interação ($kcal.mol^{-1}$) dos compostos **1-74** para cada um dos termos selecionados na **Eq.Q (BD-IV)**.

Figura 43. Gráfico de barras dos resíduos ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos dos conjuntos de (A) treinamento (**1-59**) e de (B) teste (**60-74**) de acordo com a **Eq.Q (BD-IV)**.

Figura 44. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.Q (BD-IV)**.

Figura 45. Estruturas dos NH-DABOs **76** e **77** (*R* e *S*) não incluídos nos conjuntos de treinamento e de teste para os quais foram feitas as predições teóricas das potências biológicas (pIC_{50}) aplicando as melhores equações (Bancos de Dados **I** a **IV**) de QSAR-3D dependente do receptor.

Figura 46. Modo de ligação do NH-DABO **76** (modelo bastão-e-bola, colorido por elemento) no NNBS da RT do HIV-1, destacando os resíduos (modelo bastão, colorido por elemento) **Lys101** e **Tyr181** relacionados aos termos correspondentes da **Eq.E**.

Figura 47. Proposição de novo NH-DABO por hibridação molecular entre o NH-DABO mais potente (**59**) e o fármaco NNRTI recém-lançado (etravirina) e por extensão de cadeia por homologia.

Figura 48. NH-DABO proposto (em bastão-e-bola, colorido por elemento) e possíveis interações com aminoácidos selecionados do NNBS (em bastão, coloridos por elemento)

LISTA DE TABELAS

Tabela 1. Nomes genéricos e comerciais, indústrias farmacêuticas envolvidas e anos de lançamento dos oito fármacos da classe dos NRTIs em uso clínico (Adaptada de Flexner, 2007).

Tabela 2. Principais classes de NNRTIs, abreviaturas e exemplos estruturais.

Tabela 3. Nomes genéricos e comerciais, estruturas químicas, indústrias farmacêuticas envolvidas e anos de lançamento dos quatro fármacos da classe dos NNRTIs em uso clínico (Adaptada de Flexner, 2007).

Tabela 4. Dimensões dos métodos de QSAR, principais características e possibilidade de representar a estrutura do receptor (Adaptada de Vedani *et al.*, 2006).

Tabela 5. Estruturas e potências inibitórias (pIC_{50} , M) dos derivados S- e NH-DABOs sobre a RT do HIV-1 (Ragno *et al.*, 2004; Mai *et al.*, 1999; Mai *et al.*, 1997; Mai *et al.*, 1995).

Tabela 6. Resultados estatísticos dos modelos de CoMFA obtidos testando três Alinhamentos (1, 2 e 3) e quatro tipos de cargas atômicas parciais (DFT, HF, AM1 e PM3). Opções padrões: valor de corte de energia (30 kcal.mol^{-1} para os campos estérico e eletrostático), átomo de prova (C_{sp}^3 , carga +1) e espaçamento da grade ($2,0 \text{ \AA}$).

Tabela 7. Resultados estatísticos dos modelos de CoMFA (Alinhamento 1 e cargas atômicas parciais PM3) obtidos testando três átomos de prova (C_{sp}^3 , carga +1; O_{sp}^3 , carga -1 e H, carga +1). Opções padrões: valor de corte de energia (30 kcal.mol^{-1} para os campos estérico e eletrostático) e espaçamento da grade ($2,0 \text{ \AA}$).

Tabela 8. Resultados estatísticos dos modelos de CoMFA (Alinhamento 1 e cargas atômicas parciais PM3) obtidos testando três valores de corte de energia (30 , 20 e 10 kcal.mol^{-1}) para os campos estérico e eletrostático. Opções padrões: átomo de prova (C_{sp}^3 , carga +1) e espaçamento da grade ($2,0 \text{ \AA}$).

Tabela 9. Valores de pIC_{50} (M) observados e preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) dos composto dos conjuntos de treinamento (1-59) e de teste (60-74) do melhor modelo de CoMFA (Alinhamento 1, cargas atômicas parciais PM3, átomo de prova C_{sp^3} e carga +1, valor de corte de 30 kcal.mol⁻¹ e espaçamento da grade de 2,0 Å).

Tabela 10. Resumo das características dos quatro bancos de dados (BD) (59 compostos do conjunto de treinamento) usados nas análises de QSAR-3D dependente do receptor.

Tabela 11. Resultados estatísticos dos cinco melhores modelos (A-E) de QSAR-3D (dependente do receptor) obtidos a partir do Banco de Dados I (BD-I).

Tabela 12. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (1-59) e de teste (60-74) de acordo com a Eq.E (BD-I).

Tabela 13. Matriz de correlação cruzada entre os descritores da Eq.E (BD-I).

Tabela 14. Resultados estatísticos dos cinco melhores modelos (F-J) de QSAR-3D (dependente do receptor) obtidos a partir do Banco de Dados II (BD-II).

Tabela 15. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (1-59) e de teste (60-74) de acordo com a Eq.J (BD-II).

Tabela 16. Matriz de correlação cruzada entre os descritores da Eq.J (BD-II).

Tabela 17. Resultados estatísticos dos quatro melhores modelos (K-N) de QSAR-3D (dependente do receptor) obtidos a partir do Banco de Dados III (BD-III).

Tabela 18. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (1-59) e de teste (60-74) de acordo com a Eq.L (BD-III).

Tabela 19. Matriz de correlação cruzada entre os descritores da Eq.L (BD-III).

Tabela 20. Resultados estatísticos dos cinco melhores modelos (**O-S**) de QSAR-3D dependente do receptor obtidos a partir do Banco de Dados IV (**BD-IV**).

Tabela 21. Valores de pIC_{50} observados, ϵ preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**) de acordo com a **Eq.Q** (**BD-IV**).

Tabela 22. Matriz de correlação cruzada entre os descritores da **Eq.Q** (**BD-IV**).

Tabela 23. Melhores equações dos Bancos de Dados **I** a **IV**.

Tabela 24. Resultados estatísticos das melhores equações dos Bancos de Dados **I** a **IV**.

Tabela 25. Compostos *outliers* (e respectivos valores residuais) identificados nas melhores equações dos Bancos de Dados **I** a **IV**.

Tabela 26. Matriz de correlação cruzada entre os valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) das melhores equações dos Bancos de Dados **I** a **IV**.

Tabela 27. Valores de pIC_{50} (M) observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os NH-DABOs **57**, **59**, **76** e **77** (*R*, *S* e racemato) de acordo com as melhores equações dos Bancos de Dados **I** a **IV**.

Tabela 28. Atividades preditas pelas melhores equações dos **BD-I** a **IV** para o NH-DABO proposto.

LISTA DE ABREVIATURAS E SIGLAS

ADAM – *Alkenyldiarylmethane* (Alquenil-diarilmetano)

α -APA – *Alpha-anilinophenylacetamide* (α -Anilino-fenilacetamida)

AM1 – *Austin Model 1*

AZT – *Azidothymidine* (Azidotimidina)

BHAP – *Bis(heteroaryl)piperazine* (Bis-heteroaril-piperazina)

COMBINE – *Comparative Binding Energy Analysis* (Análise Comparativa da Energia de Ligação)

CoMFA – *Comparative Molecular Field Analysis* (Análise Comparativa do Campo Molecular)

DFT – *Density Funcional Theory* (Teoria do Funcional de Densidade)

DNA – *Desoxiribonucleic acid* (Ácido desoxiribonucleico)

dNTP – 2'-desoxinucleotídeo-5'-trifosfato

DABO – *Dihydro-alkyloxy-benzyl-oxopyrimidine* (Diidroalcoxi-benzil-oxopirimidina)

DAPY – *Diarylaminopyrimidine* (Diarilaminopirimidina)

DATA – *Diaryltriazine* (Diariltriazina)

DM – Dinâmica Molecular

EA – *Evolutionary Algorithm* (Algoritmo Evolucionário)

GA – *Genetic Algorithm* (Algoritmo Genético)

GFA – *Genetic Function Approximation* (Aproximação da Função Genética)

HAART – *Highly Active Anti-Retroviral Therapy* (Terapia Anti-Retroviral Altamente Ativa)

HEPT – *1-[2-Hydroxyethoxy)methyl]-6-(phenylthio)-thymine* (Hidróxi-etóxi-metil feniltiotimina)

HF – *Hartree-Fock*

IC₅₀ – *Inhibition Concentration at 50%* (Concentração de Inibição a 50%)

ITU – *Imidoylthiourea* (Imidoil-tiouréia)

LOF – *Lack-of-Fit* (Falta de Ajuste)

LOO_{cv} – *Leave-One-Out Crossvalidation* (Validação Cruzada Deixe-Um-Fora)

LSE – *Least Squares Error* (Erro de Mínimos Quadrados)

MSA – *Molecular Shape Analysis* (Análise da Forma Molecular)

NMR¹H – *Nuclear Magnetic Resonance* (Ressonância Magnética Nuclear de Hidrogênio¹)

NMR¹³C – *Nuclear Magnetic Resonance* (Ressonância Magnética Nuclear de Carbono¹³)

NMR¹⁵N – *Nuclear Magnetic Resonance* (Ressonância Magnética Nuclear de Nitrogênio¹⁵)

NNBS – *Non-Nucleoside Binding Site* (Sítio de Ligação Não-Nucleosídeo)

NNRTI – *Non-Nucleoside Reverse Transcriptase Inhibitor* (Inibidor Não-Nucleosídeo da Transcriptase Reversa)

NRTI – *Nucleoside/Nucleotide Reverse Transcriptase Inhibitor* (Inibidor Nucleosídeo/Nucleotídeo da Transcriptase Reversa)

PC – *Principal Component* (Componente Principal)

PDB – *Protein Data Bank* (Banco de Dados de Proteína)

PETT – *Phenylethyl)-N-(2-thiazolyl)thiourea* (Fenil-etil-tiazolil-tiouréia)

PLS – *Partial Least Squares* (Mínimos Quadrados Parciais)

PME – *Particle-Mesh Ewald*

PM3 – *Parametric Model 3*

PRESS – *Predictive Sum of Squares* (Soma dos Quadrados Preditiva)

QSAR – *Quantitative Structure-Activity Relationship* (Relação Quantitativa Estrutura-Atividade)

RESP – *Restrained Molecular Electrostatic Potential* (Potencial Eletrostático Molecular Restrito)

RMS – *Root Mean Square* (Raiz Média Quadrada)

RNA – *Ribonucleic acid* (Ácido Ribonucléico)

RT – *Reverse Transcriptase* (Transcriptase Reversa)

SAR – *Structure-Activity Relationship* (Relação Estrutura-Atividade)

SEE – *Standard Error of Estimate* (Erro Padrão de Estimativa)

TIBO – *Tetrahydro-imidazo-benzodiazepinone* (Tetraidro-imidazo-benzodiazepinona)

SUMÁRIO

RESUMO

ABSTRACT

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

SUMÁRIO

1. INTRODUÇÃO	26
1.1. ASPECTOS BIOLÓGICOS E ESTRUTURAIS.....	26
1.1.1. HIV e AIDS	26
1.1.2. Estatística de Infectados e Óbitos pelo HIV	27
1.1.3. Formas de Transmissão do HIV	29
1.1.4. Alvos de Infecção do HIV em Humanos	29
1.1.5. Organização Estrutural do HIV	30
1.1.6. Ciclo Replicativo do HIV	30
1.1.7. Coquetel de Fármacos (HAART)	32
1.1.8. Transcriptase Reversa do HIV-1	33
1.1.8.1. Inibidores da Transcriptase Reversa do HIV-1	35
A. Inibidores Nucleosídeos.....	36
B. Inibidores Não-Nucleosídeos	38
B.1. Diversidade Estrutural dos Inibidores Não-Nucleosídeos	38
B.2. Inibidores Não-Nucleosídeos Utilizados na Prática Clínica....	40
B.3. Sítio de Ligação dos NNRTIs	41
B.4. Conformação Bioativa dos NNRTIs	42
B.5. Principais Tipos de Interações dos NNRTIs	43
1.1.9. Estudos de Cristalografia de Raios-X e Modelagem Molecular com a Transcriptase Reversa do HIV-1.....	44
1.1.10. Resistência do HIV-1 aos NNRTIs	45
1.1.11. Principais Mutações Descritas para os NNRTIs	46
1.1.12. Classe dos DABOs	47
1.1.13. Rota Sintética dos DABOs	49
1.1.14. Citotoxicidade dos DABOs	52
1.2. ASPECTOS TEÓRICOS	52

1.2.1. SAR e QSAR Clássico: Análises de Hansch & Fujita.....	52
1.2.2. Importância dos Estudos de QSAR.....	53
1.2.3. Evolução dos Estudos de QSAR: do QSAR-1D ao QSAR-6D	54
1.2.4. QSAR-3D Independente e Dependente da Estrutura do Receptor.....	56
A. QSAR-3D Independente da Estrutura do Receptor	57
A.1. QSAR-3D Independente do Receptor Mais Aplicado no Planejamento de Fármacos: CoMFA.....	58
A.1.1. Requerimentos Estruturais e Conformação Bioativa.....	58
A.1.2. Escolha dos Átomos para o Alinhamento	59
A.1.3. Cargas Atômicas Parciais	60
A.1.4. Caixa para o Alinhamento	60
A.1.5. Construção dos Campos Moleculares	61
A.1.6. Técnicas Estatísticas Aplicadas no CoMFA.....	62
A.1.6.1. Mínimos Quadrados Parciais	64
A.1.6.2. Validação Cruzada <i>Leave-One-Out</i>	65
A.1.7. Vantagens e Limitações do CoMFA.....	66
B. QSAR-3D Dependente da Estrutura do Receptor.....	67
B.1. Algoritmos Genéticos em Estudos de QSAR-3D Dependente do Receptor.....	69
B.1.1. Algoritmos Genéticos	69
B.1.2. As Bases da Seleção Natural Aplicadas aos GAs	69
B.1.3. Vantagens e Aplicações dos Algoritmos Genéticos na Modelagem Molecular	71
B.1.4. Aproximação da Função Genética com Mínimos Quadrados Parciais	71
2. OBJETIVOS.....	73
3. METODOLOGIA	74
3.1. BANCO DE DADOS ESTRUTURAL E BIOLÓGICO.....	74
3.1.1. Perfil Estrutural dos DABOs	76
3.1.2. Definição dos Conjuntos de Treinamento e de Teste.....	77
3.2. QSAR-3D INDEPENDENTE E DEPENDENTE DO RECEPTOR	79
3.2.1. Construção e Otimização das Estruturas 3D dos Ligantes	79
3.2.2. Parte A – QSAR-3D Independente do Receptor (CoMFA).....	80

3.2.2.1. Definição da Hipótese Farmacofórica.....	80
3.2.2.2. Assinalamento das Cargas Atômicas Parciais	81
3.2.2.3. Definição dos Alinhamentos	81
3.2.2.4. Definição dos Níveis de <i>Cutoff</i>	83
3.2.2.5. Definição da Caixa e Obtenção das Variáveis Independentes .	83
3.2.2.6. Obtenção dos Modelos de CoMFA.....	83
3.2.2.7. Validação Interna dos Modelos de CoMFA	84
3.2.2.8. Validação Externa dos Modelos de CoMFA	86
3.2.2.9. Seleção do Melhor Modelo de CoMFA.....	86
3.2.2.10. Seleção dos <i>Outliers</i>	87
3.2.3. Parte B. QSAR-3D Dependente do Receptor.....	87
3.2.3.1. Construção e Otimização dos Complexos Ligante-Enzima	87
3.2.3.2. Simulação por Dinâmica Molecular dos Complexos Ligante-Enzima.....	88
3.2.3.3. Definição do Raio de Corte e Obtenção das Energias de Interação Ligante-Enzima (Variáveis Independentes)	90
3.2.3.4. Definição dos Bancos de Dados de Variáveis Independentes .	91
3.2.3.5. Obtenção das Equações de QSAR-3D Dependentes do Receptor	92
3.2.3.6. Validação Interna dos Modelos de QSAR-3D Dependentes do Receptor	93
3.2.3.7. Validação Externa dos Modelos de QSAR-3D Dependentes do Receptor	94
3.2.3.8. Seleção dos Melhores Modelos de QSAR-3D Dependente do Receptor	94
3.2.3.9. Seleção dos <i>Outliers</i>	95
3.2.3.10. Análise da Matriz de Correlação Cruzada dos Resíduos	95
3.2.3.11. Análise da Matriz de Correlação Cruzada dos Descritores	96
4. RESULTADOS E DISCUSSÃO	97
4.1. PARTE A – QSAR-3D INDEPENDENTE DO RECEPTOR (CoMFA).....	98
4.1.1. Análise da Sobreposição dos DABOs em Função dos Alinhamentos.....	98
4.1.2. Seleção do Melhor Alinhamento e Carga Atômica Parcial	99
4.1.3. Seleção do Tipo de Átomo de Prova.....	103

4.1.4. Seleção do Valor de Corte de Energia	104
4.1.5. Análise dos Resíduos dos Compostos dos Conjuntos de Treinamento e de Teste	106
4.1.6. Análise dos Mapas de Contorno do Melhor Modelo de CoMFA.....	109
4.1.7. Análise dos Derivados Mais e Menos Potentes	114
4.1.8. Análise dos Compostos <i>Outliers</i>	115
4.2. PARTE B – QSAR-3D DEPENDENTE DO RECEPTOR.....	118
4.2.1. Análise da Simulação por Dinâmica Molecular dos Complexos Ligantes-Enzima.....	118
4.2.2. Análise do Recorte dos Complexos Ligantes-Enzima	119
4.2.3. Avaliação dos Bancos de Dados	121
4.2.3.1. Análise do Banco de Dados I	122
A) Análise dos Índices Estatísticos do BD-I	122
B) Análise da Melhor Equação do BD-I (Eq.E)	123
C) Análise dos Valores Residuais da Melhor Equação do BD-I (Eq.E)	130
D) Análise dos Compostos <i>Outliers</i> da Eq.E (BD-I)	133
E) Análise do MKC-442 Segundo a Eq.E (BD-I)	134
F) Análise da Matriz de Correlação Cruzada entre os Descritores da Eq.E (BD-I)	135
4.2.3.2. Análise do Banco de Dados II	136
A) Análise dos Índices Estatísticos do BD-II	136
B) Análise da Melhor Equação do BD-II (Eq.J)	137
C) Análise dos Valores Residuais da Eq.J (BD-II)	141
D) Análise dos Compostos <i>Outliers</i> da Eq.J (BD-II)	145
E) Análise do MKC-442 Segundo a Eq.J (BD-II)	146
F) Análise da Matriz de Correlação Cruzada entre os Descritores da Eq.J (BD-II)	147
4.2.3.3. Análise do Banco de Dados III	148
A) Análise dos Índices Estatísticos do BD-III	148
B) Análise da Melhor Equação do BD-III (Eq.L)	149
C) Análise dos Valores Residuais da Eq.L (BD-III)	154
D) Análise dos Compostos <i>Outliers</i> da Eq.L (BD-III)	157
E) Análise do MKC-442 Segundo a Eq.L (BD-III)	158

F) Análise da Matriz de Correlação Cruzada dos Descritores da	
Eq.L (BD-III)	158
4.2.3.4. Análise do Banco de Dados IV	159
A) Análise dos Índices Estatísticos do BD-IV	161
B) Análise da Melhor Equação do BD-IV (Eq.Q)	163
C) Análise dos Valores Residuais da Eq.Q (BD-IV)	168
D) Análise dos Compostos <i>Outliers</i> da Eq.Q (BD-IV)	172
E) Análise do MKC-442 Segundo a Eq.Q (BD-IV)	173
F) Análise da Matriz de Correlação Cruzada dos Descritores da	
Eq.Q (BD-IV)	174
4.2.4. Análise Global das Melhores Equações dos Bancos de Dados	
I a IV	175
4.2.5. Predição das Potências Biológicas (pIC ₅₀) de DABOs não	
Incluídos nos Conjuntos de Treinamento e de Teste	176
4.3. Planejamento e Proposição de Novos DABOs	184
4.3.1. Predição da Potência Biológica (pIC ₅₀) do NH-DABO Proposto	187
5. CONCLUSÕES	190
6. PERSPECTIVAS	192
7. REFERÊNCIAS BIBLIOGRÁFICAS	193
8. ANEXOS	228
Tabela A. Matriz de correlação cruzada entre os descritores das melhores	
equações dos BD-I a -IV	228
Artigo de QSAR-3D Independente do Receptor (CoMFA)	
Capítulo de Livro de Química Medicinal sobre QSAR	

1. INTRODUÇÃO

1.1. ASPECTOS BIOLÓGICOS E ESTRUTURAIS

1.1.1. HIV e AIDS

O vírus da imunodeficiência humana (*Human Immunodeficiency Virus*, HIV), um lentivírus pertencente à família *Retroviridae* (NCBI, 2008; Reid, 2002; Garg *et al.*, 1999), é um parasita do sistema imunológico humano, causando uma doença infecciosa conhecida como síndrome da imunodeficiência adquirida (*Acquired Immuno Deficiency Syndrome*, AIDS) (CDC, 2008; Raffanti & Haas, 2006; De Clercq, 2002; Gallo *et al.*, 1984; Barre-Sinoussi *et al.*, 1983).

Dois tipos de HIV foram identificados, o HIV-1 e o HIV-2 (Hightower & Kallas, 2003). O HIV-1 é predominante, mais disseminado pelo mundo e apresenta maior taxa de mutação. O HIV-2 apresenta índices de patogenicidade e de transmissibilidade inferiores aos do HIV-1, como consequência, os indivíduos infectados apenas com o HIV-2 possuem, em geral, sobrevida mais longa (Bird *et al.*, 2003). O HIV-2 está mais restrito à África Ocidental, a países como Senegal, Guiné, Gâmbia e Cabo Verde (Hightower & Kallas, 2003).

A AIDS é uma doença que afeta o sistema imunológico, onde o organismo torna-se incapacitado de se defender. A doença é caracterizada por perda de peso acentuada, astenia (fraqueza) e pela suscetibilidade a infecções (AIDS Education Global Information System, www.aegis.com/).

1.1.2. Estatística de Infectados e Óbitos pelo HIV

Em todo o mundo, a infecção pelo HIV já matou mais de 26 milhões de pessoas desde quando foi oficialmente reconhecida, em 1981, mostrando-se uma das epidemias mais destrutivas da história (AIDS Epidemic Update, 2007).

O número total de pessoas que vivem com o vírus alcançou um nível elevado, estimado em cerca de 33 milhões de pessoas até 2007 (Figura 1A) (AIDS Epidemic Update, 2007).

Considerando apenas o ano de 2007, e a despeito do acesso ao tratamento anti-retroviral em muitas regiões do mundo, a epidemia da AIDS atingiu 2,5 milhões de pessoas, das quais 420 mil são crianças (Figura 1B) (AIDS Epidemic Update, 2007).

Estes números mostram o desafio que os pesquisadores têm pela frente no sentido de combater o HIV. Além dos esforços sociais, políticos e econômicos, as Indústrias Farmacêuticas e Universidades de grandes centros mundiais estão colocando em prática pesquisas para identificar novos alvos e novas classes de fármacos para combater o HIV (AIDS Education & Training Centers National Resource Center, 2008; AIDS Education Global Information System, 2008; Albany Medical College Division of HIV Medicine, 2008; CDC National Prevention Information Network, 2008; HIV Clinical Resource, New York State Department of Health AIDS Institute, 2008; Johns Hopkins AIDS Service; Debnath, 2005).

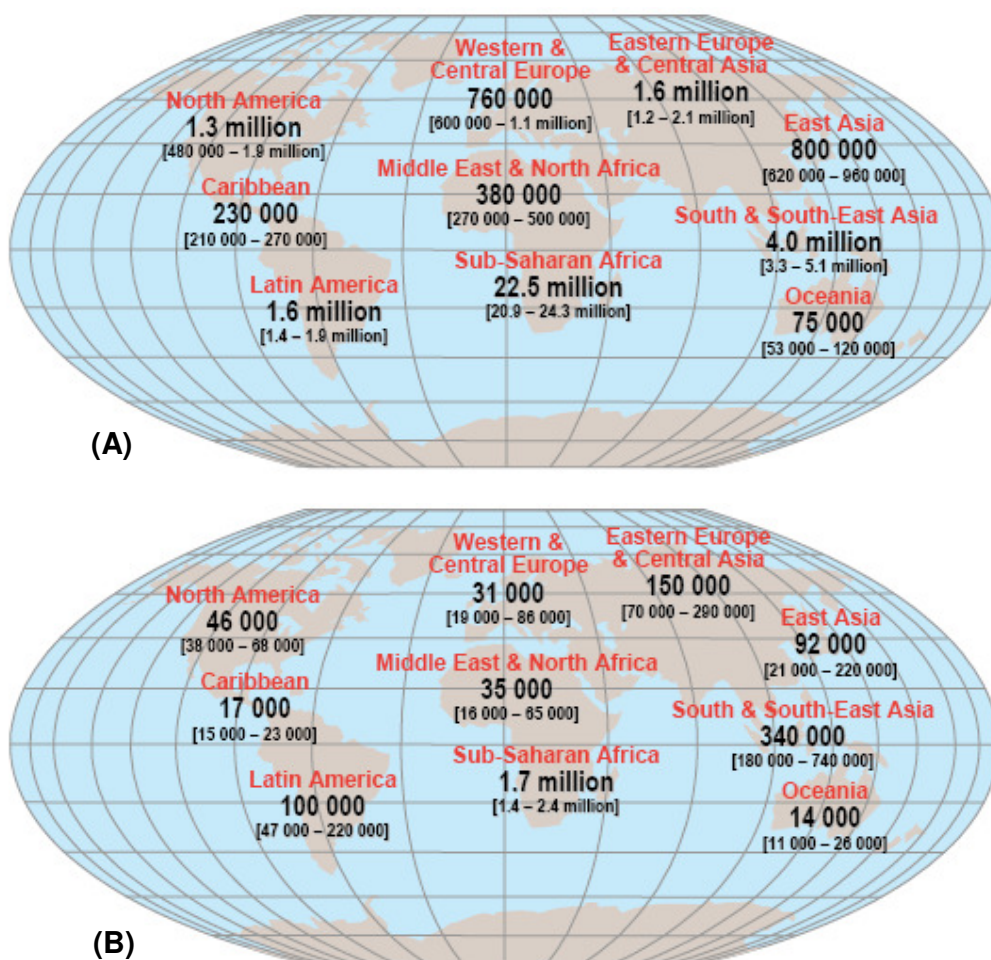


Figura 1. Mapa-múndi dividido por regiões indicando o número estimado de pessoas (crianças e adultos) com o HIV. (A) Total de infectados até 2007. (B) Total de infectados no ano de 2007 (Reproduzido de AIDS Epidemic Update, 2007).

1.1.3. Formas de Transmissão do HIV

O HIV pode ser encontrado em diversos órgãos, tecidos e fluidos humanos, como sangue, líquido cefalorraquidiano, leite materno, líquido amniótico, sêmen e secreções vaginais (Harms & Feldmeier, 2002). Dessa forma, as principais formas de transmissão do vírus são as que envolvem contato entre algum desses fluidos; e incluem a sexual, a sangüínea (pessoas que recebem sangue ou hemoderivados e usuários de drogas injetáveis) e a vertical (quando a criança é infectada pelo vírus HIV durante a gestação, o parto ou o aleitamento). Além dessas formas mais freqüentes, também pode ocorrer a transmissão ocupacional, ocasionada por acidente de trabalho, em que profissionais da área da saúde sofrem ferimentos com instrumentos perfuro-cortantes contaminados com sangue de pacientes infectados com o HIV (CDC, 2008; WHO/UNAIDS, 2008; Harms & Feldmeier, 2002).

1.1.4. Alvos de Infecção do HIV em Humanos

Os alvos de infecção do HIV nos humanos são células que contêm uma glicoproteína de membrana denominada CD4, como macrófagos e células dendríticas e, principalmente, linfócitos T, que são células do sistema imunológico diretamente envolvidas na defesa do organismo contra microorganismos invasores (Raffanti & Haas, 2006; De Clercq, 2004; Souza & Almeida, 2003; Thomas, 2003). Ao infectar os linfócitos T_{CD4}, o HIV conduz à falta de coordenação do sistema imunológico e à sua progressiva inoperância, acabando por estabelecer um quadro de imunodeficiência e permitindo o desenvolvimento de infecções oportunistas (De Clercq, 2002; Hayden, 2001).

1.1.5. Organização Estrutural do HIV

Como ilustrado na Figura 2, o HIV possui uma estrutura relativamente simples, composta por duas moléculas de ácido ribonucleico (*Ribonucleic Acid*, RNA) envoltas por um capsídeo e um tegumento e uma camada lipoproteica externa (CDC, 2008a; Balzarini, 2004; Souza & Almeida, 2003). Na superfície externa do vírus encontra-se a glicoproteína gp120, ligada à glicoproteína gp41, que atravessa o envoltório lipoproteico, cuja função é o reconhecimento molecular a receptores CD4 nas células hospedeiras (Debnath, 2005).

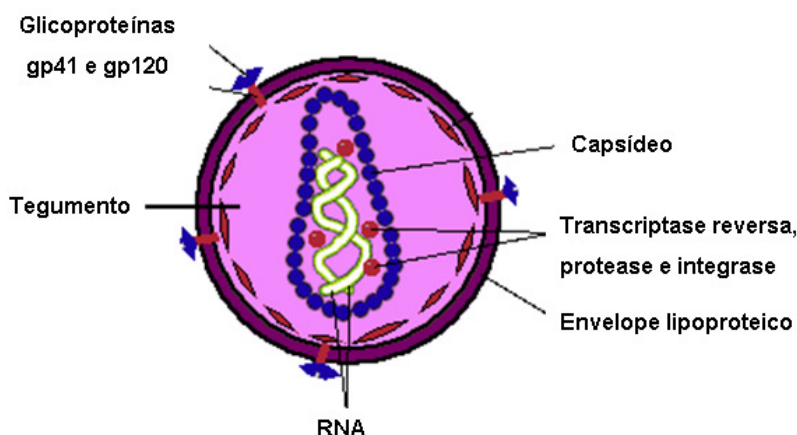


Figura 2. Organização estrutural esquemática do vírus HIV (Adaptada de www.mclld.co.uk/hiv).

1.1.6. Ciclo Replicativo do HIV

O estudo do ciclo replicativo do HIV possibilitou a identificação de alguns alvos macromoleculares suscetíveis à intervenção terapêutica. Os fármacos anti-retrovirais em uso clínico têm como alvos principais as enzimas virais transcriptase reversa, integrase e protease, mas há fármacos em desenvolvimento para outros alvos (Raffanti & Haas, 2006; Debnath, 2005).

O ciclo replicativo do HIV pode ser dividido esquematicamente em cinco etapas (Figura 3): fusão, transcrição reversa, integração, transcrição e tradução/formação de proteínas virais.

Na Etapa 1, o vírus funde-se à membrana celular após reconhecimento molecular via interações entre as glicoproteínas do envelope viral (gp41 e gp120) e os receptores CD4 e co-receptores (quimiocinas, como CXCR4 e CXCR5) da célula hospedeira (Raffanti & Haas, 2006; Debnath, 2005; Souza & Almeida, 2003). Um fármaco inibidor de fusão, cujo nome comercial é Fuzeon[®] (enfuvirtida), foi aprovado em 2003 pelo órgão norte-americano de controle sobre produtos alimentícios e farmacêuticos (*Food and Drug Administration*, FDA) como fármaco anti-retroviral (De Clercq, 2004; FDA, 2003).

Na Etapa 2, após a liberação do conteúdo viral no citosol, ocorre a transcrição reversa da fita simples de RNA viral para DNA, catalisada pela enzima transcriptase reversa (*Reverse Transcriptase*, RT). Essa enzima constitui o alvo de duas classes de fármacos anti-retrovirais, os inibidores nucleosídeo/nucleotídeo (*Nucleoside/Nucleotide Reverse Transcriptase Inhibitors*, NRTIs) e os inibidores não-nucleosídeos (*Non-Nucleoside Reverse Transcriptase Inhibitors*, NNRTIs) (Souza & Almeida, 2003; De Clercq, 2002; De Clercq, 1998).

Na Etapa 3, após a transcrição reversa, ocorre a incorporação da fita dupla de DNA pró-viral ao DNA humano no núcleo celular, catalisada pela enzima integrase (Balzarini, 2004). O fármaco raltegravir (Isentress[®]) foi aprovado em 2007 pelo FDA como inibidor desta enzima (FDA, 2007; Raffanti & Haas, 2006).

Na Etapa 4, uma vez incorporado ao cromossomo humano, o DNA pró-viral é transcrito em RNA viral, que pode sofrer tradução em poliproteínas virais (Etapa 5) ou ser incorporado a vírions imaturos (Etapa 6), que sofrem maturação e brotamento

pela membrana celular. A maturação exige a clivagem de poliproteínas virais pela enzima protease viral. Os vírions maduros podem então infectar outras células suscetíveis (Raffanti & Haas, 2006; Debnath, 2005).

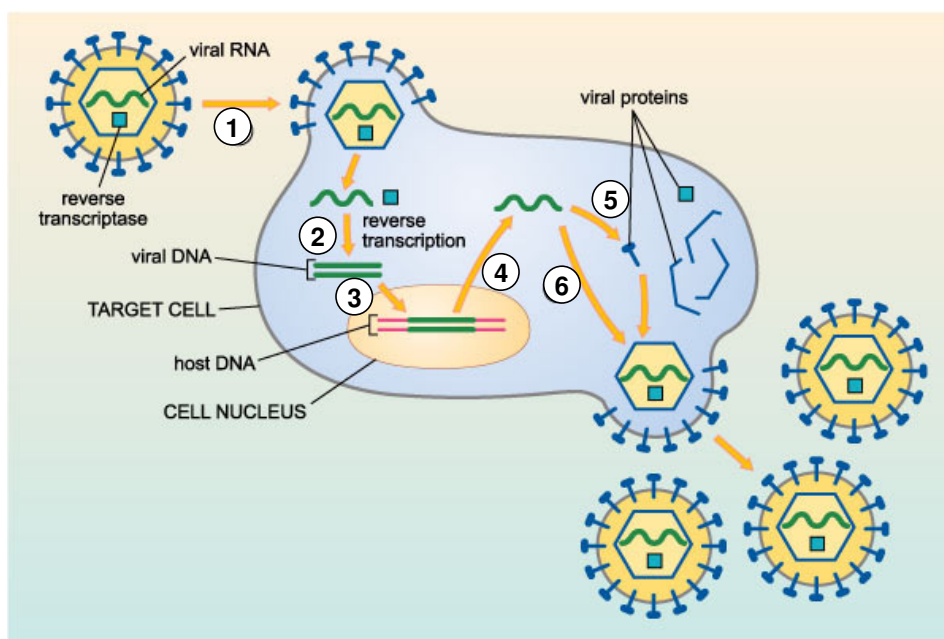


Figura 3. Esquema do ciclo replicativo do HIV (Adaptada de <http://www.tthivclinic.com/lifecycle4.htm>).

1.1.7. Coquetel de Fármacos (HAART)

O tratamento que utiliza uma combinação de antivirais potentes tendo como alvos a RT e a protease do HIV é abreviado HAART (*Highly Active Anti-Retroviral Therapy*), popularmente conhecido como “coquetel” (Janssen *et al.*, 2005).

O uso de combinações de fármacos é uma estratégia mais eficiente do que os regimes monoterápicos (Fattorusso *et al.*, 2005; Das *et al.*, 2004; De Clercq, 2004) por retardar o surgimento de variantes resistentes (UNAIDS, 2008; Fattorusso *et al.*, 2005; Kaufmann & Cooper, 2000; Vella & Palmisano, 2000). As combinações mais comuns incluem dois ou três inibidores da RT e um ou dois inibidores de protease (Janssen *et al.*, 2005).

1.1.8. Transcriptase Reversa do HIV-1

A RT é uma enzima essencial para o HIV, pois ela catalisa a transcrição reversa da fita simples de RNA viral para DNA. Estruturalmente, a RT é heterodimérica e consiste de duas subunidades proteicas de 66 kDa (p66) e 51 kDa (p51) (Figura 4) (Castro *et al.*, 2006; Balzarini, 2004; Huang *et al.*, 1998; Rodgers *et al.*, 1995; Jacobo-Molina *et al.*, 1993).

A subunidade p66 contém dois domínios, o domínio polimerase N-terminal (440 resíduos) e o domínio ribonuclease H (RNaseH) C-terminal (120 resíduos). O sítio catalítico polimerase é composto por uma tríade de resíduos de ácido aspártico, **Asp110**, **Asp185** e **Asp186** (Yadav & Singh, 2005).

A subunidade p51, derivada da p66 após clivagem por uma protease, contém outro domínio polimerase, que não é funcional (Balzarini, 2004; Rodgers *et al.*, 1995; Jacobo-Molina *et al.*, 1993). Acredita-se que a subunidade p51 forneça suporte estrutural à subunidade p66, uma vez que a forma dimérica da RT é um requerimento para sua atividade enzimática (Camarasa *et al.*, 2006; Tachedjian *et al.*, 2005).

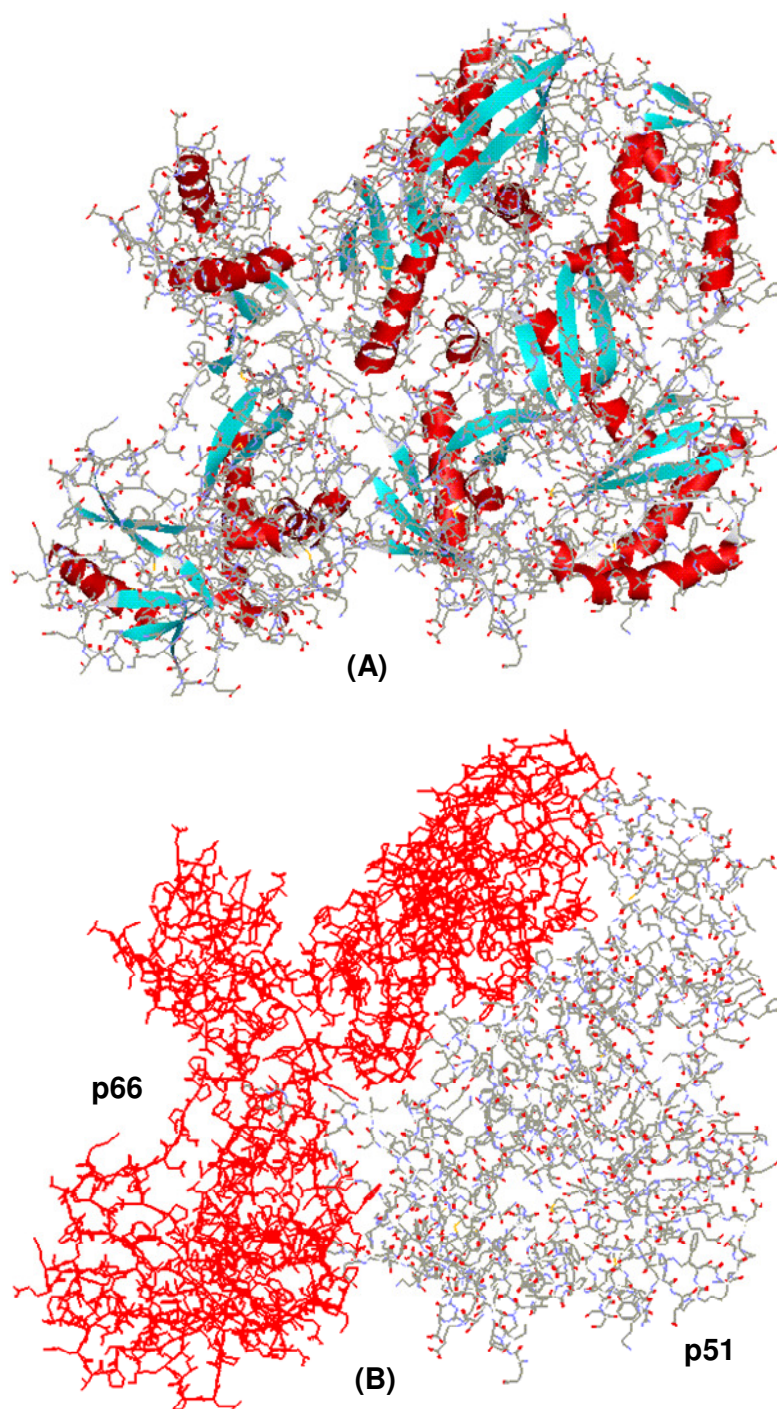


Figura 4. Estrutura 3D da transcriptase reversa do HIV-1. (A) Representação da estrutura secundária em modelo de fita (hélices-alfa em vermelho e folhas-beta em azul claro). (B) Representação das subunidades p66 (em vermelho) e p51 (colorido por elemento).

1.1.8.1. Inibidores da Transcriptase Reversa do HIV-1

Os inibidores da RT do HIV-1 podem ser classificados em dois grupos principais, de acordo com o sítio de ligação preferencial na enzima. Os inibidores da classe dos nucleosídeos, NRTIs, são assim denominados por possuírem analogia estrutural com os substratos naturais do sítio polimerase, dNTP (2'-desóxinucleotídeo-5'-trifosfato), competindo com estes nucleotídeos na ligação ao sítio catalítico (Castro *et al.*, 2006; Baba *et al.*, 1989; Miyasaka *et al.*, 1989; Baba *et al.*, 1991; Tanaka *et al.*, 1991).

Os inibidores da classe dos não-nucleosídeos, NNRTIs, apresentam estruturas químicas diversificadas (De Corte, 2005; Ren & Stammers, 2005), possuem cinética de inibição do tipo não competitiva em relação ao substrato natural e ligam-se a um sítio alostérico na RT (Yadav & Singh, 2005). A Figura 5 ilustra a estrutura 3D da RT do HIV-1, destacando a tríade catalítica do domínio polimerase (sítio de ligação dos NRTIs) e alguns resíduos do sítio alostérico (sítio de ligação dos NNRTIs).

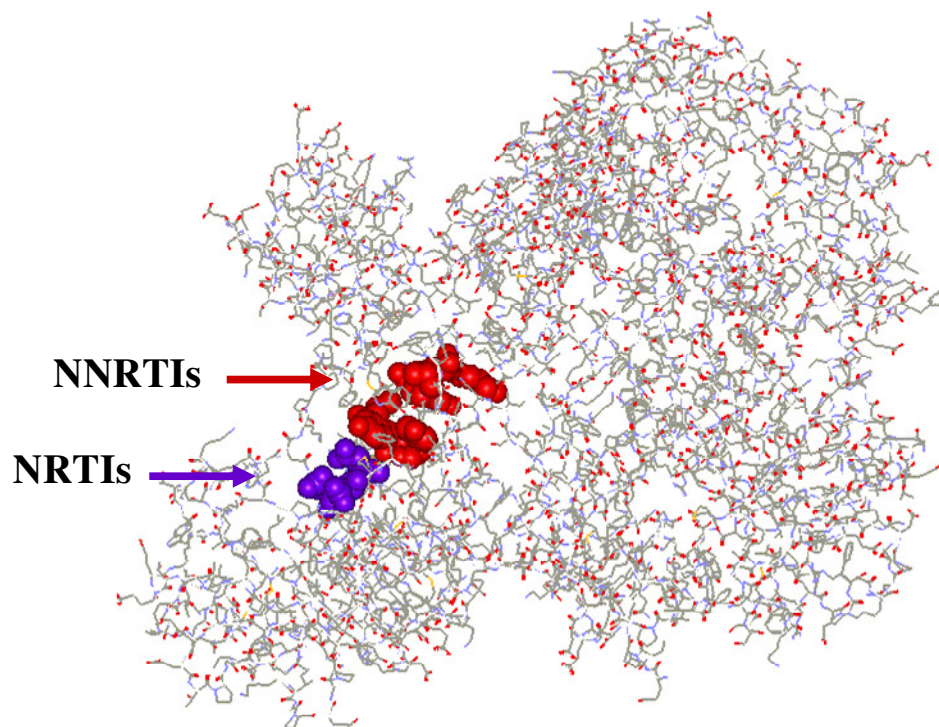


Figura 5. Estrutura da RT do HIV-1 destacando os sítios dos NRTIs e NNRTIs. Em azul os resíduos catalíticos (**Asp110**, **Asp185** e **Asp186**) e em vermelho alguns resíduos do sítio alostérico (**Lys101**, **Lys103**, **Val106**, **Tyr181** e **Tyr188**), todos em modelo CPK.

A. Inibidores Nucleosídeos

A terapia da infecção pelo HIV-1 começou com os NRTIs, com derivados 2',3'-didesoxinucleosídeos, particularmente com o 3'-azido-2',3'-didesoxitimidina (zidovudina ou AZT), aprovado pelo FDA em 1986 e comercializado a partir de 1987 (Souza & Almeida, 2003). Em seguida, foram lançados comercialmente outros fármacos da classe dos inibidores nucleosídeos: didanosina, zalcitabina, estavudina, lamivudina, abacavir, tenofovir e, mais recentemente, a emtricitabina (De Clercq, 2004).

Os inibidores nucleosídeos ligam-se ao sítio ativo da RT (domínio polimerase), mimetizando os substratos naturais da enzima (dNTPs), mas falta-lhes

um grupo 3'-OH necessário para a elongação do DNA, o que causa o término prematuro da fita de DNA em crescimento (Rizzo *et al.*, 2002).

A principal desvantagem dos NRTIs é que eles precisam ser convertidos intracelularmente aos seus metabólitos ativos 5'-trifosfatos, *i.e.*, são pró-fármacos (Castro *et al.*, 2006; Balzarini, 2004; Baba *et al.*, 1991).

A Tabela 1 lista os fármacos em uso clínico da classe dos NRTIs pelo nome genérico, nome comercial, nome da indústria farmacêutica envolvida e ano de lançamento.

Tabela 1. Nomes genéricos e comerciais, indústrias farmacêuticas envolvidas e anos de lançamento dos oito fármacos da classe dos NRTIs em uso clínico (Adaptada de Flexner, 2007).

Nome Genérico	Nome Comercial[®]	Indústria Farmacêutica	Ano de Lançamento
Zidovudina	Retrovir	GlaxoSmithKline	1987
Didanosina	Videx	Bristol-Myers Squibb	1991
Zalcitabina	Hivid	Roche	1992
Estavudina	Zerit	Bristol-Myers Squibb	1995
Lamivudina	Epivir	GlaxoSmithKline, Shire Pharmac.	1998
Abacavir	Ziagen	GlaxoSmithKline	1999
Tenofovir	Viread	Gilead	2001
Emtricitabina	Emtriva	Gilead	2003

B. Inibidores Não-Nucleosídeos

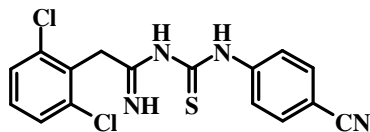
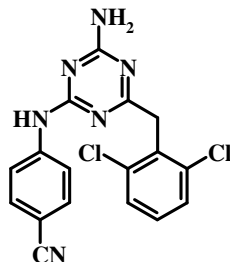
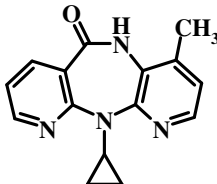
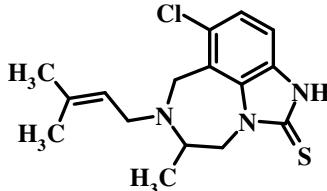
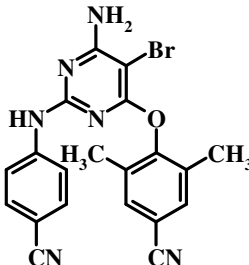
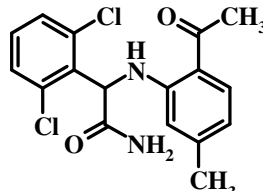
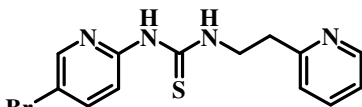
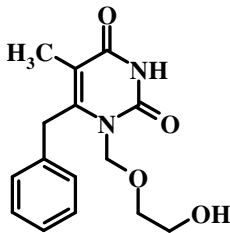
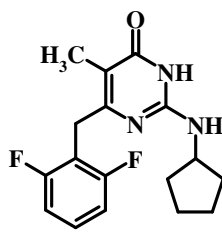
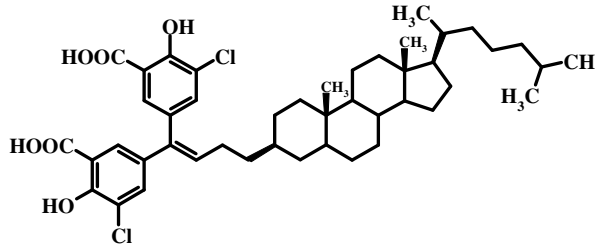
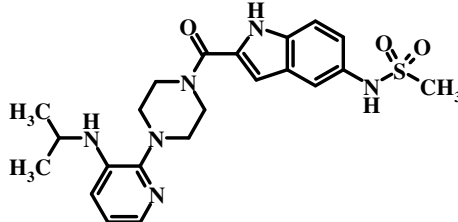
A era dos NNRTIs começou a aproximadamente 18 anos com a descoberta da classe do hidróxi-etóxi-metil-feniltiotimina (HEPT) (Baba *et al.*, 1991; Tanaka *et al.*, 1991; Baba *et al.*, 1989; Miyasaka *et al.*, 1989) e do tetraidroimidazo-benzodiazepinona (TIBO) (De Corte, 2005; Schafer *et al.*, 1993; Debyser *et al.*, 1991; Pauwels *et al.*, 1990).

Os NNRTIs possuem especificidade para a RT do HIV-1, sendo inativos para o HIV-2 e outros retrovírus (Castro *et al.*, 2006; Smith *et al.*, 1995; Baba *et al.*, 1991), embora as RTs do HIV-1 e do HIV-2 tenham aproximadamente 60% de identidade na seqüência primária de aminoácidos (Smerdon *et al.*, 1994; Baba *et al.*, 1991).

B.1. Diversidade Estrutural dos Inibidores Não-Nucleosídeos

Mais de 30 classes de compostos estruturalmente diferentes têm sido identificadas como NNRTIs, tais como a classe dos DAPY (diarilpirimidina), DABO (diidroalcóxi-benzil-oxopirimidina), α -APA (α -anilino-fenilacetamida), entre outras (Tabela 2) (Castro *et al.*, 2006; Deng *et al.*, 2006; Zhang *et al.*, 2006; De Corte, 2005; Balzarini, 2004; De Clercq, 1998). Sabe-se que o fator hidrofóbico desses compostos é muito importante para a atividade dos mesmos, já que o sítio de ligação dos NNRTIs, abreviado NNBS (*Non-Nucleoside Binding Site*), é essencialmente constituído por resíduos de aminoácidos com caráter hidrofóbico (Venezia *et al.*, 2006; De Corte, 2005; Balzarini, 2004).

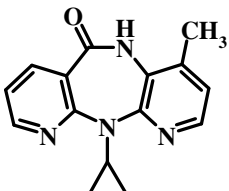
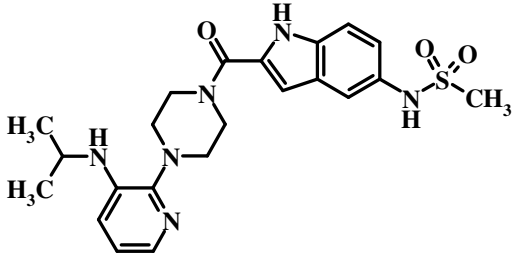
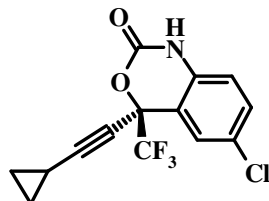
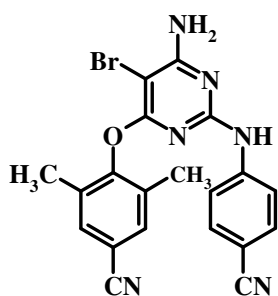
Tabela 2. Principais classes de NNRTIs, abreviaturas e exemplos estruturais.

Classe (Abreviatura) & Exemplo		
<p>Imidoil-tiouréia (ITU)</p> 	<p>Diariltiazina (DATA)</p> 	<p>Dipirido-diazepinona</p> 
<p>Tetraidroimidazo-benzodiazepin -ona e -tiona (TIBO)</p> 	<p>Diarilpirimidina (DAPY)</p> 	<p>α-anilino-fenil-acetamida (α-APA)</p> 
<p>Fenil-etil-tiazolil-tiouréia (PETT)</p> 	<p>Hidróxi-etóxi-metil-fenil-tiotimina (HEPT)</p> 	<p>Diidroalcóxi-benzil-oxopirimidina (DABO)</p> 
<p>Alquenil-diarilmetano (ADAM)</p> 	<p>Bis-heteroaril-piperazina (BHAP)</p> 	

B.2. Inibidores Não-Nucleosídeos Utilizados na Prática Clínica

Atualmente, há no mercado quatro fármacos da classe dos NNRTIs: nevirapina (dipiridodiazepinona), delavirdina (BHAP), efavirenz e etravirina (DAPY) (Haubrich *et al.*, 2008; Souza & Almeida, 2003). A etravirina é o fármaco mais recente aprovado pelo FDA, em janeiro de 2008 (FDA, 2008; Haubrich *et al.*, 2008). Em testes clínicos, a etravirina mostrou eficácia contra variantes do HIV-1 resistentes ao efavirenz e à nevirapina (Scott *et al.*, 2008). A Tabela 3 lista os NNRTIs comerciais pelo nome genérico, estrutura química, nome comercial, indústria farmacêutica envolvida e ano de lançamento.

Tabela 3. Nomes genéricos e comerciais, estruturas químicas, indústrias farmacêuticas envolvidas e anos de lançamento dos quatro fármacos da classe dos NNRTIs em uso clínico (Adaptada de Flexner, 2007).

Nome Genérico (Comercial [®])	Estrutura Química	Indústria Farmacêutica	Ano de Lançamento
Nevirapina (Viramune)		Boehringer Ingelheim	1996
Delavirdina (Rescriptor)		Pharmacia & Upjohn	1997
Efavirenz (Sustiva)		Bristol-Myers Squibb	1998
Etravirina (Intence)		Tibotec	2008

B.3. Sítio de Ligação dos NNRTIs

Os NNRTIs ligam-se a um sítio alostérico composto principalmente por resíduos hidrofóbicos, como **Pro95**, **Leu100**, **Val106**, **Val179**, **Tyr181**, **Tyr188**, **Phe227**, **Trp229**, **Leu234**, **His235**, **Pro236** e **Tyr318**. Este sítio está localizado na

subunidade p66 da RT do HIV-1, a aproximadamente 10 Å de distância do sítio catalítico polimerase (Yadav & Singh, 2005; Balzarini, 2004; Hopkins *et al.*, 2004; De Clercq, 1998; Ren *et al.*, 1998; Ren *et al.*, 1995; Kohlstaedt *et al.*, 1992).

A cavidade do sítio alostérico característico dos NNRTIs (NNBS) não existe fisicamente na enzima na ausência do ligante, sendo formada após a ligação do inibidor (Das *et al.*, 2004; Hopkins *et al.*, 1996; Ren *et al.*, 1995). Quando um NNRTI liga-se ao NNBS, ocorre uma desorganização da tríade catalítica (resíduos Asp do sítio ativo) devido à reorientação das cadeias laterais dos resíduos **Tyr181** e **Tyr188** em direção aos resíduos catalíticos, resultando em uma diminuição da atividade catalítica da enzima (Ren *et al.*, 1998; Miyasaka *et al.*, 1996; Esnouf *et al.*, 1995; Ren *et al.*, 1995; Rodgers *et al.*, 1995). Após a ligação do NNRTI, os resíduos **Pro225** e **Pro226** fecham o sítio (Ren *et al.*, 1998).

B.4. Conformação Bioativa dos NNRTIs

Apesar de apresentarem estruturas químicas diversas, os NNRTIs ligam-se à RT de maneira semelhante (Yadav & Singh, 2005; Balzarini, 2004). A análise de complexos cristalográficos desses inibidores revela que compostos como a delavirdina e derivados das classes do TIBO, HEPT e DABO assumem uma conformação do tipo “borboleta” (alguns autores referem-se também como “V”), onde as “asas” seriam representadas por anéis aromáticos e estariam separadas por aproximadamente 7,5 Å, ligadas por um “corpo” representado por grupos espaçadores diversos (Barreca *et al.*, 2005; Balzarini, 2004; Das *et al.*, 2004; De Clercq, 2004; Parreira *et al.*, 2001; Ren *et al.*, 1995; Schafer *et al.*, 1993).

Estudos conformacionais de modelagem molecular de algumas classes de NNRTIs (Yadav & Singh, 2005; Parreira *et al.*, 2001; Mota-Neto *et al.*, 1992a; Mota-Neto *et al.*, 1992b) corroboram estes dados experimentais.

B.5. Principais Tipos de Interações dos NNRTIs

Estudos experimentais com enzimas resistentes (Geitmann & Danielson, 2007; Ren *et al.*, 2006; Petersen *et al.*, 2005; Hopkins *et al.*, 2004; Ragno *et al.*, 2004; Rao *et al.*, 2004; Mai *et al.*, 2001; Hopkins *et al.*, 1996) e de modelagem molecular (Almerico *et al.*, 2008; Ragno *et al.*, 2005; Shen *et al.*, 2003; Udier-Blagovic *et al.*, 2003; Mai *et al.*, 2001; Smith *et al.*, 1995) permitem inferir que a estabilização da ligação dos NNRTIs ao sítio alostérico da RT do HIV-1 é alcançada por quatro tipos principais de interações:

- I. Interações do tipo *pi-stacking* com os anéis aromáticos das cadeias laterais dos resíduos **Tyr181**, **Tyr188**, **Phe227**, **Trp229** e **Tyr318**;
- II. Interações eletrostáticas com os grupos ionizados das cadeias laterais dos resíduos **Lys101**, **Lys103** e **Glu138**;
- III. Interações de van der Waals com os grupos apolares das cadeias laterais dos resíduos **Leu100**, **Val106**, **Val179**, **Tyr181**, **Gly190**, **Trp229**, **Leu234** e **Tyr318**;
- IV. Interações por ligação hidrogênio com os grupos amida da cadeia principal dos resíduos **Lys101** e **Lys103**.

1.1.9. Estudos de Cristalografia de Raios-X e Modelagem Molecular com a Transcriptase Reversa do HIV-1

Inúmeros trabalhos de modelagem molecular utilizam estruturas de cristalografia de raios-X da RT nativa e de mutantes em estudos de *docking* (Carlsson *et al.*, 2008; Zhang *et al.*, 2006; Barreca *et al.*, 2005; Ragno *et al.*, 2005; Rizzo *et al.*, 2004; Medina-Franco *et al.*, 2004; Xu *et al.*, 2004; Zhou & Madura, 2004; Ranise *et al.*, 2003; Silvestri *et al.*, 2002; Rizzo *et al.*, 2000; Hopkins *et al.*, 1999; Smith *et al.*, 1995), dinâmica molecular (Carlsson *et al.*, 2008; Su *et al.*, 2007; Rungrotmongkol *et al.*, 2006; Rodríguez-Barrios *et al.*, 2005; Zhou *et al.*, 2005; Udier-Blagovic *et al.*, 2003; Smith *et al.*, 2003; Shen *et al.*, 2003; Rizzo *et al.*, 2000; Eriksson *et al.*, 1999) e QSAR (Bak & Polanski, 2006; Zhang *et al.*, 2006; Daszykowski *et al.*, 2005; Zhou & Madura, 2004; Weekes & Fogel, 2003; Gáudio & Montanari, 2002; Hannongbua *et al.*, 2001; Luco & Ferretti, 1997).

A primeira estrutura cristalográfica da RT corresponde ao complexo com um potente NNRTI, a nevirapina (código no Banco de Dados de Proteínas, PDB 1HVT) (Kohlstaedt *et al.*, 1992). Desde então, inúmeras estruturas cristalográficas da RT livre ou em complexo com diferentes classes de inibidores tem sido relatadas (Castro *et al.*, 2006; Balzarini, 2004; De Clercq, 1998).

A análise das estruturas cristalográficas de RT do HIV-1 tem revelado importantes características da função e da estrutura da enzima, incluindo detalhes do sítio de ligação dos inibidores. Essas estruturas incluem a RT livre (Hsiou *et al.*, 1996; Esnouf *et al.*, 1995; Rodgers *et al.*, 1995), em complexo com substratos (Sarafianos *et al.*, 2001; Sarafianos *et al.*, 1999; Huang *et al.*, 1998; Jacobo-Molina *et al.*, 1993) e com diferentes inibidores não-nucleosídeos, como nevirapina (Ren *et*

al., 1995; Kohlstaedt *et al.*, 1992), delavirdina (Esnouf *et al.*, 1997), efavirenz (Lindberg *et al.*, 2002; Ren *et al.*, 2000), HBV097 (Hsiou *et al.*, 1998), PETT (Hogberg *et al.*, 1999; Ren *et al.*, 2000), MKC-442 (Hopkins *et al.*, 1996), e, mais recentemente, a etravirina (Das *et al.*, 2004).

1.1.10. Resistência do HIV-1 aos NNRTIs

O HIV tem uma alta taxa de replicação, que pode alcançar até 10^{10} partículas virais por dia num indivíduo não tratado (Perelson *et al.*, 1996). Adicionalmente, no processo de transcrição reversa do vírus, observa-se um número alto de mutações genéticas, estimadas na ordem de 10^4 a 10^5 vezes por dia (Coffin, 1995). A elevada taxa de mutação aliada à conseqüente variabilidade genética tem como principal conseqüência a seleção e a predominância de cepas resistentes aos fármacos anti-HIV empregados no combate à AIDS.

O surgimento de variantes virais resistentes a fármacos em pacientes infectados com o HIV é a principal causa da falha no tratamento (Das *et al.*, 2004; De Clercq, 2004; De Clercq, 2002). O desenvolvimento de resistência é um fator importante quando se considera a administração de um fármaco por um período prolongado (De Clercq, 2004). Nesse caso, o inibidor passa a atuar como um elemento de pressão seletiva para a sobrevivência do vírus, e o processo de mutação torna-se acelerado em função do uso inadequado dos fármacos anti-HIV (De Clercq, 2004).

As mutações alteram a afinidade dos inibidores pela RT, resultando, geralmente, em diminuição das interações de van der Waals entre inibidor e enzima (Himmel *et al.*, 2005; Rodríguez-Barrios *et al.*, 2005; Zhou *et al.*, 2005; Das *et al.*, 2004; Hopkins *et al.*, 2004; Hsiou *et al.*, 1998; Smerdon *et al.*, 1994).

1.1.11. Principais Mutações Descritas para os NNRTIs

O desenvolvimento de resistência é um problema significativo da classe dos NNRTIs (Rodríguez-Barrios *et al.*, 2005; Das *et al.*, 2004; De Clercq, 2004). Mutações simples tais como **Leu100Ile**, **Lys101Glu**, **Lys103Asn**, **Val106Ala**, **Val108Ile**, **Tyr181Cys**, **Tyr188Leu**, **Gly190Ala**, **Pro225His** e **Phe227Leu** (Jorgensen *et al.*, 2006; Ren *et al.*, 2006; Himmel *et al.*, 2005; Rodríguez-Barrios *et al.*, 2005; Das *et al.*, 2004; El-Brollosy *et al.*, 2002; Rizzo *et al.*, 2000; Corbett *et al.*, 1999; Hopkins *et al.*, 1999) e mutações duplas como **Lys103Asn/Tyr181Cys** (Gagnon *et al.*, 2007; El-Brollosy *et al.*, 2002), **Lys103Asn/Val108Ile**, **Lys103Asn/Pro225His** e **Lys103Asn/Leu100Ile** (Corbett *et al.*, 1999) têm sido descritas na literatura em função do uso prolongado de NNRTIs.

A Figura 6 mostra o sítio de ligação do fármaco etravirina, um NNRTI, na RT mutante **Lys103Asn** a partir da estrutura de cristalografia de raios-X deste composto (código PDB 1SV5) (Das *et al.*, 2004).

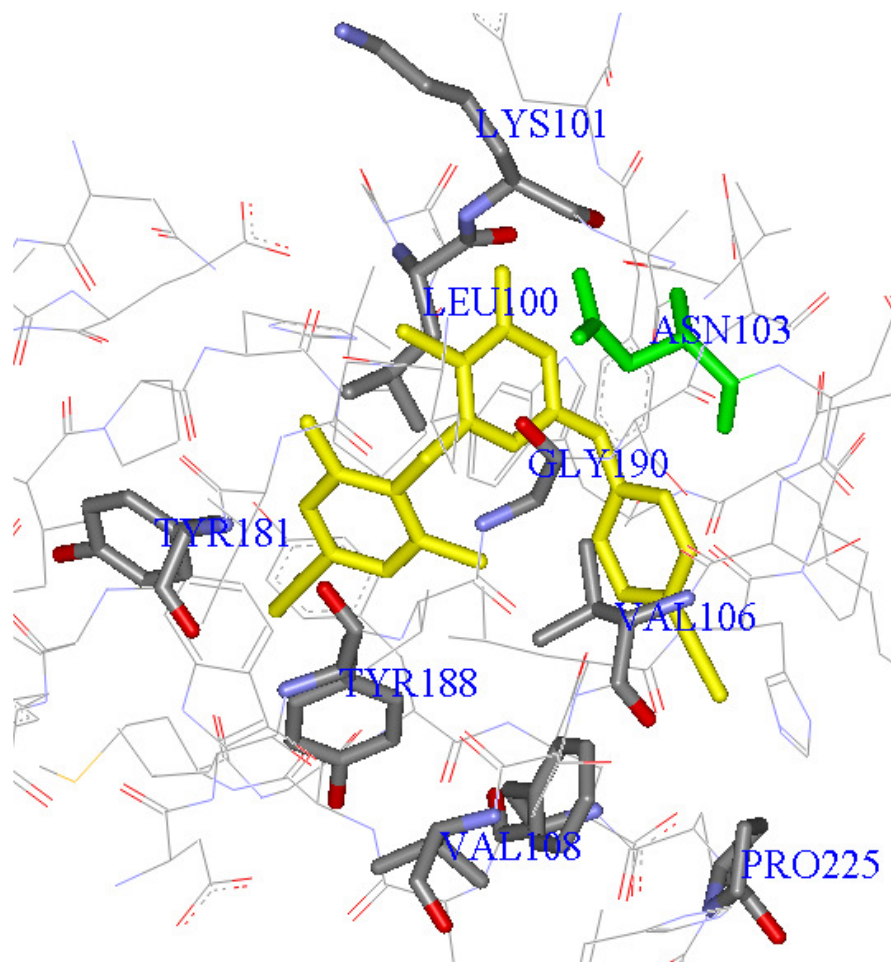


Figura 6. Etravirina (em bastão amarelo) no NNBS (Código PDB 1SV5). Os resíduos nomeados (em bastão) são os mais suscetíveis a mutações. Todos estão coloridos por elemento, exceto **Asn103** (em verde) que é uma mutação de **Lys103**.

1.1.12. Classe dos DABOs

Os DABOs são uma classe potente de NNRTIs, estruturalmente semelhantes à classe dos HEPTs, como ilustrado na Figura 7 (Mugnaini *et al.*, 2006; Togninelli *et al.*, 2006; Mai *et al.*, 2005; Manetti *et al.*, 2005; He *et al.*, 2004; Ragno *et al.*, 2004; Tranter, 2001; Mai *et al.*, 1995; Artico *et al.*, 1993).

As principais modificações estruturais na classe dos HEPTs que deram origem aos DABOS foram a retirada do substituinte da posição **N1**; a substituição da carbonila da posição **C2** por um átomo de oxigênio ligado a grupos alquila; a substituição facultativa do grupo metila na posição **C5** por um átomo de hidrogênio; a manutenção do grupo benzila ligado à posição **C6**, porém com substituintes no anel aromático.

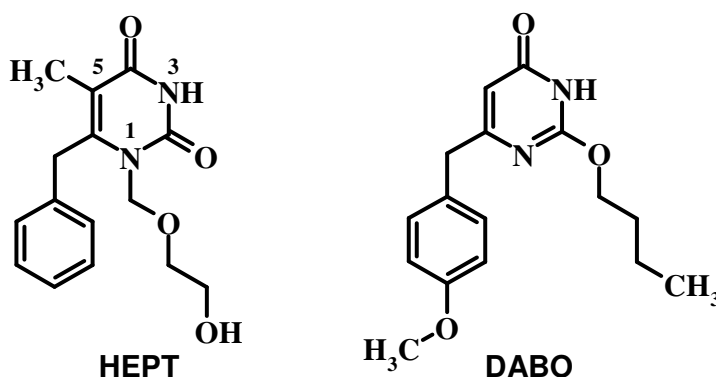


Figura 7. Estruturas de um HEPT e de um DABO.

Os DABOs foram inicialmente sintetizados como inibidores da enzima diidrofolato redutase de fungos e bactérias, mas a semelhança estrutural com a classe dos HEPTs fez com que fossem testados contra o HIV, mostrando atividade contra o vírus e seletividade para o HIV-1 (Botta *et al.*, 1992).

Estruturalmente, são caracterizados da seguinte forma (Figura 8):

- (A) Base pirimidínica uracila (substituinte em **C5**=H) ou timina (substituinte em **C5**=Me, sendo os compostos com timina, em geral, mais potentes;
- (B) Na posição **C2**, os isósteros -O-alquila, -S-alquila ou -NH-alquila, dos quais o grupo -NH-alquila é o que confere, em geral, maior potência aos derivados e o grupo -O-alquila é o que confere, em geral, menor potência;

(C) Na posição **C6**, um grupo arila, em geral, um grupo benzila não-substituído, mono ou di-substituído, sendo os compostos com o grupo 2,6-difluoro-benzila os mais potentes da série.

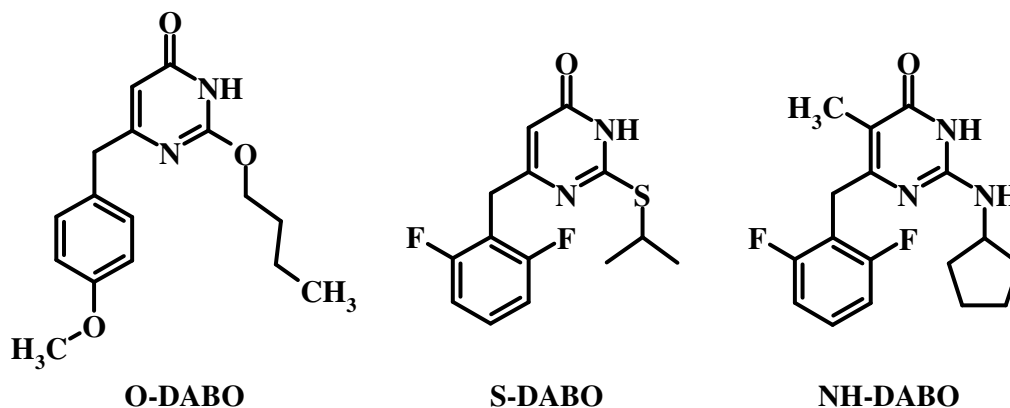


Figura 8. Estruturas de O-, S- e NH-DABOs.

Diversos trabalhos de QSAR-3D (Roy *et al.*, 2008; Guo *et al.*, 2006; Polanski *et al.*, 2004; Gayen *et al.*, 2004; Toropov *et al.*, 2003; Tang & Li, 2002; Hannongbua *et al.*, 2001; Kireev *et al.*, 1997; Luco & Ferretti, 1997; Tronchet *et al.*, 1997), QSAR-4D (Bak & Polanski, 2006) e Redes Neurais (Arakawa *et al.*, 2006; Chiu & So, 2004; Weekes & Fogel, 2003; Douali *et al.*, 2003; Jalali-Heravi & Parastar, 2000) são descritos na literatura para a classe dos HEPTs, uma das classes de NNRTIs mais estudadas. Porém, não há trabalhos de QSAR envolvendo apenas a classe dos DABOs.

1.1.13. Rota Sintética dos DABOs

A síntese dos DABOs está esquematizada na Figura 9, conforme descrita por Massa *et al.*, 1995 e Artico *et al.*, 1993. Inicialmente ocorre a reação de acilação da 2,2-dimetil-1,3-dioxano-4,6-diona (**A**) com cloreto fenilacético (**B**) na presença de piridina anidra em diclorometano e subsequente metanólise sob refluxo dos

intermediários **(C)** formados gerando os respectivos β -cetoésteres **(D)** que são tratados com quantidade equimolar de iodeto de metila na presença de metóxido de sódio, fornecendo os respectivos derivados β -cetoésteres α -metilados **(E)**. A condensação destes com o sulfato de O-metil-isouréia na presença de hidróxido de cálcio, fornece as respectivas 3,4-diidro-2-metóxi-4-oxopirimidinas **(F)**, intermediários-chaves na preparação dos DABOs. A reação de **(F)** com o correspondente alcóxido ou cicloalcóxido de potássio sob refluxo fornece as correspondentes 3,4-diidro-2-alcóxi-6-benzil-4-oxopirimidinas **(G)**.

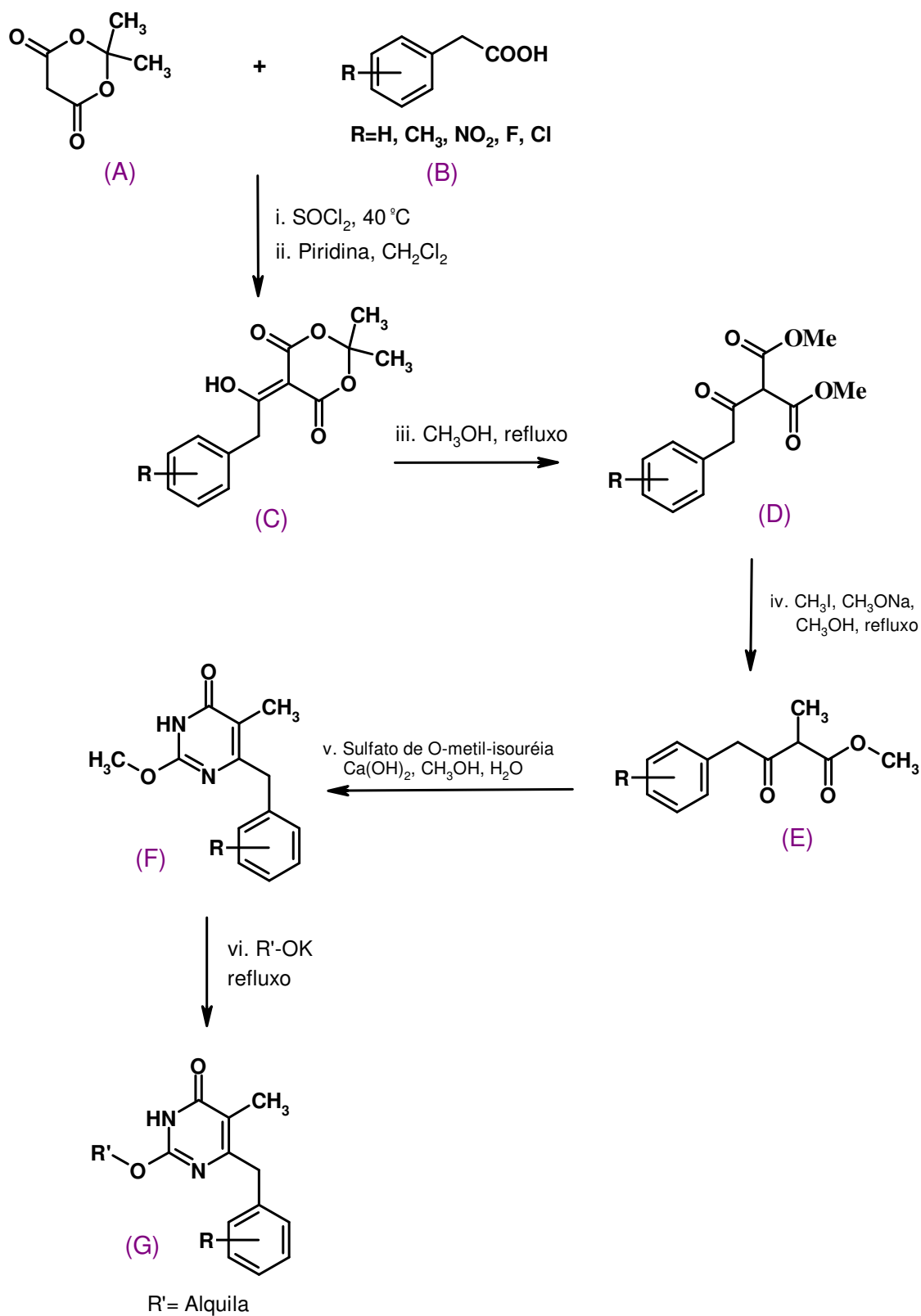


Figura 9. Esquema da rota sintética dos derivados DABOs (Massa *et al.*, 1995 e Artico *et al.*, 1993).

1.1.14. Citotoxicidade dos DABOs

Ensaio de citotoxicidade de derivados da classe dos DABOs com doses de até 300 μ M mostraram um índice de seletividade (razão entre a citotoxicidade, CC_{50} , e a atividade, IC_{50}) alto para a maioria dos derivados (Ragno *et al.*, 2004; Mai *et al.*, 1999; Massa *et al.*, 1995). Estes ensaios foram realizados em células MT-4 com o reagente MTT (brometo de 3-(4,5-dimetil-tiazol-2-il)-2,5-difenil-tetrazolium), que sofre redução em células vivas formando cristais, e pode ser quantificado espectrofotometricamente, de onde se estima o número de células vivas em uma amostra (Freimoser *et al.*, 1999; Mai *et al.*, 1999; Pauwels *et al.*, 1988).

1.2. ASPECTOS TEÓRICOS

1.2.1. SAR e QSAR Clássico: Análises de Hansch & Fujita

Desde 1868 os cientistas observam que alterações na estrutura química de compostos bioativos resultam em mudanças no perfil de atividade desses compostos, nascendo então a metodologia de “Relação Estrutura-Atividade” (*Structure-Activity Relationship*, SAR) (Livingstone, 1995; Kubinyi, 1993; Hansch, 1969; Hansch & Fujita, 1964; Crum-Brown & Fraser, 1868). Estudos de SAR ajudam a decidir que características de uma molécula refletem em aumento na sua atividade e ajudam a modificar compostos com propriedades ruins (Leach, 2001, pp. 695).

Entretanto, a metodologia conhecida como “Relação Quantitativa Estrutura-Atividade” (*Quantitative Structure-Activity Relationship*, QSAR) somente foi racionalizada na década de 1960 com os trabalhos pioneiros de Corwin Hansch e

Toshio Fujita (Hansch, 1969; Hansch & Fujita, 1964), que demonstraram que a resposta biológica pode ser correlacionada quantitativamente com parâmetros físico-químicos, expressos de acordo com a Equação 1.

$$\text{Log (1/C)} = k_1 \text{LogP} - k_2 (\text{LogP})^2 + k_3 \sigma + k_4 \quad \text{Equação 1}$$

onde **C** corresponde à concentração molar do composto necessária para produzir uma resposta biológica definida; **LogP** é o logaritmo do coeficiente de partição de um composto entre octanol e água (corresponde à contribuição hidrofóbica), **σ** é o índice de Hammett dos substituintes, **k₁-k₃** são os respectivos coeficientes dos termos (determinados por análise de regressão) e **k₄** é uma constante.

Em estudos de QSAR é prática comum dividir o banco de dados original em dois grupos, um denominado conjunto de treinamento e o outro conjunto de teste. O conjunto de treinamento contém os compostos com os quais se constrói o modelo de QSAR, enquanto o conjunto de teste contém os compostos usados no processo de validação externa, determinando a preditividade externa do modelo. Assim, o conjunto de teste corresponde a uma percentagem de cerca de 10 a 30% de compostos que não foi incluída no processo de geração do modelo (Albuquerque *et al.*, 2007; Romeiro *et al.*, 2005; Brito, 2004; Albuquerque *et al.*, 1998; Hopfinger *et al.*, 1997).

1.2.2. Importância dos Estudos de QSAR

Os estudos de QSAR são métodos importantes na história do desenvolvimento de fármacos, sendo intensamente relatados na química medicinal

farmacêutica como fundamentais na otimização de protótipos de fármacos (van de Waterbeemd & Rose, 2003; Albuquerque *et al.*, 2007).

Uma pesquisa no banco de dados “Web of Science” procurando no título do artigo a palavra QSAR no “Science Citation Index Expanded” desde o ano de 1945, resultou em 3.080 trabalhos publicados. A mesma pesquisa considerando apenas os últimos 5 anos resultou em 1.337 trabalhos (ISI Web of Knowledge, <http://www.isiwebknowledge.com/>).

Alguns fármacos mundialmente comercializados (*e.g.* captopril, pravastatina, ciprofoxacin) foram planejados racionalmente e muitas classes de fármacos (*e.g.* anti-hipertensivos, antibióticos β -lactâmicos, bloqueadores de canais de cálcio) foram otimizadas com a ajuda de estudos de QSAR (van de Waterbeemd & Rose, 2003; Thomas, 2003; Leach, 2001; Kubinyi, 1997).

1.2.3. Evolução dos Estudos de QSAR: do QSAR-1D ao QSAR-6D

A metodologia de Hansch, denominada QSAR clássico ou QSAR-1D/2D (1D, uma dimensão e 2D, duas dimensões), utiliza como descritores parâmetros experimentais (LogP, pKa, etc) ou calculados da estrutura bidimensional (Albuquerque *et al.*, 2007; Vedani *et al.*, 2006). A partir da década de 1980, com o uso mais constante da estrutura molecular tridimensional (3D) proveniente das técnicas de modelagem molecular em desenvolvimento, os descritores 3D começaram a substituir os descritores 1D/2D (Debnath, 2001; Cho *et al.*, 1996; Cramer *et al.*, 1988; Hopfinger, 1980), dando origem aos métodos de QSAR-3D, onde são construídos modelos quantitativos, estatisticamente validados, relacionando matematicamente a variação da potência biológica de uma série de

compostos com a variação da estrutura tridimensional (Brito *et al.*, 2008; Brito, 2004; Sant'Anna, 2002; Rodrigues, 1999).

Posteriormente, surgiram outras metodologias de QSAR adicionando novas dimensões, como a 4D, que considera múltiplas conformações do ligante. A metodologia de QSAR-4D é uma abordagem de QSAR-3D que utiliza uma amostragem conformacional obtida por meio de simulação por Dinâmica Molecular (DM), reduzindo a dificuldade em identificar a conformação bioativa, isto é, aquela que efetivamente interage com o receptor (Albuquerque *et al.*, 2007; Albuquerque *et al.*, 1998; Hopfinger *et al.*, 1997).

Com a intenção de aperfeiçoar o poder preditivo das equações de QSAR, dimensões adicionais foram incorporadas aos métodos, como a 5D, onde a teoria do encaixe induzido ligante-receptor é considerada, ou seja, a adaptação mútua entre cada ligante individual e o respectivo sítio de ligação no receptor, como descrito por Vedani e colaboradores (Vedani *et al.*, 2005a; Vedani & Dobler, 2002; Vedani *et al.*, 2000). Recentemente, o grupo de Vedani desenvolveu o método de QSAR-6D, que considera o efeito do solvente no complexo ligante-receptor (Lill *et al.*, 2005; Vedani *et al.*, 2005b). A Tabela 4 lista as principais características dos métodos de QSAR de acordo com a dimensão incorporada e a possibilidade de incluir a estrutura do receptor, como resumido por Vedani e colaboradores (Vedani *et al.*, 2006).

Tabela 4. Dimensões dos métodos de QSAR, principais características e possibilidade de representar a estrutura do receptor (Adaptada de Vedani *et al.*, 2006).

Dimensão do QSAR	Característica	Receptor
1D	Atividade correlacionada com propriedades eletrônicas e estéricas	Não
2D	Atividade correlacionada com propriedades estruturais	Não
3D	Atividade correlacionada com estruturas 3D	Possível
4D	Ligantes representados como conjunto de conformações, orientações, diferentes estados de protonação	Sim
5D	Representação de diferentes modelos de encaixe induzido	Sim
6D	Representação de diferentes cenários de solvatação	Sim

1.2.4. QSAR-3D Independente e Dependente da Estrutura do Receptor

Os métodos de QSAR-3D podem ser classificados como independentes e dependentes do receptor (Albuquerque *et al.*, 2007; Albuquerque *et al.*, 2006; Sippl, 2002; Hopfinger *et al.*, 1997; Cho *et al.*, 1996; Loew *et al.*, 1993).

Os descritores do QSAR-3D são provenientes do cálculo de propriedades relacionadas apenas à estrutura 3D dos ligantes quando o método é independente do receptor (*Receptor-Independent*, RI, ou *Ligand-Based*), ou à estrutura 3D do complexo ligante-receptor quando o método é dependente do receptor (*Receptor-Dependent*, RD, ou *Structure-Based*) (Albuquerque *et al.*, 2006; Sippl, 2002; Hopfinger *et al.*, 1997; Cho *et al.*, 1996; Loew *et al.*, 1993).

A. QSAR-3D Independente da Estrutura do Receptor

Na abordagem RI, faz-se a análise comparativa das estruturas dos compostos de potência variada em relação a um determinado receptor, utilizando-se o conceito de complementaridade molecular para o desenvolvimento de um modelo topográfico hipotético do sítio receptor, denominado modelo farmacofórico (Marshall & Cramer, 1988).

Duas metodologias de QSAR-3D RI conhecidas são a “Análise da Forma Molecular” (*Molecular Shape Analysis*, MSA), desenvolvida por Hopfinger em 1980, que incorpora a análise conformacional dos ligantes e a obtenção de descritores globais 3D da forma molecular nos estudos de QSAR (Hopfinger, 1980) e a “Análise Comparativa do Campo Molecular” (*Comparative Molecular Field Analysis*, CoMFA), desenvolvida por Cramer e colaboradores em 1988, onde a variação da resposta biológica é correlacionada com as variações de energia estérica e eletrostática quando um átomo-prova interage com os ligantes (Cramer *et al.*, 1988).

Comparados à análise de Hansch (QSAR clássico), o MSA e o CoMFA são mais adequados para descrever as interações ligante-receptor, pois consideram as propriedades dos ligantes na sua suposta conformação bioativa (Kubinyi, 1997). Diferente dos métodos de QSAR clássicos, esses métodos podem ser aplicados a conjuntos de compostos estruturalmente diversos (Marshall & Cramer, 1988). Além disso, em alguns modelos, as propriedades moleculares que aumentam ou diminuem a potência biológica podem ser visualizadas na forma de gráficos, como no CoMFA (Cramer *et al.*, 1988) e no QSAR-4D (Cunha, 2006).

A.1. QSAR-3D Independente do Receptor Mais Aplicado no Planejamento de Fármacos: CoMFA

O CoMFA é atualmente a metodologia de QSAR-3D mais utilizada e relatada na literatura, sendo usado em centenas de estudos de planejamento de fármacos (Brito *et al.*, 2008; Hannongbua *et al.*, 2001; Rodrigues *et al.*, 2000; Guner, 2002; Chen, 1998; Golbraikh *et al.*, 2000). Uma pesquisa no banco de dados “Web of Science” procurando no título do artigo a palavra CoMFA no “Science Citation Index Expanded” a partir do ano de 1988 resultou em 393 trabalhos. A mesma pesquisa considerando apenas os últimos 5 anos resultou em 205 referências (ISI Web of Knowledge, <http://www.isiwebofknowledge.com/>).

Aplicações do CoMFA feitas com sucesso nos últimos anos provaram o valor dessa metodologia, especialmente em casos onde métodos clássicos de QSAR falham (Kubinyi, 1997). Em contraste com as análises de Hansch e Free & Wilson, o método CoMFA é mais adequado para descrever interações ligante-receptor porque ele considera as propriedades dos ligantes nas suas supostas conformações bioativas (Kubinyi, 1997; Free & Wilson, 1964; Hansch & Fujita, 1964).

A.1.1. Requerimentos Estruturais e Conformação Bioativa

O primeiro passo na construção de um modelo de CoMFA é a seleção dos compostos, que devem ser estruturalmente relacionados e atuarem pelo mesmo mecanismo de ação. A escolha das prováveis conformações responsáveis pela bioatividade desses compostos é feita com base nos modelos farmacofóricos, nos resultados de *docking* ou em dados cristalográficos relatados na literatura (Brito, 2004; Golbraikh *et al.*, 2000; Kubinyi, 1997).

Em contraste com métodos clássicos de QSAR, os compostos selecionados devem ter um farmacóforo comum, mas não necessariamente o mesmo esqueleto molecular (Todeschini & Consonni, 2000; Marshall & Cramer, 1988). Quando há moléculas com muitos graus de liberdade, podendo gerar várias conformações, e não estão disponíveis informações sobre as suas conformações bioativas, alguns autores têm utilizado a conformação de menor energia, isto é, o confórmero mais estável, para a construção do modelo de CoMFA (Brito, 2004; Rodrigues, 1999; Burt *et al.*, 1991).

A.1.2. Escolha dos Átomos para o Alinhamento

Após a seleção dos compostos e da provável conformação bioativa, é feita a escolha dos átomos para o alinhamento dos compostos, uma vez que o CoMFA é um método dependente do alinhamento. A escolha dos átomos para o alinhamento é difícil, pois pode influenciar o modelo final, especialmente quando o conjunto de moléculas escolhido possui diversidade estrutural (Ivanciuc, 2000). A escolha pode ser feita de acordo com regras de orientação que seguem dos farmacóforos maximizando a sobreposição estérica dos compostos. Em muitas situações um alinhamento perfeito é difícil, por exemplo, para moléculas que não são de uma série congênere ou aquelas com grande número de ligações com rotação livre.

No primeiro estudo de CoMFA, publicado pelos idealizadores do método, uma série de esteróides foi sobreposta pelo alinhamento de seus esqueletos rígidos (Cramer *et al.*, 1988). A escolha desse grupo de compostos foi um exemplo bem sucedido, porque esteróides não têm flexibilidade conformacional e o alinhamento tornou-se fácil, sem ambigüidades (Kubinyi, 1997).

A.1.3. Cargas Atômicas Parciais

Escolhida a hipótese farmacofórica e os átomos para o alinhamento dos compostos, segue-se o cálculo das cargas atômicas parciais. A escolha do método de geração das cargas pode ser feita levando-se em conta os resultados da literatura para séries de compostos estruturalmente semelhantes aos estudados (Brito, 2004). Em alguns estudos, porém, faz-se o cálculo das cargas por dois ou mais métodos para compará-los e selecionar o que leva a melhores resultados (Brito, 2004; Rodrigues, 2000; Cunha, 2006).

Alguns autores enfatizam que as interações eletrostáticas são as principais forças responsáveis pela interação ligante-receptor e que os campos eletrostáticos resultantes de diferentes métodos de cálculo de cargas podem influenciar o modelo CoMFA significativamente (Ivanciuc, 2000; Kroemer, 1996).

A.1.4. Caixa para o Alinhamento

Escolhidos os átomos para o alinhamento e as cargas atômicas parciais, uma caixa é gerada ao redor das moléculas alinhadas de tal maneira que a caixa cubra inteiramente os compostos e haja um espaço entre eles e a caixa, como ilustrado na Figura 10. A caixa é dividida em pequenos cubos (grades), e a distância padrão entre eles é de 2 Å.

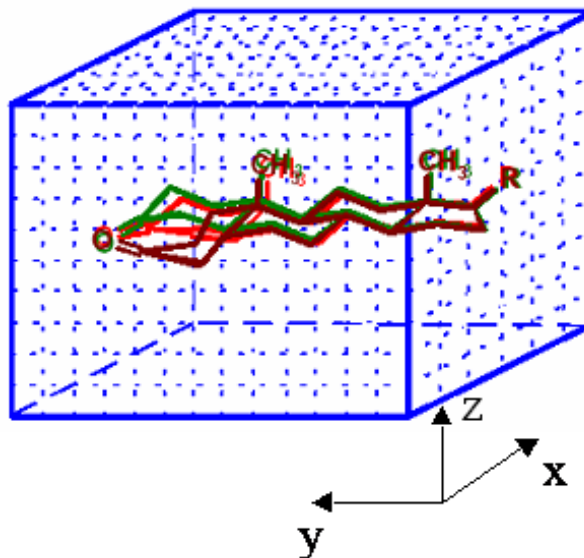


Figura 10. Representação da caixa 3D na qual os compostos alinhados são inseridos no processo do CoMFA (Adaptado de Kubinyi, 1997).

A.1.5. Construção dos Campos Moleculares

A premissa do método CoMFA é que as interações mais importantes entre o ligante e o receptor são de natureza estérica e eletrostática (Cramer *et al.*, 1988), visto que os fatores estéricos e eletrônicos das moléculas desempenham um papel importante nas interações com o receptor (Huuskonen, 2001; Klebe, 1994).

As energias de interação estéricas e eletrostáticas são calculadas com um átomo de prova que mimetiza átomos do receptor. Como átomos de prova são usados geralmente oxigênio com hibridação sp^3 e carga -1, carbono com hibridação sp^3 e carga +1 e hidrogênio com carga +1 (Brito *et al.*, 2008; Brito, 2004; Rodrigues *et al.*, 2000; Cho & Tropsha, 1995).

As energias estéricas (de van der Waals) são obtidas pelo potencial de Lennard-Jones (Equação 2) e as energias eletrostáticas através do potencial de Coulomb (Equação 3).

$$V_{LJ} = \sum \left[\frac{C_{12}(ij)}{r_{ij}^{12}} - \frac{C_6(ij)}{r_{ij}^6} \right] \quad \text{Equação 2}$$

Na Equação 2, **C** é um parâmetro específico de Lennard-Jones que difere para cada átomo e **r** é a distância entre os dois átomos que interagem. O termo **r**¹² descreve o potencial repulsivo de curto alcance devido às distorções das nuvens eletrônicas e predomina quando a separação entre os átomos que interagem é pequena. O termo **r**⁶ descreve o potencial atrativo de longo alcance e predomina quando a separação aumenta em magnitude (Levine, 2000, pg. 670; Martínez *et al.*, 2007, pg. 424; Leach, 2001, 271; Scott *et al.*, 1999).

$$V_C = \frac{1}{4\pi\epsilon} \sum \frac{q_i q_j}{r_{ij}} \quad \text{Equação 3}$$

Na Equação 3, **qi** e **qj** são as cargas elétricas dos átomos **i** e **j** que estão interagindo a uma distância **rij** e **ε** é a permissividade elétrica do meio (Levine, 2000, pg.671; Martínez *et al.*, 2007, pg. 424; Leach, 2001, 271; Scott *et al.*, 1999).

Na análise de CoMFA, são gerados contornos que definem regiões do espaço como sendo favoráveis ou desfavoráveis para a interação ligante-receptor (Kubinyi, 1997).

A.1.6. Técnicas Estatísticas Aplicadas no CoMFA

As energias de interação obtidas em cada ponto da grade da caixa 3D são tabeladas para cada composto, junto às respectivas atividades biológicas. O resultado disso é uma tabela com centenas de colunas, conforme ilustrado na Figura 11. Posteriormente, as energias obtidas serão relacionadas às atividades biológicas

pela técnica de mínimos quadrados parciais (*Partial Least Squares*, PLS), conforme ilustrado na Figura 11 (Glenn *et al.*, 1989; Cramer *et al.*, 1988).

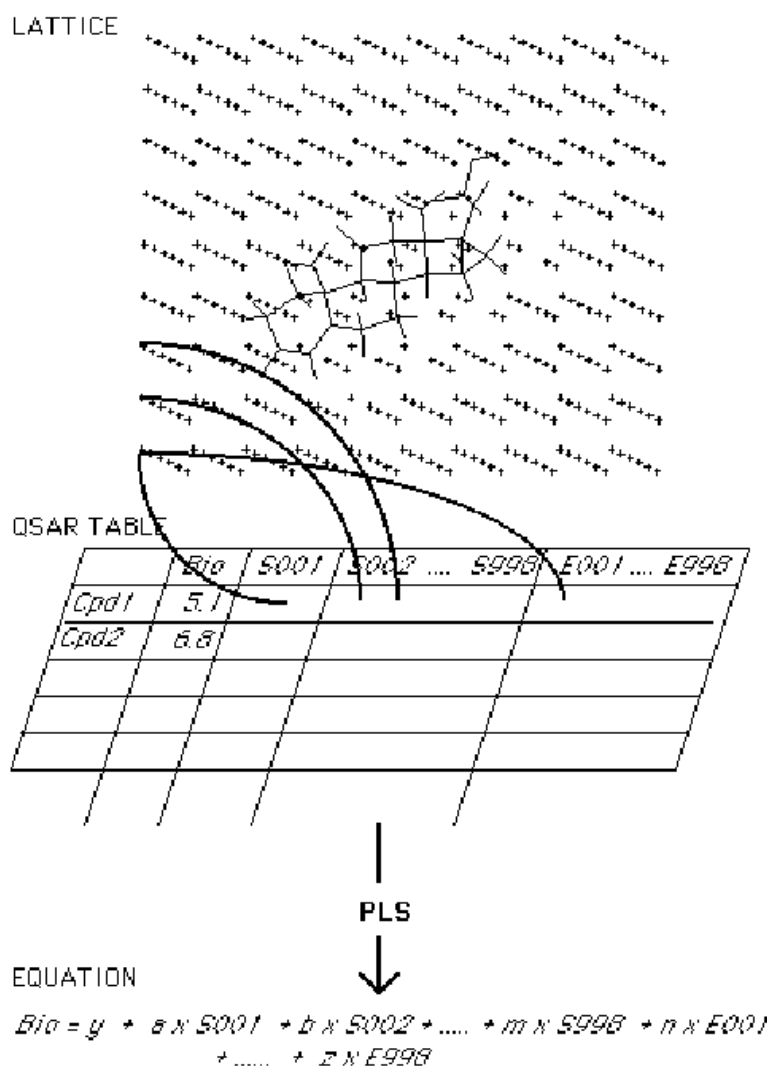


Figura 11. Esquema do processo de geração das energias no CoMFA e tabelamento dos resultados (Reproduzido de Kubinyi, 1997).

A abordagem adotada no CoMFA gera um grande número de variáveis que precisam ser analisadas estatisticamente. A análise dos dados por regressão padrão é praticamente impossível. Correlacionam-se as energias com a atividade biológica utilizando o método de Mínimos Quadrados Parciais (*Partial Least Squares*, PLS) associado à validação cruzada por exclusão exaustiva (*Leave-One-Out Cross-*

validation, LOO_{CV}), que assegura a significância estatística da equação final gerada pelo método (Leach, 2001, pp. 706; Livingstone, 1995, pp. 134).

A.1.6.1. Mínimos Quadrados Parciais

O método PLS é similar à regressão múltipla e é usado especialmente quando o número de variáveis é superior ao número de graus de liberdade. Pode ser melhor descrito como uma tentativa de correlacionar a variável dependente (resposta biológica) com as variáveis independentes (descritores) de uma só vez, em contraste com o método de regressão múltipla, onde em cada etapa é identificada apenas uma variável independente que melhor correlaciona-se à variável dependente (Leach, 2001, pp. 706; Livingstone, 1995, pp. 153).

PLS é uma técnica muito utilizada em análises de QSAR em que um grande número de variáveis são geradas, pois reduz o problema construindo variáveis latentes (componentes principais), que são combinações lineares das variáveis independentes originais (Migliavacca, 2003; Hopfinger *et al.*, 1997; Livingstone, 1995; Rogers & Hopfinger, 1994; Glenn *et al.*, 1989). PLS expressa a variável dependente (Y) por combinações lineares das variáveis independentes originais (X), como mostrado na Equação 4, onde t_1 , t_2 , etc, são as variáveis latentes (ou componentes principais) (Leach, 2001, pg. 706).

$$Y = b_1 t_1 + b_2 t_2 + b_3 t_3 + \dots + b_m t_m$$

Equação 4

Onde:

$$t_1 = c_{11} X_1 + c_{12} X_2 + \dots + c_{1p} X_p$$

$$t_2 = c_{21} X_1 + c_{22} X_2 + \dots + c_{2p} X_p$$

$$t_m = c_{m1} X_1 + c_{m2} X_2 + \dots + c_{mp} X_p$$

A.1.6.2. Validação Cruzada *Leave-One-Out*

A técnica LOO_{cv} é usada na grande maioria dos estudos de CoMFA e em estudos de QSAR em geral (Bak & Polanski, 2006; Romeiro *et al.*, 2006), pois assegura a significância estatística do modelo. Na técnica LOO_{cv} cada composto é sistematicamente excluído uma vez do banco de dados e sua atividade é prevista por um modelo gerado a partir dos demais compostos que compõem o banco, conforme ilustrado na Figura 12.

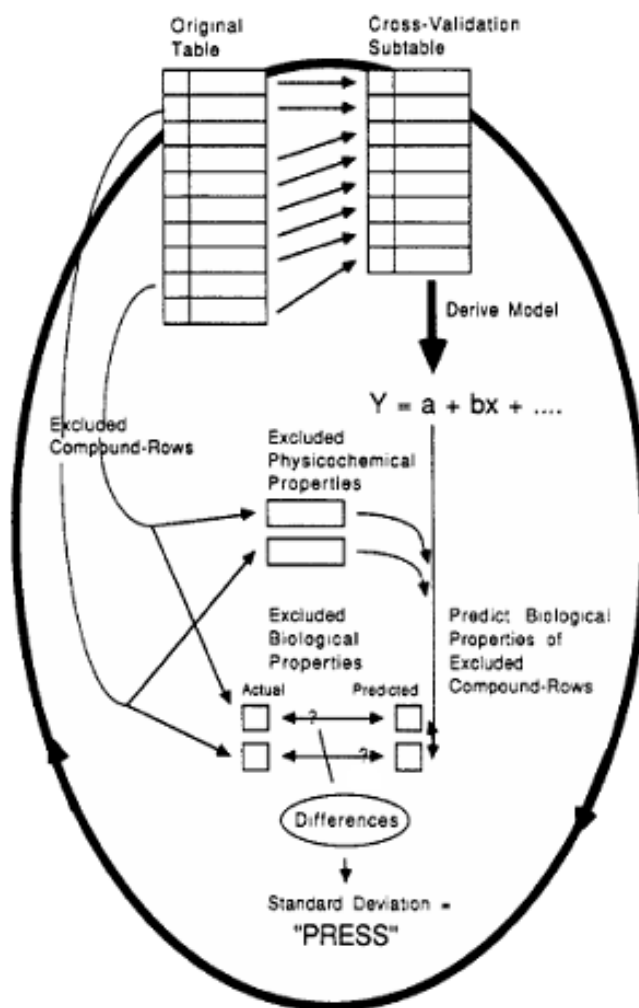


Figura 12. Esquema do processo de validação cruzada pelo método *Leave-One-Out*. (Reproduzido de Kubinyi, 1997).

A.1.7. Vantagens e Limitações do CoMFA

As principais vantagens do CoMFA são a aplicação a conjunto de dados heterogêneos e a visualização das interações estéricas e eletrostáticas favoráveis e desfavoráveis com o receptor pelos mapas de contorno 3D (Bhongade & Gadad, 2004). Estes mapas de contorno devem ser interpretados em termos da estrutura do receptor, se ele for conhecido, e caso não seja, pode-se inferir possíveis sub-sítios de interação a partir desses mapas.

Como limitações do CoMFA, existem as incertezas a respeito da conformação bioativa e dos modos possíveis de ligação dos ligantes com o receptor, os níveis de *cutoff* utilizados (valores de corte nas energias de interação do átomo de prova com os ligantes) e o fato de ser aplicável somente a dados biológicos e farmacológicos *in vitro*, uma vez que nestes não estão envolvidos aspectos relacionados à fase farmacocinética (absorção, distribuição, metabolismo e excreção), enquanto em dados *in vivo* esses aspectos são relevantes.

B. QSAR-3D Dependente da Estrutura do Receptor

A abordagem RD é aplicada nos casos onde a estrutura do receptor é conhecida experimentalmente por técnicas de cristalografia de raios-X ou de ressonância magnética nuclear de hidrogênio, carbono ou nitrogênio (*Nuclear Magnetic Resonance*, NMR¹H, NMR¹³C, NMR¹⁵N), ou pode ser obtida por técnicas de modelagem comparativa, onde se faz o ajuste da biomacromolécula desejada a uma estrutura similar do receptor por meio da modelagem molecular (Marshall & Cramer, 1988).

Como exemplos de metodologias de QSAR-3D RD, há a “Análise Comparativa da Energia de Ligação” (*Comparative Binding Energy Analysis*, COMBINE), desenvolvida por Ortiz e colaboradores (Ortiz *et al.*, 1995) e o “Campo de Força de Energia Livre” (*Free-Energy Force Field*, FEF), desenvolvido por Tokarski & Hopfinger. Na metodologia FEF são utilizados como descritores para cada ligante e resíduos selecionados no receptor, as energias ligadas (comprimento de ligação, ângulo e torsão) e não ligadas (Lennard-Jones e Coulomb), antes e após a complexação com o receptor (Venkatarangan & Hopfinger, 1999; Tokarski & Hopfinger, 1997).

De fato, o conhecimento a respeito das forças de interação entre o ligante e resíduos do receptor é importante para compreender a variação da potência biológica de uma série de compostos (Vinkers *et al.*, 2003). A força das interações químicas pode ser estimada com campos de força de mecânica molecular, ancorando-se os compostos no sítio de ligação do receptor (Leach, 2002). Essas interações químicas podem ser expressas pelas interações de van der Waals (na forma do potencial de Lennard-Jones, Equação 2) e interações eletrostáticas (na forma do potencial de Coulomb, Equação 3).

Todd e Freire (1999), em um estudo experimental por calorimetria do processo termodinâmico da ligação de inibidores na protease do HIV-1, observaram que o efeito da estabilização do inibidor na enzima não apresenta a mesma ordem de magnitude em relação a todos os resíduos da enzima. Resíduos em certas regiões da enzima são significativamente mais afetados do que outros em outras regiões, isto é, a interação ocorre com resíduos específicos. Em particular, a ligação do inibidor tem um efeito significativo nas regiões que definem o sítio.

Kulkarni e Kulkarni (1999), trabalhando com a predição experimental da afinidade de ligação de inibidores não peptídicos na enzima protease do HIV-1, observaram que a energia de interação total é particionada em interações em diferentes sub-sítios e que a interação com apenas alguns resíduos contribui para diferenças nas potências desses inibidores.

Esses dados corroboram que se pode fazer estudos de QSAR-3D RD de inibidores enzimáticos considerando apenas a interação dos ligantes com resíduos específicos da enzima, sem incorrer em erros graves.

B.1. Algoritmos Genéticos em Estudos de QSAR-3D Dependente do Receptor

B.1.1. Algoritmos Genéticos

Os Algoritmos Genéticos (*Genetic Algorithms*, GA) são derivados de uma classe de métodos estocásticos, os Algoritmos Evolucionários (*Evolutionary Algorithms*, EA) (Magalhães *et al.*, 2007, pg. 508). Os GA são inspirados no princípio de seleção natural e evolução Darwiniana (Linden, 2006, pg. 32; Li *et al.*, 1999; Devillers, 1996, pg. 3).

B.1.2. As Bases da Seleção Natural Aplicadas aos GAs

Na natureza, apenas os organismos mais aptos conseguem vencer a competição por recursos como alimento e espaço, sobrevivem e se reproduzem, gerando proles e permitindo a transmissão de sua herança genética por meio dos genes contidos nos cromossomos. A adaptação às mudanças do meio-ambiente é essencial para a sobrevivência dos indivíduos das espécies. Assim, a seleção natural leva à sobrevivência dos melhores indivíduos, e implicitamente dos melhores genes. O processo de reprodução sexuada permite a diversificação dos genes de uma espécie (Russel & Norvig, 2004, p. 117; Devillers, 1996, pg. 4-5).

Em um estudo de QSAR empregando GA são criados modelos aleatoriamente e aqueles modelos com melhores valores estatísticos propagam seu “material genético”, ou seja, suas características, por operação de cruzamento (do inglês, *crossover*), onde há a combinação de variáveis independentes de dois bons modelos (pais) para criar um novo modelo (filho) (Linden, 2006, pp. 67; Russel & Norvig, 2004, p. 117; Hemmateenejad *et al.*, 2002).

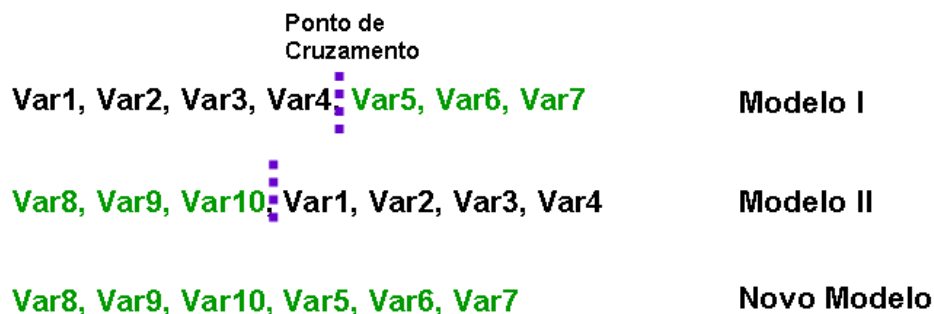


Figura 13. Esquema da operação de cruzamento dentro da técnica de algoritmos genéticos. **Var** representa um descritor no estudo de QSAR.

Na geração seguinte, os modelos com melhores pontuações são mantidos e novos modelos são criados por operações de cruzamento e de mutação (do inglês, *mutation*). A operação de mutação consiste na criação de um novo modelo pela introdução aleatória de uma nova variável no modelo recém-criado por cruzamento, o que ajuda a manter diversidade suficiente na população (Linden, 2006, pp. 69; Niculescu, 2003).

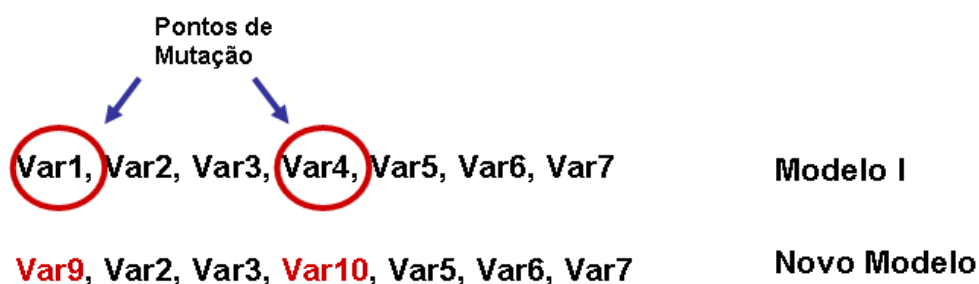


Figura 14. Esquema da operação de mutação dentro da técnica de algoritmos genéticos. **Var** representa um descritor no estudo de QSAR.

B.1.3. Vantagens e Aplicações dos Algoritmos Genéticos na Modelagem Molecular

Os GAs são especialmente úteis na resolução de problemas com um grande número de variáveis, por possibilitarem uma amostragem eficiente (Linden, 2006, pg. 49.; Niculescu, 2003; Romeiro *et al.*, 2006). Eles vêm sendo aplicados a estudos diversos como buscas conformacionais, que são essenciais na metodologia de *docking*; estudos de SAR, onde auxiliam a busca por farmacóforos; e QSAR, ajudando a correlação de descritores com atividades biológicas e o planejamento de fármacos (Magalhães *et al.*, 2007, pg. 508; Romeiro *et al.*, 2006; Leach, 2001, pg. 479).

B.1.4. Aproximação da Função Genética com Mínimos Quadrados Parciais

O algoritmo aproximação da função genética (*Genetic Function Approximation*, GFA) é um algoritmo genético utilizado para a criação de modelos de QSAR (Rogers, 1996, pp. 88-92; Hahn & Rogers, 1995; Rogers & Hopfinger, 1994), onde as variáveis são chamadas de funções de base.

O GFA aplica os mesmos procedimentos descritos anteriormente para os GAs, combinando à técnica de PLS, que vem se tornando o método estatístico de escolha para a maioria dos estudos de QSAR (Brito, 2004; Romeiro *et al.*, 2006; Luco & Ferretti, 1997; Dunn & Rogers, 1996, pg.109-118; Hahn & Rogers, 1995).

A técnica de GFA-PLS tem como característica principal a geração de múltiplos bons modelos, ao invés da otimização de apenas um único modelo (Rogers & Hopfinger, 1994). Outros autores relatam o uso da análise combinada de

algoritmos genéticos e PLS (Sodero, 2007; Pita, 2006; Cunha, 2006; Romeiro, 2002; Albuquerque, 1997; Hasegawa, 1997).

2. OBJETIVOS

Inserido no contexto da Química Medicinal, que engloba o planejamento racional de substâncias bioativas e, tendo em vista a importância da transcriptase reversa (RT) como alvo terapêutico no combate à AIDS, foi selecionada uma série inédita em estudos de QSAR-3D de 74 compostos pertencente à classe química das diidro-alcóxi-benzil-oxopirimidinas (DABOs), inibidores não-nucleosídeos da RT do HIV-1 (Ragno *et al.*, 2004; Mai *et al.*, 1999; Mai *et al.*, 1997; Mai *et al.*, 1995), objetivando a construção e a avaliação de modelos de QSAR-3D independente e dependente da estrutura da enzima, que serão úteis na compreensão do modo de interação dos DABOs e, conseqüentemente, no planejamento e proposição de novos inibidores não nucleosídeos da RT para o tratamento da AIDS.

3. METODOLOGIA

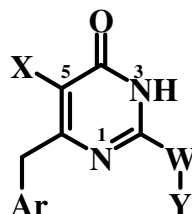
3.1. BANCO DE DADOS ESTRUTURAL E BIOLÓGICO

Uma série de 74 compostos da classe dos DABOs foi selecionada para os estudos de QSAR-3D independente e dependente do receptor, a partir de uma série de trabalhos publicada pelo grupo de Mai, Ragno e colaboradores (Ragno *et al.*, 2004; Mai *et al.*, 1999; Mai *et al.*, 1997; Mai *et al.*, 1995).

O perfil de atividade biológica desses compostos foi avaliado em ensaios *in vitro* frente a enzima RT do HIV-1, segundo o mesmo protocolo farmacológico (Mai *et al.*, 1995). As potências inibitórias dadas em IC_{50} (μM) foram transformadas em pIC_{50} (M), que corresponde ao logaritmo do inverso da concentração mínima capaz de inibir 50% da atividade enzimática ($-\log IC_{50}$). A Tabela 5 mostra as estruturas químicas e os respectivos valores de atividade biológica (pIC_{50}) dos 74 compostos da série dos DABOs.

Os compostos contendo um centro estereogênico (Y=*sec*-butil, Tabela 5), correspondendo, portanto, a uma mistura racêmica, foram definidos na configuração absoluta *R* e os respectivos valores de IC_{50} originais foram multiplicados por dois. Como este centro estereogênico está localizado em uma cadeia alquílica lateral, ele foi considerado de menor importância e o enantiômero *R* foi arbitrariamente definido como o eutômero.

Tabela 5. Estruturas e potências inibitórias (pIC_{50} , M) dos derivados S- e NH-DABOs sobre a RT do HIV-1 (Ragno *et al.*, 2004; Mai *et al.*, 1999; Mai *et al.*, 1997; Mai *et al.*, 1995).



# ^a	X ^b	Ar	W-Y ^c	pIC_{50} ^d	# ^a	X ^b	Ar	W-Y ^c	pIC_{50} ^d
1	Me	2-naftil	S- <i>sec</i> -Bu	4,23	38	H	2,6-di-F-Ph	S-Me	6,10
2	H	1-naftil	S-ciclopentil	4,31	39	Me	2-Cl-Ph	S- <i>sec</i> -Bu	6,10
3	Me	1-naftil	S-ciclopentil	4,35	40	Me	2-F-Ph	S- <i>sec</i> -Bu	6,10
4	Me	4-F-Ph	S- <i>sec</i> -Bu	4,59	41	Me	3-NO ₂ -Ph	S- <i>sec</i> -Bu	6,10
5	Me	4-Cl-Ph	S- <i>sec</i> -Bu	4,77	42	H	2-F-Ph	S- <i>sec</i> -Bu	6,22
6	H	1-naftil	S- <i>sec</i> -Bu	4,79	43	H	3-NO ₂ -Ph	S- <i>sec</i> -Bu	6,22
7	H	2-naftil	S- <i>sec</i> -Bu	4,83	44	H	2,6-di-Cl-Ph	S- <i>terc</i> -Bu	6,22
8	H	4-F-Ph	S- <i>sec</i> -Bu	4,83	45	H	2,6-di-Cl-Ph	S- <i>n</i> -Bu	6,30
9	H	4-Cl-Ph	S- <i>sec</i> -Bu	5,02	46	H	2,6-di-Cl-Ph	S-ciclopentil	6,40
10	H	Ph	S- <i>terc</i> -Bu	5,07	47	H	2,6-di-F-Ph	S- <i>n</i> -Bu	6,70
11	H	3-Me-Ph	S- <i>terc</i> -Bu	5,09	48	H	2,6-di-F-Ph	S- <i>terc</i> -Bu	6,70
12	Me	3-Me-Ph	S- <i>sec</i> -Bu	5,27	49	H	2,6-di-Cl-Ph	S- <i>sec</i> -Bu	6,70
13	Me	2,6-di-Cl-Ph	S-ciclohexil	5,31	50	Me	2,6-di-Cl-Ph	S- <i>sec</i> -Bu	6,92
14	Me	Ph	S-Me	5,31	51	H	2,6-di-F-Ph	S- <i>sec</i> -Bu	7,00
15	Me	Ph	S- <i>sec</i> -Bu	5,32	52	Me	2,6-di-F-Ph	S- <i>sec</i> -Bu	7,00
16	Me	3-Me-Ph	S- <i>terc</i> -Bu	5,34	53	H	2,6-di-F-Ph	S-ciclohexil	7,05
17	Me	Ph	S-ciclohexil	5,37	54	Me	2,6-di-F-Ph	S- <i>terc</i> -Bu	7,05
18	H	3-Cl-Ph	S- <i>sec</i> -Bu	5,42	55	H	2,6-di-F-Ph	S-ciclopentil	7,10
19	Me	4-NO ₂ -Ph	S- <i>sec</i> -Bu	5,44	56	Me	2,6-di-F-Ph	S-ciclopentil	7,10
20	Me	3-Me-Ph	S-ciclopentil	5,47	57	H	2,6-di-F-Ph	NH-ciclopentil	7,15
21	H	2-Cl-Ph	S- <i>sec</i> -Bu	5,49	58	H	2,6-di-F-Ph	S- <i>iso</i> -Pr	7,30
22	Me	3-F-Ph	S- <i>sec</i> -Bu	5,52	59	Me	2,6-di-F-Ph	NH-ciclopentil	7,52
23	H	2,6-di-Cl-Ph	S-Me	5,52	60	Me	1-naftil	S- <i>sec</i> -Bu	4,35
24	H	Ph	S-ciclohexil	5,52	61	H	2-naftil	S-ciclohexil	4,48
25	H	3-Me-Ph	S- <i>iso</i> -Pr	5,54	62	H	Ph	S- <i>sec</i> -Bu	5,27
26	H	Ph	S-ciclopentil	5,55	63	Me	Ph	S-ciclopentil	5,47
27	H	3-Me-Ph	S-ciclohexil	5,59	64	H	3-Me-Ph	S-ciclopentil	5,59
28	Me	3-Me-Ph	S-Me	5,60	65	Me	Ph	S- <i>iso</i> -Pr	5,60
29	Me	3-Me-Ph	S- <i>iso</i> -Pr	5,60	66	H	3-Me-Ph	S- <i>sec</i> -Bu	5,62
30	H	4-NO ₂ -Ph	S- <i>sec</i> -Bu	5,62	67	Me	3-Cl-Ph	S- <i>sec</i> -Bu	5,74
31	Me	3-Me-Ph	S-ciclohexil	5,66	68	H	3-F-Ph	S- <i>sec</i> -Bu	5,92
32	Me	Ph	S- <i>terc</i> -Bu	5,72	69	H	2-NO ₂ -Ph	S- <i>sec</i> -Bu	6,22
33	Me	2,6-di-Cl-Ph	S-ciclopentil	5,80	70	H	2,6-di-Cl-Ph	S-ciclohexil	6,40
34	H	2,6-di-Cl-Ph	S- <i>iso</i> -Pr	5,89	71	Me	2,6-di-F-Ph	S-Me	6,70
35	Me	2,6-di-Cl-Ph	S- <i>iso</i> -Pr	5,94	72	Me	2,6-di-F-Ph	S- <i>n</i> -Bu	7,05
36	Me	2,6-di-Cl-Ph	S- <i>n</i> -Bu	5,94	73	Me	2,6-di-F-Ph	S-ciclohexil	7,15
37	Me	2,6-di-Cl-Ph	S- <i>terc</i> -Bu	5,96	74	Me	2,6-di-F-Ph	S- <i>iso</i> -Pr	7,30

a) Números sublinhados correspondem aos compostos do conjunto de teste (60-74). b) Base pirimidínica uracila (X=H) ou timina (X=Me). c) Séries S-DABO (W=S; Y=alquila) e NH-DABO (W=NH; Y=alquila). d) Os valores de IC_{50} originais dos compostos contendo centro estereogênico (W=S; Y=*sec*-Bu) foram multiplicados por dois e apenas os isômeros *R* foram considerados nos estudos de QSAR-3D.

3.1.1. Perfil Estrutural dos DABOs

Os DABOs estudados possuem o perfil estrutural ilustrado na Figura 15. Tendo como referência o núcleo pirimidínico dos DABOs, as variações estruturais são:

- (A) na posição **C2**, onde **W** pode ser –S– (S-DABOs) ou –NH– (NH-DABOs) e **Y** pode ser alquila ou cicloalquila (**W–Y** = S–Me, S–*iso*-Pr, S–*n*-Bu, S–*sec*-Bu, S–*terc*-Bu, S–ciclopentila, S–ciclohexila e NH–ciclopentila);
- (B) na posição **C5**, onde **X** pode ser –H (uracila) ou –CH₃ (timina);
- (C) na posição **C6**, onde **Ar** pode ser fenila (com diversos substituintes), 1-naftila ou 2-naftila.

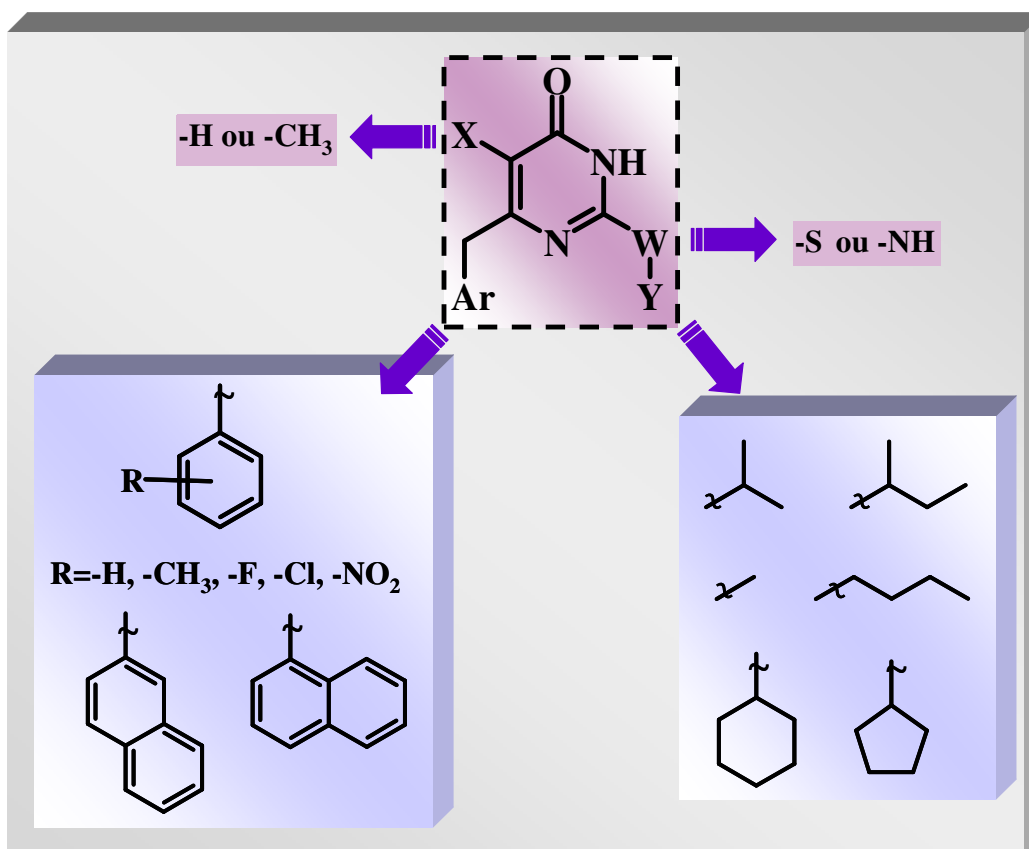


Figura 15. Perfil estrutural dos S- e NH-DABOs utilizados nos estudos de QSAR-3D.

3.1.2. Definição dos Conjuntos de Treinamento e de Teste

Para os estudos de QSAR-3D, o banco de dados de 74 compostos da série dos DABOs foi dividido em um conjunto de treinamento, contendo 59 compostos (**1-59**), e um conjunto de teste, contendo 15 compostos (**60-71**), mostrados na Tabela 5. Neste trabalho, os compostos do conjunto de teste representam cerca de 20% do total de compostos.

A Figura 16 mostra a distribuição dos valores de atividade biológica (pIC_{50}) dos compostos dos conjuntos de treinamento (A) e de teste (B). No conjunto de treinamento os valores de pIC_{50} variam de 4,23 a 7,52 M e no conjunto de teste os valores de pIC_{50} variam de 4,35 a 7,30 M. Em ambos os conjuntos, os compostos estão regularmente distribuídos por toda a faixa de atividade, que compreende cerca de quatro unidades logarítmicas, e apresentam a mesma diversidade estrutural.

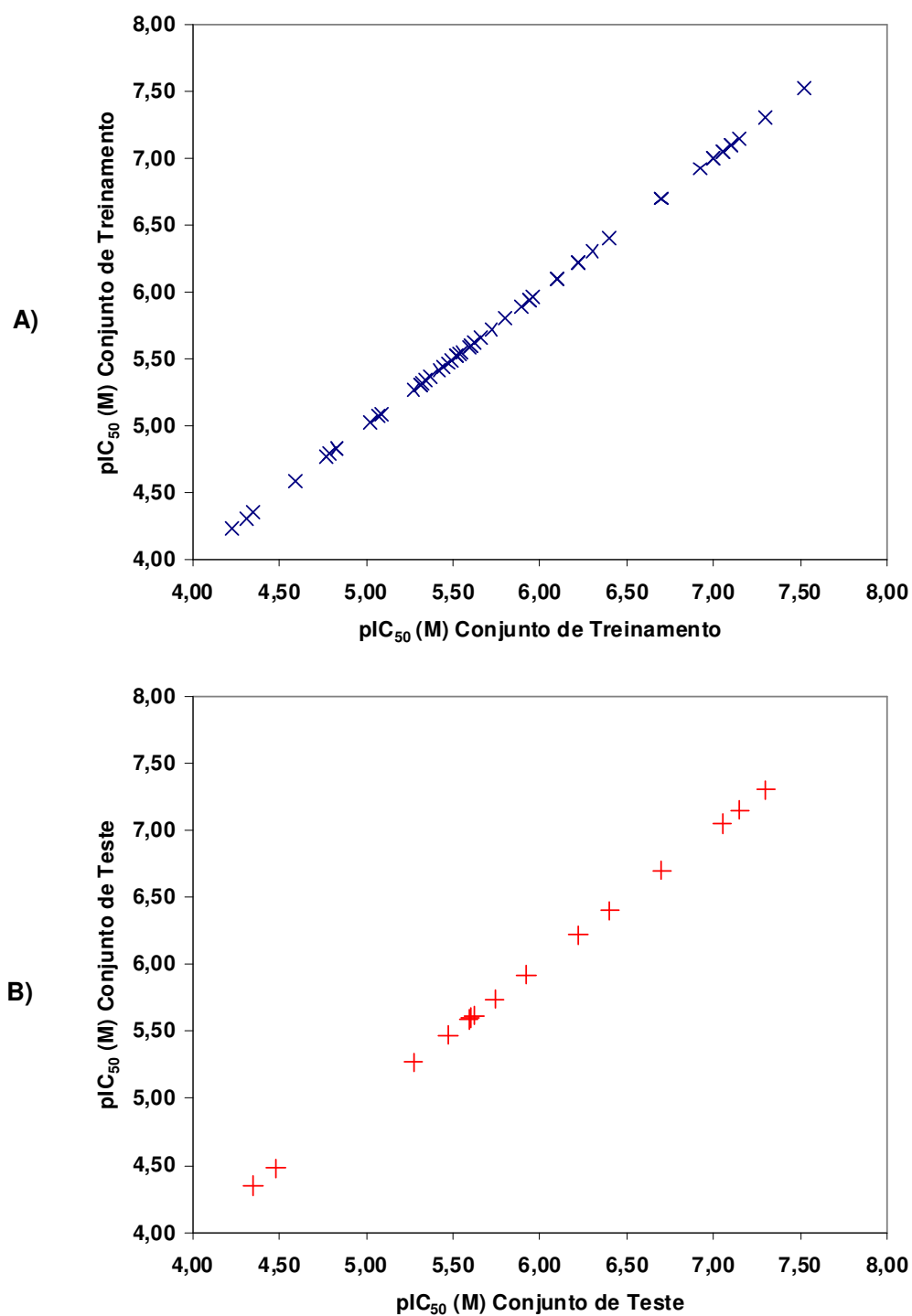


Figura 16. Distribuição dos valores de atividade biológica (pIC₅₀) dos compostos do conjunto de treinamento (A) e do conjunto de teste (B).

3.2. QSAR-3D INDEPENDENTE E DEPENDENTE DO RECEPTOR

A descrição da metodologia das duas abordagens de QSAR-3D empregadas neste trabalho, *i.e.* os estudos independente e dependente da estrutura do receptor, está dividida em Parte A (QSAR-3D Independente do Receptor, CoMFA) e Parte B (QSAR-3D Dependente do Receptor).

Mas antes serão descritas a construção e a otimização das estruturas 3D dos ligantes, etapas comuns aos dois métodos.

3.2.1. Construção e Otimização das Estruturas 3D dos Ligantes

Na ausência de um composto da classe dos DABOs co-cristalizado com a RT do HIV-1, foi usada como conformação bioativa de referência na etapa de construção das estruturas dos DABOs, a estrutura do MKC-442 (emivirina, Figura 17), co-cristalizado à enzima RT nativa, disponível no PDB sob o código 1RT1, com resolução de 2,55 Å (Hopkins et al., 1996). O MKC-442 é um inibidor não-nucleosídico da classe dos HEPTs, estruturalmente relacionado aos DABOs.

Assim, as estruturas 3D dos 74 DABOs foram construídas a partir da estrutura 3D do MKC-442, extraída do arquivo 1RT1, usando a opção *Build* do programa Spartan'06 (Wavefunction, Inc.). Todos os compostos foram otimizados geometricamente pelo método de mecânica molecular utilizando o campo de força MMFF94 e em seguida pelo método semi-empírico utilizando o Hamiltoniano AM1 (Dewar, 1985).

A Figura 17 mostra a sobreposição entre as estruturas 3D do MKC-442 e do NH-DABO **59**, o mais potente da série. Esta etapa de construção e otimização das

estruturas dos ligantes foi comum às duas metodologias de QSAR-3D (independente e dependente do receptor) empregadas neste trabalho.

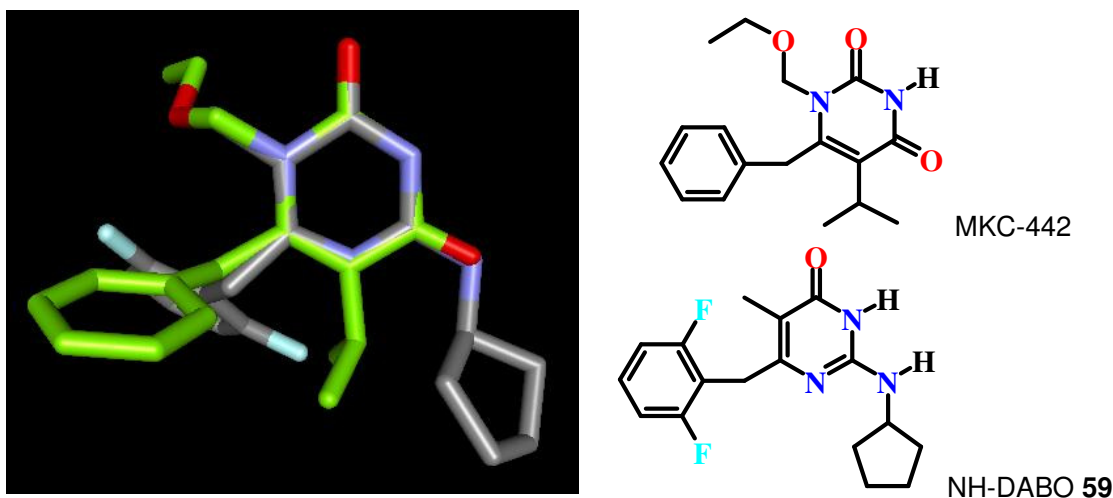


Figura 17. Sobreposição entre as estruturas 3D do MKC-442 (átomos de carbono coloridos em verde) e do NH-DABO 59 (átomos de carbono coloridos em cinza). Para melhor visualização, os átomos de hidrogênio foram omitidos.

3.2.2. PARTE A – QSAR-3D INDEPENDENTE DO RECEPTOR (COMFA)

3.2.2.1. Definição da Hipótese Farmacofórica

A hipótese farmacofórica utilizada na orientação e sobreposição dos derivados DABOs para conseguir um alinhamento consistente, teve como referência o modo de ligação do composto MKC-442 na enzima RT (código no PDB 1RT1), conforme descrito no item “Construção e Otimização das Estruturas 3D dos Ligantes”.

3.2.2.2. Assinalamento das Cargas Atômicas Parciais

No estudo de CoMFA, a contribuição eletrostática no modelo de QSAR-3D depende do esquema usado no assinalamento das cargas atômicas parciais, devido a variação das cargas em função das diferentes abordagens teóricas que podem ser empregadas.

Desta forma, para verificar o efeito das cargas atômicas parciais sobre a qualidade dos modelos de CoMFA, o assinalamento das cargas dos inibidores foi feito pelo ajuste das cargas a um potencial eletrostático molecular restringido (*Restrained Molecular Electrostatic Potential*, RESP), calculado em quatro níveis de teoria, todos disponíveis no programa *Spartan`06* (Wavefunction, Inc.): (a) DFT (*Density Funcional Theory*), usando o funcional B3LYP e o conjunto de bases 6-31G*; (b) *ab initio* Hartree-Fock (HF), usando o conjunto de bases 6-31G*; e semi-empíricos (c) AM1 e (d) PM3. O conjunto dos compostos (estruturas otimizadas e cargas assinaladas) foi exportado para o programa *Sybyl v.7.2* (Tripos, Inc.), onde o módulo CoMFA está disponível.

3.2.2.3. Definição dos Alinhamentos

As etapas de definição da hipótese farmacofórica, da conformação bioativa e do alinhamento são as etapas mais importantes do CoMFA, uma vez que as energias de interação dependem, significativamente, da posição relativa dos grupos funcionais presentes nos ligantes (Peterson *et al.*, 2006; Cramer *et al.*, 1988).

O alinhamento consiste na escolha de um conjunto de átomos que irá orientar o posicionamento relativo das moléculas dentro da caixa virtual. A escolha dos átomos pode ser feita selecionando-se átomos ou grupos funcionais conservados

em todas as estruturas químicas. Para obter a melhor sobreposição dos compostos, foram testados três alinhamentos (Figura 18), usando como referência o composto mais potente da série, **59**.

No **Alinhamento 1**, os seis átomos do anel 4-oxo-pirimidina foram selecionados (**N1**, **C2**, **N3**, **C4**, **C5** e **C6**). No **Alinhamento 2**, três átomos do anel 4-oxo-pirimidina (**N1**, **N3** e **C5**) e três átomos do anel aromático do grupo 6-benzila (**C1'**, **C3'** e **C5'**) foram selecionados. No **Alinhamento 3**, foram selecionados três átomos do anel 4-oxo-pirimidina (**N1**, **N3** e **C5**), o átomo de oxigênio do grupo carbonila, o átomo de carbono da posição *ipso* (**C1'**) do anel aromático do grupo 6-benzila e o átomo de carbono do anel ciclopentila (**C1**).

Os substituintes do grupo 6-benzila foram orientados na mesma direção relativa para uma sobreposição ótima entre os compostos. As etapas de alinhamento e sobreposição foram realizadas usando o comando *Alignment Database* do programa *Sybyl v.7.2 (Tripos, Inc.)*.

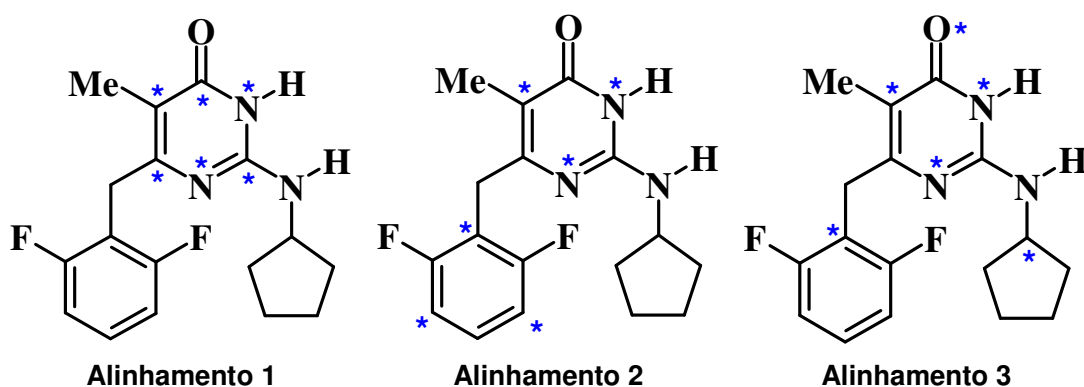


Figura 18. Composto de referência NH-DABO **59** marcado com asteriscos nos átomos usados para os três Alinhamentos testados.

3.2.2.4. Definição dos Níveis de *Cutoff*

Mantendo-se fixos o melhor alinhamento e as cargas atômicas parciais que forneceram os melhores resultados, estudou-se ainda a influência de diferentes níveis de *cutoff* sobre o q^2 resultante do melhor modelo. Os níveis de *cutoffs* foram truncados para os campos estéricos e eletrostáticos em 30 kcal.mol⁻¹ (*default*), 20 kcal.mol⁻¹ e 10 kcal.mol⁻¹.

3.2.2.5. Definição da Caixa e Obtenção das Variáveis Independentes

O conjunto de compostos previamente sobrepostos segundo os Alinhamentos 1, 2 e 3, foi inserido numa caixa 3D virtual, regularmente dividida em grades de 2,0 Å de espaçamento (valor padrão) e com limites de 4,0 Å, em todas as direções da caixa.

As variáveis independentes na metodologia de CoMFA são as energias de interação estéricas e eletrostáticas calculadas quando um tipo definido de átomo de prova, posicionado em espaços regulares da caixa, interage com cada um dos compostos inseridos na caixa.

Inicialmente, os modelos de CoMFA foram construídos mantendo-se fixos o tipo de átomo de prova (C_{sp}³ com carga +1, átomo de prova padrão) e o valor de corte de energia (*cutoff*) para ambas as interações estéricas e eletrostáticas (valor de corte padrão de 30 kcal.mol⁻¹), e testando os três alinhamentos e os quatro tipos de assinalamento de cargas atômicas parciais para os ligantes.

3.2.2.6. Obtenção dos Modelos de CoMFA

As equações ou modelos de CoMFA são obtidas pelo uso da técnica estatística de análise de regressão múltipla de mínimos quadrados parciais, PLS

(Glenn *et al.*, 1989), correlacionando a variação nas variáveis independentes (*i.e.*, valores de energia de interação estérica e eletrostática) com a variação na variável dependente (*i.e.*, valores de pIC_{50}) do conjunto de treinamento dos inibidores.

Como as metodologias de QSAR-3D geram um grande número de variáveis independentes, o PLS é uma técnica muito utilizada neste tipo de análise, por ser capaz de reduzir o problema a ser tratado, construindo variáveis latentes (componentes principais), que são combinações lineares das variáveis independentes originais (Migliavacca, 2003; Hopfinger *et al.*, 1997; Livingstone, 1995; Rogers & Hopfinger, 1994; Glenn *et al.*, 1989).

3.2.2.7. Validação Interna dos Modelos de CoMFA

A técnica LOO_{CV} (descrita na Introdução) foi usada para acessar a capacidade preditiva interna dos modelos de CoMFA (Kubinyi, 1997a). Nessa técnica, cada um dos compostos é sistematicamente excluído do conjunto de treinamento e a potência de cada composto excluído (pIC_{50}) é predita por um novo modelo (com as mesmas variáveis, porém com novos coeficientes) derivado a partir dos compostos remanescentes do banco de dados (Kubinyi, 1997a). A validação cruzada fornece o número ótimo de componentes principais (PC) e o valor de q^2 após a validação cruzada (r^2_{cv}), índice que traduz a capacidade preditiva do modelo (Equações 5 e 6).

Os resultados da validação cruzada foram analisados considerando que valores de r^2_{cv} acima de 0,3 indicam que a probabilidade de correlação ao acaso é menor do que 5% (Peterson *et al.*, 2006).

$$r^2_{cv} = 1 - \frac{PRESS}{\sum(Y - Y_{média})^2}$$
Equação 5

$$PRESS = \sum(Y - Y_{pred})^2$$
Equação 6

Nas Equações 5 e 6, **PRESS** é a soma dos quadrados do erro de predição proveniente da validação cruzada (LOO_{cv}), **Y** é o valor da potência observada, **Y_{média}** é o valor médio da potência e **Y_{pred}** é o valor da potência predita.

Em seguida, a análise de PLS é repetida sem a validação cruzada, usando o número ótimo de PC, gerando o modelo final de CoMFA e os respectivos valores de r^2 convencional, erro padrão da estimativa (SEE), valor de Fischer (F) e as percentagens de contribuição das interações estérica e eletrostática. Como recomendado por Kybinyi e Abraham (1993), o número ótimo de PC deve corresponder ao primeiro mínimo do erro padrão da validação cruzada (SE_{cv}).

Para evitar superajuste (*overfitting*) do modelo, não mais do que N/5 PC são extraídos, onde N é o número de compostos do conjunto de treinamento (Melville & Hirst, 2004).

Qualquer coluna com variação de energia menor do que 2,0 kcal.mol⁻¹ foi excluída para reduzir o tempo de cálculo sem afetar negativamente a qualidade dos modelos. O modelo de CoMFA é representado como um mapa de contorno colorido ao redor de cada molécula, descrevendo os campos estéricos e eletrostáticos que contribuem significativamente para aquele modelo.

3.2.2.8. Validação Externa dos Modelos de CoMFA

Na validação externa dos modelos de CoMFA gerados, a capacidade preditiva total é avaliada em termos de r^2_{pred} , calculado de acordo com a Equação 7, usando os compostos do conjunto de teste.

$$r^2_{\text{pred}} = \frac{\text{SE} - \text{PRESS}}{\text{SE}} \quad \text{Equação 7}$$

Na Equação 7, **SE** é a soma dos desvios quadrados entre o valor de atividade do conjunto de teste e o valor de atividade média do conjunto de treinamento e **PRESS** é a soma dos quadrados do erro de predição proveniente da validação cruzada.

A incerteza da predição é definida pela Equação 8, onde **k** é o número de termos (variáveis) incluídos no modelo e **n** é o número de compostos usado no estudo.

$$\text{SPRESS} = \left[\frac{\text{PRESS}}{(n - k - 1)} \right]^{1/2} \quad \text{Equação 8}$$

3.2.2.9. Seleção do Melhor Modelo de CoMFA

Considerando as diversas equações de CoMFA obtidas a partir da variação dos alinhamentos, cargas atômicas parciais, tipos de átomo de prova e cortes de energia, foram selecionadas as equações de QSAR para análise qualitativa aquelas com os maiores valores de q^2 e de r^2 , os menores erros (SE_{cv} e SEE) e com menor número de *outliers* (Gáudio & Zandonade, 2001; Kubinyi, 1993).

3.2.2.10. Seleção dos *Outliers*

Em estudos de QSAR, alguns autores na literatura consideram como compostos *outliers* aqueles com valor residual modular maior do que uma unidade logarítmica (Sobhia, 2005; Brito, 2004; Kim, 2004), ao passo que outros consideram como *outliers* os compostos com valor residual duas vezes superior ao valor do desvio padrão da estimativa (*Standard Error of Estimate*, SEE) (Albuquerque *et al.*, 2007; Hopfinger, 1997; Kubinyi, 1993). Neste trabalho foi utilizada a última abordagem na detecção dos *outliers*.

3.2.3. PARTE B - QSAR-3D DEPENDENTE DO RECEPTOR

3.2.3.1. Construção e Otimização dos Complexos Ligantes-Enzima

Na ausência de um composto da classe dos DABOs co-cristalizado com a RT do HIV-1, foi usado como complexo de referência na etapa de construção dos complexos ligante-enzima no estudo de QSAR-3D dependente do receptor, a mesma estrutura do PDB do estudo de CoMFA (código 1RT1), contendo o MKC-442 (Figura 19) ligado à enzima RT nativa (Hopkins *et al.*, 1996). A estrutura 3D deste complexo, obtida por cristalografia de raios-X, foi empregada em diversos estudos de modelagem molecular, tais como ancoramento molecular, dinâmica molecular e QSAR-3D (Nervall *et al.*, 2008; Su *et al.*, 2007; Ragno *et al.*, 2005; Ragno *et al.*, 2004; Gaudio & Montanari, 2002; Rizzo *et al.*, 2002; Mao *et al.*, 2000; Rizzo *et al.*, 2000; Szczech *et al.*, 2000).

Na construção dos complexos ligantes-enzima, as estruturas otimizadas de cada um dos DABOs foram inseridas no sítio dos NNRTIs por sobreposição por

RMS (*Root Mean Square*) com a estrutura do inibidor MKC-442 do complexo de referência, usando o programa *HyperChem v.7.5* (HyperCube, Inc.).

Após a exclusão da estrutura do MKC-442 (sobreposta ao ligante), das moléculas de água presentes na estrutura original e da adição dos átomos de hidrogênio, cada um dos complexos ligante-enzima foi otimizado. Os aminoácidos básicos (Lys e Arg) e ácidos (Asp e Glu) foram ionizados.

No primeiro passo otimizou-se apenas o ligante, depois apenas a enzima e por fim todo o complexo. Estas otimizações foram feitas por 1000 passos ou até norma de gradiente menor do que 0,01 kcal/mol.Å, utilizando o algoritmo *Steepest Descent*, e em seguida, por mais 1000 passos ou até norma de gradiente menor do que 0,01 kcal/mol.Å, utilizando o algoritmo *Conjugate Gradient*. Estes cálculos foram feitos no vácuo e sem qualquer restrição geométrica, com o objetivo de minimizar os possíveis contatos de van der Waals desfavoráveis, empregando o campo de força Tripos no programa *Sybyl v.7.2* (Tripos, Inc.).

3.2.3.2. Simulação por Dinâmica Molecular dos Complexos Ligante-Enzima

Antes da dinâmica molecular (DM) as estruturas dos complexos foram submetidas a uma nova etapa de otimização usando o campo de força *Gromos87* (van Gunsteren & Berendsen, 1987), disponível no programa *GROMACS v.3.2* (Lindahl *et al.*, 2001), empregando a mesma metodologia descrita anteriormente para a otimização dos complexos no programa *Sybyl*.

Desta forma, após a etapa preliminar de otimização dos complexos no programa *Sybyl*, as topologias moleculares dos ligantes, necessárias para a etapa de DM dos complexos, foram construídas no servidor *PRODRG* (Schuettelkopf &

van Aalten, 2004), onde as cargas atômicas parciais, calculadas neste programa, foram substituídas pelas cargas calculadas pelo método semi-empírico AM1 no programa *Spartan'06* (Wavefunction, Inc.).

A etapa de simulação por DM foi realizada empregando o campo de força *Gromos87* (van Gunsteren & Berendsen, 1987), disponível no programa *GROMACS* v.3.2 (Lindahl *et al.*, 2001). A razão da escolha deste campo de força é a disponibilidade de se construir as topologias dos ligantes a partir do servidor “The Dundee PRODRG Server” (Schuettelkopf & van Aalten, 2004), o que facilita o trabalho quando se está trabalhando com um grande número de compostos. O *PRODRG* fornece de modo rápido e automático, as topologias e as coordenadas de moléculas pequenas. Testes de refinamento cristalográfico mostram que as topologias geradas são de igual ou melhor qualidade do que as obtidas por outros meios (Schuettelkopf & van Aalten, 2004).

Finalmente, os complexos otimizados foram submetidos à simulação por DM, em condições constantes de temperatura (310 K) e pressão (1 atm), utilizado o método PME (*Particle-Mesh Ewald*) (Darden *et al.*, 1992), seguindo as velocidades iniciais da distribuição de Maxwell.

Inicialmente, as simulações por DM foram executadas até 1000 ps (1 ns), entretanto, como as energias de interação a partir de 100 ps tornaram-se praticamente constantes, esse foi o tempo padrão usado para a coleta dos valores de energia, que correspondeu aos últimos 10 ps.

Para as energias eletrostáticas de longo alcance e para as interações não-ligadas, foi usado um valor de corte (*cutoff*) de 9 Å. O algoritmo SHAKE foi utilizado

para manter os comprimentos das ligações fixos (Ryckaert *et al.*, 1977). O tempo de integração foi de 1 fs e as energias foram tabeladas em intervalos de 10 ps.

3.2.3.3. Definição do Raio de Corte e Obtenção das Energias de Interação Ligante-Enzima (Variáveis Independentes)

As variáveis independentes neste estudo de QSAR-3D dependente do receptor correspondem às energias de interação estéricas e eletrostáticas, calculadas entre cada um dos ligantes e cada um dos resíduos de aminoácidos da enzima compreendidos dentro de um raio de corte de 10 Å em torno do ligante (Figura 19), que compreendeu 53 resíduos de aminoácidos.

Este recorte foi feito considerando que as interações ligante-receptor que mais significativamente contribuem para a variação da resposta inibitória ocorrem com resíduos específicos da enzima, próximos ao sítio de ligação (Kulkarni e Kulkarni, 1999; Todd e Freire, 1999). Este tipo de recorte tem analogia ao método de “pruning” (poda) desenvolvido por Tokarski e Hopfinger (1997) em estudos de QSAR-3D, onde os termos de energia de ligação do complexo ligante-receptor são calculados pelo campo de força de energia livre (*Free Energy Force Field*, FEFF) em modelos recortados do complexo, e são usados como variáveis independentes na construção de equações de FEFF-QSAR-3D (Romeiro *et al.*, 2006).

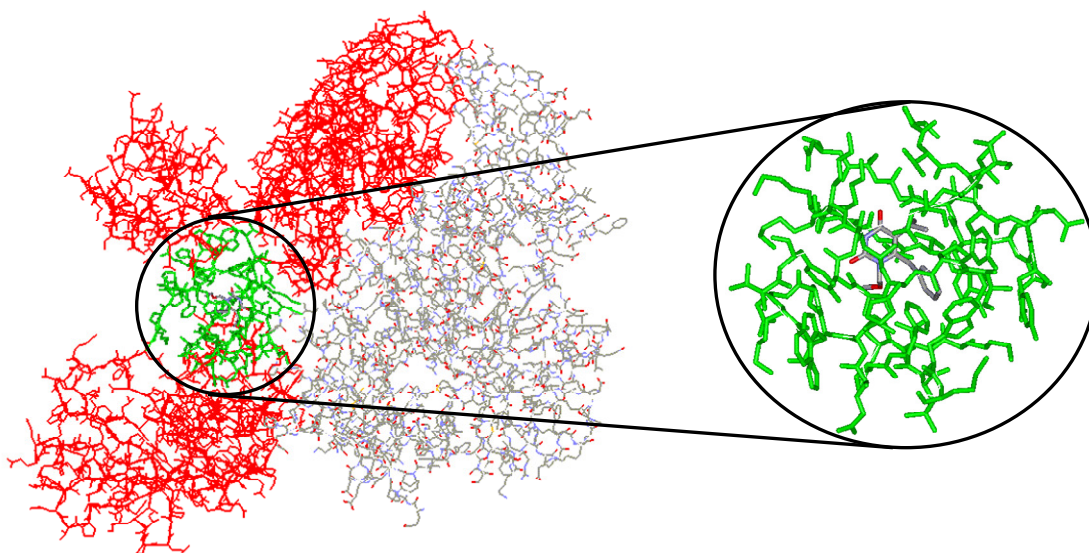


Figura 19. Representação esquemática do recorte do complexo do inibidor MKC-442 com a enzima RT do HIV-1 (código no PDB: 1RT1). À esquerda é mostrada a estrutura da RT com as subunidades p66 (colorida em vermelho) e p51 (colorida por elemento) e o círculo delimita a região do raio de corte de 10 Å a partir do inibidor. No detalhe à direita encontram-se os resíduos da enzima compreendidos no recorte (coloridos em verde) e o inibidor (colorido por elemento).

Os cálculos foram realizados no programa *GROMACS* v.3.2 (Lindahl *et al.*, 2001), que emprega os potenciais de Lennard-Jones e de Coulomb para calcular as energias de interação estéricas e eletrostáticas, respectivamente (Berendsen *et al.*, 1995).

3.2.3.4. Definição dos Bancos de Dados de Variáveis Independentes

Objetivando avaliar a influência do banco de dados de variáveis independentes (descritores) sobre a capacidade preditiva dos modelos (equações) a serem gerados, foram testados quatro bancos de dados (BDs), variando o número, a combinação e o pré-tratamento dos descritores, como descrito a seguir.

a) O primeiro banco de dados, **BD-I**, corresponde ao banco de dados original, onde os descritores são as energias de interação estéricas e eletrostáticas calculadas individualmente por resíduo, usando os potenciais de Lennard-Jones (**LJ**) e de Coulomb (**C**), respectivamente. Considerando que o recorte do complexo ligante-enzima contém 53 resíduos de aminoácidos, o **BD-I** contém, portanto, um total de 106 colunas de descritores (53 de **LJ** e 53 de **C**).

b) No segundo banco de dados, **BD-II**, os descritores correspondem ao somatório das energias de Lennard-Jones e Coulomb por resíduo, perfazendo um total de 53 colunas de descritores (53 de **LJ+C**), o que reduziu o número de descritores pela metade em relação ao BD original.

c) O terceiro banco de dados, **BD-III**, corresponde a combinação dos dois bancos de dados anteriores, perfazendo um total de 159 colunas de descritores (53 de **LJ**, 53 de **C** e 53 de **LJ+C**).

d) O quarto banco de dados, **BD-IV**, corresponde ao **BD-I** (106 descritores) após pré-tratamento, onde foram excluídas as colunas de energia (descritores) com variância menor do que 0,0001, as quais, provavelmente, não contribuem para a explicação da variação na resposta biológica. Com este pré-tratamento, foram excluídos 11 descritores, perfazendo um total de 95 colunas.

3.2.3.5. Obtenção das Equações de QSAR-3D Dependentes do Receptor

Os quatro bancos de dados de variáveis independentes (**BD-I**, **BD-II**, **BD-III** e **BD-IV**) foram submetidos individualmente ao programa *Wolf* (Rogers & Hopfinger, 1994), junto com os valores de atividade biológica (pIC_{50}) para a geração das equações de QSAR-3D. Neste programa, o conjunto de variáveis independentes,

energias de interação, foi relacionado ao conjunto de variáveis dependentes, pIC_{50} , por meio da técnica GFA-PLS (Rogers & Hopfinger, 1994), descrita anteriormente na Introdução (Dunn & Rogers, 1996).

A primeira etapa no programa *Wolf* foi a geração de uma população de partida de 100 equações, cada uma contendo inicialmente quatro descritores selecionados aleatoriamente a partir do **BD**. Foram testadas diversas combinações de opções dentro da técnica de GFA-PLS, estabelecendo-se o uso de 100% de probabilidade de mutação após cada operação de *crossover* e efetuando-se 10.000 e 50.000 dessas operações. Os coeficientes das equações foram calculados por PLS, usando-se 3, 4, 5 e 6 componentes principais. O algoritmo que ajusta o número de variáveis independentes nos modelos, fator de suavização (*smoothing-factor*), foi ajustado de 0,2 até 0,6 (com variação de 0,1). A combinação destas opções foi realizada a fim de se obter modelos contendo de cinco a doze variáveis independentes (termos) e resultou em 40 conjuntos de opções, que foram testados para cada um dos quatro bancos de dados (**BD-I**, **BD-II**, **BD-III** e **BD-IV**) dando origem a um total de 200 equações de QSAR-3D a serem analisadas.

3.2.3.6. Validação Interna dos Modelos de QSAR-3D Dependentes do Receptor

Os dez melhores modelos de cada corrida de GFA-PLS foram classificados de acordo com os valores de falta de ajuste de Friedman (*lack-of-fit score*, LOF) (Rogers & Hopfinger, 1994), e foram submetidos à técnica de validação cruzada LOO_{CV} (conforme está descrito no item correspondente na Parte A da Metodologia de CoMFA), disponível no programa *Wolf* v.6.2. LOF é a medida dos mínimos quadrados penalizada, *i.e.*, quando duas equações têm o mesmo erro de mínimos

quadrados, aquela que possuir o menor número de termos (variáveis independentes) terá o menor valor LOF e será, portanto, a melhor equação (Rogers & Hopfinger, 1994).

Para evitar o superajuste (*overfitting*) do modelo, admite-se que o número máximo de termos deve estar na razão de cinco compostos do banco de dados para cada termo da equação (Gáudio & Zandonade, 2001; Kubinyi, 1993). Assim, o número máximo de termos foi obtido dividindo-se o número total de compostos do conjunto de treinamento (N=59) por cinco, que resulta no valor de 11,8 termos. Portanto, os modelos com doze ou mais variáveis não foram considerados para análise posterior.

3.2.3.7. Validação Externa dos Modelos de QSAR-3D Dependentes do Receptor

A validação externa dos modelos de QSAR-3D dependente do receptor foi feita empregando-se os compostos do conjunto de teste, conforme está descrito no item correspondente na Parte A da Metodologia de CoMFA.

3.2.3.8. Seleção dos Melhores Modelos de QSAR-3D Dependente do Receptor

As diversas equações de QSAR-3D dependente do receptor obtidas após o processo de validação cruzada foram ordenadas pelo número crescente de termos (que variou de 5 a 12 variáveis independentes) contidos em cada equação, considerando-se para análise qualitativa aquelas com os maiores valores de q^2 , e de r^2 , os menores erros (SE_{cv} e SEE) e com menor número de *outliers* (Kubinyi, 1997b).

Para comparar equações com número de termos diferentes, os valores de q^2 foram transformados em q^2 ajustado (Livingstone, 1995), de acordo com a Equação 9.

$$q_{ajustado}^2 = 1 - (1 - q^2) \frac{n-1}{n-p} \quad \text{Equação 9}$$

Nesta Equação, q^2 representa o valor do coeficiente de correlação quadrático da validação cruzada, n é o número de compostos do conjunto de treinamento e p é o número de variáveis (termos) contidas no modelo.

3.2.3.9. Seleção dos *Outliers*

A seleção dos *outliers* dos modelos de QSAR-3D dependente do receptor foi feita conforme está descrito no item correspondente na Parte A da Metodologia de CoMFA.

3.2.3.10. Análise da Matriz de Correlação Cruzada dos Resíduos

Para os melhores modelos selecionados, foram analisados os coeficientes de correlação simples (r) entre os resíduos dos modelos obtidos na respectiva matriz de correlação cruzada. Os valores residuais correspondem a diferença entre os valores de atividade biológica observada ou experimental (pIC_{50obs}) e os valores de atividade biológica calculada (pIC_{50pred}).

Segundo Rogers (1996), espera-se que modelos equivalentes tenham distribuições idênticas dos resíduos, e que modelos distintos apresentem padrões

dos resíduos não correlacionados. Portanto, este tipo de análise é uma ferramenta valiosa para a determinação de um subconjunto de modelos distintos em um conjunto de bons modelos obtidos em análises de GFA-PLS, eliminando-se modelos com o mesmo tipo de informação estrutura-atividade (Hopfinger *et al.*, 1997).

Em trabalhos de QSAR, é importante a exclusão de modelos intercorrelacionados (Sodero, 2007; Pita, 2006; Cunha, 2006; Romeiro, 2002; Albuquerque, 1997).

3.2.3.11. Análise da Matriz de Correlação Cruzada dos Descritores

Analizou-se também para os melhores modelos selecionados a matriz de correlação cruzada das variáveis independentes (descritores), objetivando determinar se duas ou mais variáveis altamente correlacionadas aparecem simultaneamente em um mesmo modelo (Sodero, 2007; Pita, 2006; Cunha, 2006; Romeiro, 2002; Albuquerque, 1997). Com esta abordagem, é possível descartar modelos que apresentem informação redundante (Livingstone, 1995).

A exclusão de modelos com variáveis independentes intercorrelacionadas também é importante em trabalhos de QSAR (Sodero, 2007; Pita, 2006; Cunha, 2006; Romeiro, 2002; Albuquerque, 1997).

4. RESULTADOS E DISCUSSÃO

A apresentação e a discussão dos resultados das duas abordagens de QSAR-3D deste trabalho estão divididas em duas partes, Parte A (QSAR-3D Independente do Receptor, CoMFA) e Parte B (QSAR-3D Dependente do Receptor).

É importante lembrar que as duas abordagens têm em comum o uso do mesmo banco de dados de compostos (Tabela 5), dividido nos mesmos conjuntos de treinamento (**1-59**) e de teste (**60-74**), onde os respectivos valores de atividade biológica (plC_{50}) estão regularmente distribuídos ao longo da faixa de atividade, que abrange cerca de quatro unidades logarítmicas (Figura 18).

Em ambos os estudos foi usada como referência a estrutura de cristalografia de raios-X do composto MKC-442, disponível no PDB sob o código 1RT1 (Hopkins *et al.*, 1996). Nos estudos de QSAR-3D dependente do receptor foi utilizada também a estrutura da enzima RT deste mesmo complexo.

A variável dependente (Y) em ambos os estudos corresponde aos valores de plC_{50} (M), e as variáveis independentes (X) correspondem ao conjunto de valores de energia de interação estérica e eletrostática calculados de modo distinto de acordo com as metodologias de QSAR-3D independente (CoMFA) e dependente do receptor.

4.1. PARTE A – QSAR-3D INDEPENDENTE DO RECEPTOR (CoMFA)

4.1.1. Análise da Sobreposição dos DABOs em Função dos Alinhamentos

Os Alinhamentos 1, 2 e 3 testados (Figura 18, Metodologia) mostraram uma boa sobreposição entre as diferentes regiões das moléculas dos DABOs, conforme ilustra a Figura 20 para o Alinhamento 1, que resultou no melhor modelo de CoMFA. Observa-se que os anéis 4-oxopirimidina, comum a todos os compostos, ficaram perfeitamente superpostos, com exceção do substituinte **X**, que pode ser hidrogênio ou metila, enquanto houve uma maior variabilidade conformacional, já esperada, nas posições 2 (substituinte **-W-Y**) e 6 (substituinte **-CH₂-Ar**) do anel 4-oxopirimidina, referentes aos diferentes grupos alquila (**Y**) e arila (**Ar**), respectivamente, que podem ocupar estas posições.

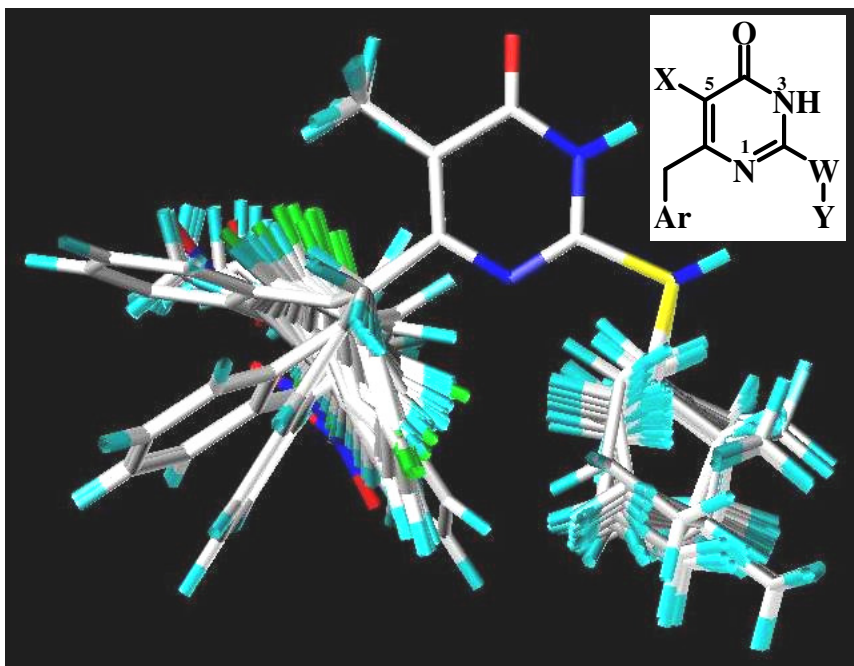


Figura 20. Sobreposição entre os 59 compostos do conjunto de treinamento da série dos S- e NH-DABOs de acordo com o Alinhamento 1.

4.1.2. Seleção do Melhor Alinhamento e Carga Atômica Parcial

A Tabela 6 mostra os resultados estatísticos dos modelos de CoMFA obtidos, mantendo-se fixos o átomo de prova (C_{sp}^3 , carga +1), o valor de corte de energia das interações estéricas e eletrostáticas (30 kcal.mol^{-1}) e o espaçamento da grade ($2,0 \text{ \AA}$), variando, porém, os Alinhamentos (1, 2 e 3) e as cargas atômicas parciais (q) dos ligantes, que foram ajustadas ao RESP calculado em três níveis de teoria: DFT (B3LYP/6-31G*), *ab initio* (HF/6-31G*) e métodos semi-empíricos AM1 e PM3.

Comparando-se os valores de r^2 , observa-se que há uma pequena diferença na % da variabilidade explicada ($r^2 \times 100$) pelos modelos de CoMFA gerados, considerando os diferentes métodos de cálculo de cargas atômicas parciais em função dos três Alinhamentos testados. Os modelos são capazes de explicar cerca de 91 a 94% da variabilidade da resposta biológica, com cinco ou seis componentes, com exceção do modelo gerado no Alinhamento 2 com cargas DFT (B3LYP/6-31G*), que explica cerca de 87% da variabilidade, com quatro componentes. Considerando apenas os modelos gerados com o mesmo número de componentes principais (PC=6), em valores absolutos, o método *ab initio* (HF/6-31G*) apresenta melhor performance (maior valor de r^2) nos três Alinhamentos, seguido dos métodos semi-empíricos (AM1 e PM3) e DFT nos Alinhamentos 1 e 3.

No entanto, comparando-se os valores de q^2 (r^2 da validação cruzada), observa-se que há uma maior diferença na capacidade preditiva dos modelos de CoMFA gerados em função dos diferentes tipos de cargas parciais e alinhamentos testados. Neste caso, os métodos semi-empíricos (AM1 e PM3) apresentam melhor performance (maior valor de q^2) do que os métodos *ab initio* e DFT, nos três Alinhamentos, onde o método PM3 se destaca, em termos de maior valor absoluto de q^2 .

Confrontando os três Alinhamentos, de um modo geral, os modelos provenientes do Alinhamento 1 são um pouco mais preditivos do que os do Alinhamento 3, enquanto que os modelos do Alinhamento 2 têm menor capacidade preditiva (Tabela 6). Este comportamento pode ser explicado porque o Alinhamento 1 permite uma maior variabilidade na disposição relativa dos substituintes das posições 2 e 6 do anel 4-oxopirimidina, visto que apenas o plano deste anel é usado na sobreposição, enquanto que o Alinhamento 2 é o que permite menor variabilidade, pois os planos dos dois sistemas aromáticos são usados na sobreposição.

Comparando-se os valores percentuais das contribuições estérica (**S**) e eletrostática (**E**) é interessante notar que, independente do alinhamento testado, os modelos gerados com cargas derivadas pelos métodos semi-empíricos (AM1 e PM3) e *ab initio* (HF/6-31G*) apresentam uma maior contribuição do campo eletrostático (**E** variando de cerca de 56 a 60%), enquanto nos modelos referentes ao método DFT, as contribuições **S** e **E** são praticamente equivalentes (cerca de 50%) nos três alinhamentos.

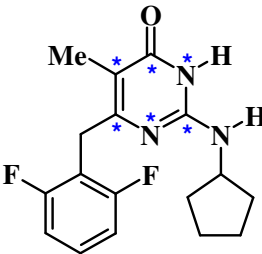
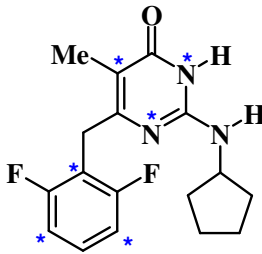
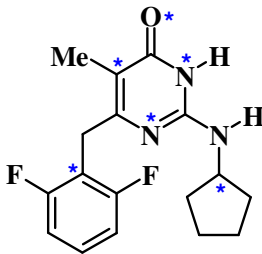
Entretanto, qualitativamente, os mapas de contorno gerados com estes modelos, independente do tipo de carga e alinhamento, são similares (dados não mostrados para todos os modelos), com contribuições positivas e negativas distribuídas em regiões semelhantes dos campos estéricos e eletrostáticos gerados ao redor das moléculas.

Considerando estes resultados em conjunto, o melhor modelo corresponde aquele obtido a partir do Alinhamento 1 e usando cargas derivadas do método semi-empírico PM3, onde foram usadas as opções padrões para: corte de energia (30 kcal.mol⁻¹ para ambos os campos estéricos e eletrostáticos), tipo de átomo de prova

(C_{sp}^3 , carga +1) e espaçamento da grade ($D=2,0\text{\AA}$) (Tabela 6). O teste F, realizado nos valores PRESS, resultou em 6 componentes principais como valor ótimo, com $SE_{cross}=0,475$ e $q^2=0,691$. Como esperado, as análises sem validação cruzada forneceram um melhor ajuste dos dados ($SEE=0,226$ e $r^2=0,930$).

O percentual de contribuição dos campos estéricos e eletrostáticos neste modelo corresponde a 43,2% e 56,8%, respectivamente, indicando uma influência um pouco maior do campo eletrostático sobre a variação da resposta biológica na relação estrutura-atividade. Isso é interessante, pois o sítio de interação não-nucleosídico (NNBS) tem características predominantemente hidrofóbicas, onde as interações estéricas devem ser predominantes. Provavelmente, esse resultado reflete uma maior influência dos campos eletrostáticos devido à semelhança entre os DABOs em termos de volumes moleculares.

Tabela 6. Resultados estatísticos dos modelos de CoMFA obtidos testando três alinhamentos (1, 2 e 3) e quatro tipos de cargas atômicas parciais (DFT, HF, AM1 e PM3). Mantendo-se fixos o valor de corte de energia (30 kcal.mol⁻¹ para os campos estérico e eletrostático), átomo de prova (C_{sp}³, carga +1) e espaçamento da grade (2,0 Å).

<div style="display: flex; justify-content: space-around; align-items: flex-end;"> <div style="text-align: center;">  <p>Alinhamento 1</p> </div> <div style="text-align: center;">  <p>Alinhamento 2</p> </div> <div style="text-align: center;">  <p>Alinhamento 3</p> </div> </div>								
q ^a	q ^{2b}	SE _{cv} ^c	PC ^d	r ^{2e}	SEE ^f	Valor F ^g	S (%) ^h	E (%) ⁱ
Alinhamento 1								
DFT	0,611	0,533	6	0,927	0,230	110,437	50,2	49,8
HF	0,617	0,528	6	0,942	0,205	141,498	41,3	58,7
AM1	0,674	0,488	6	0,931	0,225	116,223	42,7	57,3
PM3	0,691	0,475	6	0,930	0,226	115,544	43,2	56,8
Alinhamento 2								
DFT	0,575	0,546	4	0,875	0,296	94,897	50,5	49,5
HF	0,614	0,530	6	0,940	0,210	135,118	42,5	57,5
AM1	0,649	0,501	5	0,911	0,252	109,094	40,2	59,8
PM3	0,681	0,478	5	0,920	0,239	121,982	41,2	58,8
Alinhamento 3								
DFT	0,607	0,535	6	0,923	0,238	103,164	50,0	50,0
HF	0,613	0,531	6	0,940	0,209	136,485	43,8	56,2
AM1	0,659	0,499	6	0,927	0,230	110,485	43,5	56,5
PM3	0,683	0,481	6	0,929	0,228	112,782	43,6	56,4

a) q = cargas atômicas parciais ajustadas ao RESP calculado em três níveis de teoria: DFT (B3LYP/6-31G*), *ab initio* (HF/6-31G*) e métodos semi-empíricos AM1 e PM3. b) q² = r² da validação cruzada. c) SE_{cv} = desvio padrão da validação cruzada. d) PC = número ótimo de componentes principais. e) r² = coeficiente de correlação linear quadrático. f) SSE = desvio padrão de estimativa. g) F = valor de Fisher. h) S = contribuição estérica. i) E = contribuição eletrostática.

4.1.3. Seleção do Tipo de Átomo de Prova

Em seguida, após seleção do melhor alinhamento, Alinhamento 1, e do melhor método de obtenção das cargas atômicas parciais dos ligantes, método semi-empírico PM3, novos modelos foram obtidos mantendo-se estas opções constantes, testando o efeito de dois tipos de átomos de prova adicionais sobre o valor de q^2 resultante considerando os valores padrões de corte de energia e de espaçamento da grade. A Tabela 7 mostra os resultados estatísticos dos modelos de CoMFA gerados com os três átomos de prova: C_{sp}^3 , carga +1 (padrão); O_{sp}^3 , carga -1 e H, carga +1.

Embora os modelos obtidos sejam semelhantes, o maior valor absoluto de q^2 (0,695) foi obtido com o uso do átomo de prova O_{sp}^3 . Entretanto, houve uma grande diminuição nos valores de r^2 e de F (Fisher) deste modelo, quando comparado ao modelo gerado com o átomo de prova padrão, C_{sp}^3 , indicando menores níveis de consistência interna e significância para o modelo. Além do mais, o átomo de carbono reflete melhor os tipos de interações que ocorrem no NNBS do que o átomo de oxigênio, já que se sabe que este sítio é predominantemente hidrofóbico. Desse modo, o átomo de prova C_{sp}^3 com carga +1 permanece a melhor escolha para esta análise. Curiosamente, a contribuição estérica (**S**) no modelo obtido com o átomo de prova H teve um valor nulo.

Tabela 7. Resultados estatísticos dos modelos de CoMFA obtidos (Alinhamento 1 e cargas atômicas parciais PM3) testando três átomos de prova (C_{sp}^3 , carga +1; O_{sp}^3 , carga -1 e H, carga +1). Mantiveram-se fixos: valor de corte de energia (30 kcal.mol⁻¹ para os campos estérico e eletrostático) e espaçamento da grade (2,0 Å).

PA ^a	q ^{2b}	SE _{cv} ^c	PC ^d	r ^{2e}	SEE ^f	Valor F ^g	S (%) ^h	E (%) ⁱ
C_{sp}^3 (+1)	0,691	0,475	6	0,930	0,226	115,544	43,2	56,8
O_{sp}^3 (-1)	0,695	0,462	4	0,842	0,333	71,939	43,7	56,3
H (+1)	0,678	0,476	4	0,862	0,312	84,058	0	100

a) PA = átomo de prova (carga formal). b) $q^2 = r^2$ da validação cruzada. c) SE_{cv} = desvio padrão da validação cruzada. d) PC = número ótimo de componentes principais. e) r^2 = coeficiente de correlação linear quadrático. f) SSE = desvio padrão da estimativa. g) F = valor de Fisher. h) S = contribuição estérica. i) E = contribuição eletrostática.

4.1.4. Seleção do Valor de Corte de Energia

Em seguida, mantendo-se constantes o Alinhamento 1, o método PM3 para obtenção das cargas parciais, o valor padrão de espaçamento da grade (2,0 Å) e o átomo de prova padrão (C_{sp}^3), novos modelos foram obtidos testando o efeito de dois valores adicionais de corte de energia sobre o valor de q^2 resultante. A Tabela 8 mostra os resultados estatísticos dos modelos de CoMFA gerados com os três valores de corte de energia testados: 30 (padrão), 20 e 10 kcal.mol⁻¹.

Embora os modelos obtidos sejam semelhantes (Tabela 8), o maior valor absoluto de q^2 (0,726) foi obtido com o uso do valor de corte de energia de 10 kcal.mol⁻¹. Entretanto, de modo similar aos resultados anteriores, houve uma pequena diminuição nos valores de r^2 e de F deste modelo quando comparado ao modelo gerado com o valor de corte de energia padrão (30 kcal.mol⁻¹), indicando menores níveis de consistência interna e significância. Desse modo, o melhor modelo selecionado continua sendo aquele obtido segundo o Alinhamento 1, cargas

atômicas parciais dos ligantes derivadas por PM3, C_{sp}^3 (carga +1) como átomo de prova e corte de energia de 30 kcal.mol⁻¹.

Tabela 8. Resultados estatísticos dos modelos de CoMFA (Alinhamento 1 e cargas atômicas parciais PM3) obtidos testando três valores de corte de energia (30, 20 e 10 kcal.mol⁻¹) para os campos estérico e eletrostático. Mantiveram-se fixos o átomo de prova (C_{sp}^3 , carga +1) e o espaçamento da grade (2,0 Å).

Corte^a	q^{2b}	SE_{cv}^c	PC^d	r^{2e}	SEE^f	Valor F^g	S (%)^h	E (%)ⁱ
30	0,691	0,475	6	0,930	0,226	115,544	43,2	56,8
20	0,685	0,478	6	0,921	0,240	100,727	46,0	54,0
10	0,726	0,447	6	0,924	0,235	105,724	47,2	52,8

a) Corte = valor de corte de energia para ambos os campos estérico e eletrostático (kcal.mol⁻¹). b) q² = r² da validação cruzada. c) SE_{cv} = desvio padrão da validação cruzada. d) PC = número ótimo de componentes principais. e) r² = coeficiente de correlação linear quadrático. f) SSE = desvio padrão da estimativa. g) F = valor de Fisher. h) S = contribuição estérica. i) E = contribuição eletrostática.

4.1.5. Análise dos Resíduos dos Compostos dos Conjuntos de Treinamento e de Teste

A Tabela 9 mostra os valores de pIC_{50} observados (experimentais) e os preditos pelo melhor modelo de CoMFA para os compostos dos conjuntos de treinamento (1-59) e de teste (60-74), e os respectivos valores residuais, que correspondem à diferença entre os valores de pIC_{50} observado e calculado. As Figuras 21.A e 21.B mostram gráficos com os valores de pIC_{50} preditos e observados para os compostos do conjunto de treinamento e de teste, respectivamente.

Considerando os valores de pIC_{50} preditos para os compostos do conjunto de teste, a performance do modelo pode ser considerada excelente ($r^2_{pred} = 0,918$). Oitenta por cento (80%) dos valores de pIC_{50} dos compostos do conjunto de teste foram preditos com resíduos menores que 0,50 (em valores modulares). Predições estatisticamente significativas suportam a validade do modelo gerado na predição das potências de novos compostos.

Tabela 9. Valores de pIC_{50} (M) observados e preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) dos composto dos conjuntos de treinamento (**1-59**) e de teste (**60-74**) do melhor modelo de CoMFA (Alinhamento 1, cargas atômicas parciais PM3, átomo de prova C_{sp^3} e carga +1, valor de corte de 30 kcal.mol⁻¹ e espaçamento da grade de 2,0 Å).

#	pIC_{50Obs}	pIC_{50Pred}	Res.	#	pIC_{50Obs}	pIC_{50Pred}	Res.
1	4,23	4,23	0,00	38	6,10	6,26	-0,16
2	4,31	4,54	-0,23	39	6,10	5,93	0,17
3	4,35	4,23	0,12	40	6,10	6,06	0,04
4	4,59	4,64	-0,05	41	6,10	5,89	0,21
5	4,77	5,02	-0,25	42	6,22	6,07	0,15
6	4,79	4,64	0,15	43	6,22	6,32	-0,10
7	4,83	4,90	-0,07	44	6,22	6,08	0,14
8	4,83	4,58	0,25	45	6,30	6,15	0,15
9	5,02	4,93	0,09	46	6,40	5,92	0,48
10	5,07	5,19	-0,12	47	6,70	6,73	-0,03
11	5,09	5,45	-0,36	48	6,70	6,83	-0,13
12	5,27	5,53	-0,26	49	6,70	6,27	0,43
13	5,31	5,67	-0,36	50	6,92	6,20	0,72
14	5,31	5,15	0,16	51	7,00	7,01	-0,01
15	5,32	5,53	-0,21	52	7,00	7,17	-0,17
16	5,34	5,47	-0,13	53	7,05	7,06	-0,01
17	5,37	5,42	-0,05	54	7,05	6,89	0,16
18	5,42	5,41	0,01	55	7,10	7,05	0,05
19	5,44	5,47	-0,03	56	7,10	7,13	-0,03
20	5,47	5,46	0,01	57	7,15	7,26	-0,11
21	5,49	5,85	-0,36	58	7,30	7,11	0,19
22	5,52	5,55	-0,03	59	7,52	7,44	0,08
23	5,52	5,87	-0,35	60	4,35	4,40	-0,05
24	5,52	5,55	-0,03	61	4,48	4,73	-0,25
25	5,54	5,47	0,07	62	5,27	5,43	-0,16
26	5,55	5,56	-0,01	63	5,47	5,44	0,03
27	5,59	5,49	0,10	64	5,59	5,54	0,05
28	5,60	5,35	0,25	65	5,60	5,42	0,18
29	5,60	5,47	0,13	66	5,62	5,39	0,23
30	5,62	5,69	-0,07	67	5,74	5,65	0,09
31	5,66	5,53	0,13	68	5,92	5,41	0,51
32	5,72	5,39	0,33	69	6,22	5,68	0,54
33	5,80	5,83	-0,03	70	6,40	5,78	0,62
34	5,89	6,25	-0,36	71	6,70	6,78	-0,08
35	5,94	6,19	-0,25	72	7,05	6,95	0,10
36	5,94	6,19	-0,25	73	7,15	7,15	0,00
37	5,96	6,10	-0,14	74	7,30	7,24	0,06

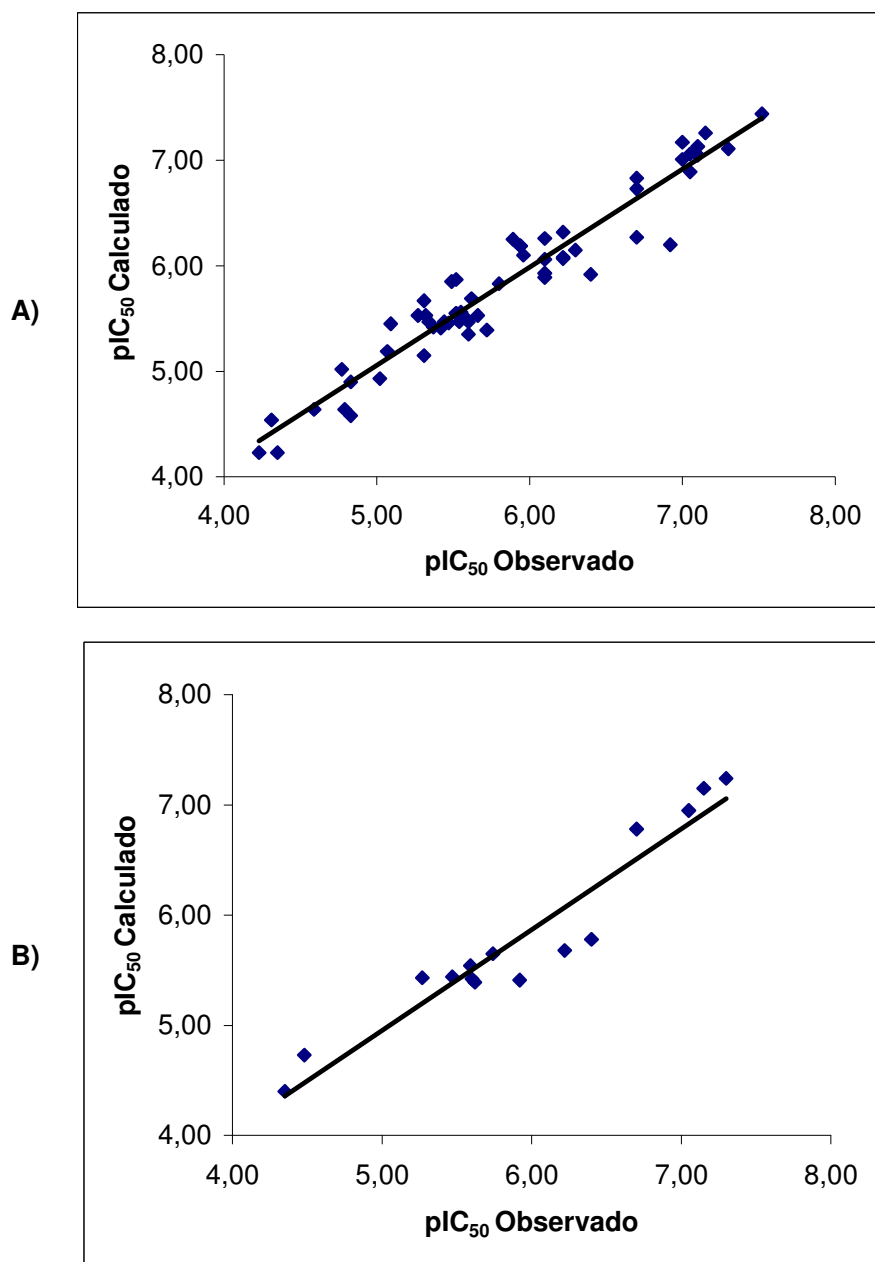


Figura 21. Gráficos dos valores de pIC₅₀ observados *versus* preditos para os conjuntos de (A) treinamento (1-59) e de (B) teste (60-74) do melhor modelo de CoMFA obtido usando Alinhamento 1, cargas PM3 e valores padrões de corte de energia (30 kcal.mol⁻¹), átomo de prova (C_{sp}³, carga +1) e espaçamento da grade (2,0 Å).

4.1.6. Análise dos Mapas de Contorno do Melhor Modelo de CoMFA

Os mapas de contorno dos campos moleculares devem ser considerados como uma representação dos pontos da grade onde as diferenças nos valores dos campos estão fortemente associadas com as diferenças na afinidade de ligação ao receptor. Embora os mapas de contorno não possam ser usados como mapas do receptor, eles permitem interpretações úteis no contexto das interações ligante-receptor.

Para ilustrar o melhor modelo de CoMFA, os mapas de contorno estérico e eletrostático ao redor do derivado mais potente, **59** (resíduo = 0,08), são mostrados nas Figuras 22.A e 22.B, respectivamente.

No mapa de CoMFA estérico (Figura 22.A), as áreas representadas por poliedros verdes são aquelas onde substituintes volumosos aumentam a potência, enquanto as áreas representadas por poliedros amarelos são aquelas onde substituintes volumosos diminuem a potência.

Usando o composto **59** como referência, observa-se campos estéricos ao redor dos substituintes das posições **C2**, **C5** e **C6** do anel 4-oxopirimidina.

Em relação à posição **C2**, o mapa estérico mostra uma região estericamente favorável (verde), relativamente extensa, ao redor do grupo ciclopentila do composto **59**, porém, nesta mesma área, há também regiões estericamente desfavoráveis fragmentadas (amarelo), indicando cuidado na seleção de substituintes volumosos para esta posição.

Em relação à posição **C5**, há uma região estericamente favorável (verde), relativamente grande, próxima ao grupo metila do composto **59**, indicando que substituintes volumosos aumentariam a potência biológica, tornando esta posição

adequada para futuras modificações. Entretanto, o tamanho desse grupo não deve ser muito grande devido ao volume limitado do NNBS na RT.

Finalmente, em relação à posição **C6** existem diversas regiões estericamente desfavoráveis (amarelo), relativamente grandes, ao redor do grupo arila do composto **59**, indicando que grupos arila volumosos nessa posição diminuem a potência. De fato, os compostos que possuem um anel naftila nessa posição, ao invés de um anel fenila, têm os menores valores de atividade biológica do banco de dados, pIC_{50} variam de 4,23 a 4,83 M).

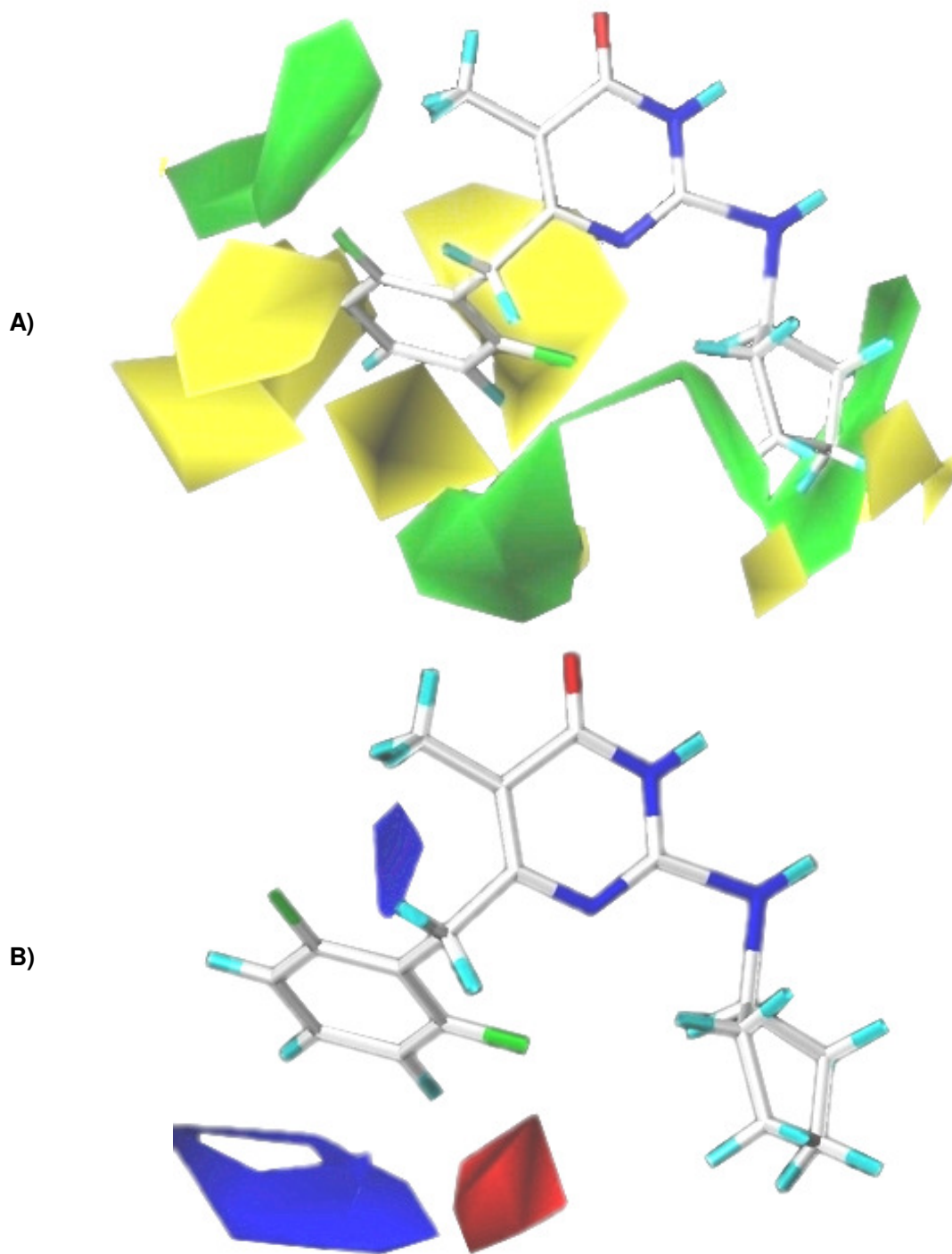


Figura 22. Mapas de contorno do melhor modelo de CoMFA exemplificado para o composto **59** (em modelo bastão e colorido por elemento). (A) O mapa estérico mostra as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a potência. (B) O mapa eletrostático mostra as áreas onde grupos com alta densidade eletrônica ou carga negativa aumentam (vermelho) ou diminuem (azul) a potência.

Ainda considerando a posição **C6** do anel 4-oxopirimidina, a análise dos S- e NH-DABOs ajustados no NNBS, permite inferir que os melhores substituintes nesta posição devem ser realmente anéis aromáticos, devido à possibilidade de interação do tipo “empilhamento” pi-pi (π - π *stacking*) entre esses sistemas aromáticos e os anéis aromáticos dos resíduos **Tyr181** e **Tyr188**.

De fato, o composto mais potente (**59**), ajustado e otimizado no NNBS, mostrou uma distância de 3,75 Å e 3,67 Å entre o anel aromático e os anéis dos resíduos **Tyr181** e **Tyr188**, respectivamente (Figura 23). Esses aminoácidos têm papel-chave no processo inibitório, uma vez que o sítio catalítico da RT é adjacente ao NNBS (sítio alostérico), onde os resíduos **Tyr181** e **Tyr188** estão próximos aos resíduos catalíticos (**Asp110**, **Asp185** e **Asp186**). Assim, quando o inibidor se liga ao sítio alostérico, ocorre uma reorganização do sítio catalítico, impedindo a função catalítica (Castro *et al.*, 2006; Flexner, 2006; De Clercq, 2005).

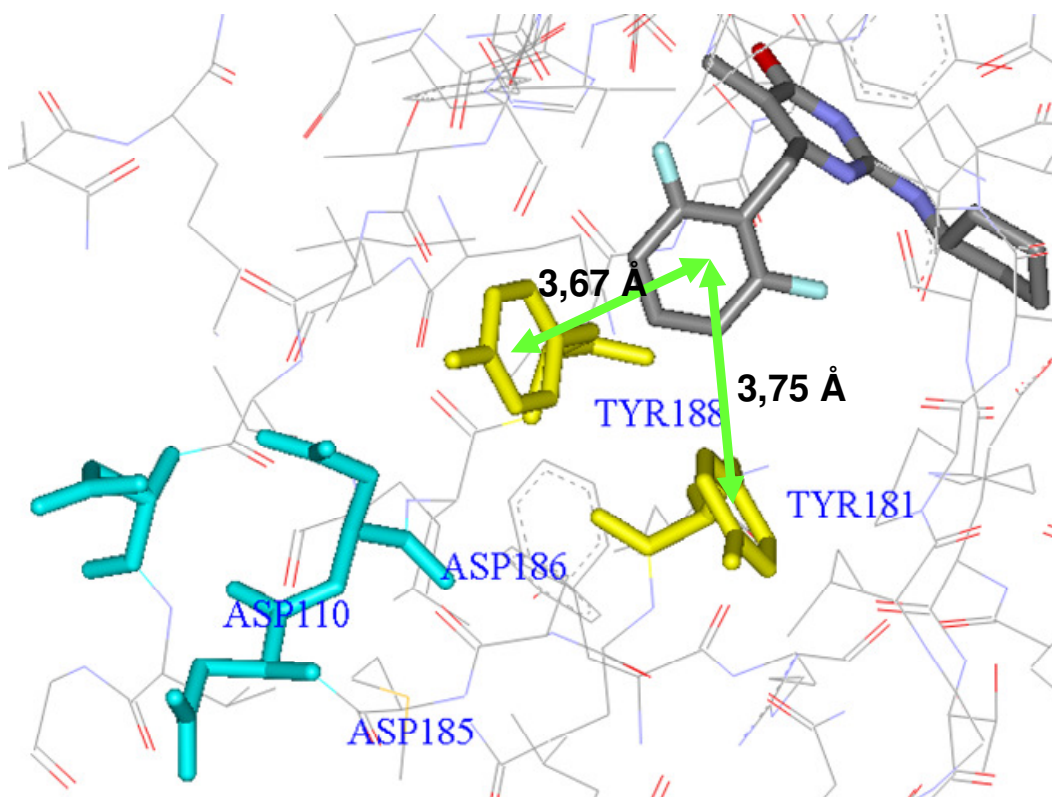


Figura 23. Modo de ligação do NH-DABO **59** (em modelo bastão e colorido por elemento) no sítio não-nucleosídeo (NNBS) da RT do HIV-1. Todos os resíduos estão representados em linhas e coloridos por elemento, exceto **Tyr181** e **Tyr188** (próximo ao grupo benzila de **59**), que está representado em modelo bastão e colorido em amarelo, e **Asp110**, **Asp185** e **Asp186** (resíduos catalíticos), que estão representados em modelo bastão e coloridos em azul. Os átomos de hidrogênio foram omitidos para melhor visualização.

No mapa de CoMFA eletrostático (Figura 22.B), as áreas representadas por poliedros vermelhos são aquelas onde substituintes com alta densidade eletrônica aumentam a potência, enquanto que as áreas representadas por poliedros azuis são aquelas onde substituintes com alta densidade eletrônica diminuem a potência.

Usando o composto **59** como referência, observam-se campos eletrostáticos ao redor do anel 2,4-difluór-fenila ligado à posição **C6** do anel 4-oxopirimidina. Este mapa mostra, próximo à posição *meta* deste anel, uma região favorável (vermelho) a

substituintes com alta densidade eletrônica e, próximo às posições *orto* e *para*, regiões desfavoráveis (azul) a substituintes com alta densidade eletrônica, indicando uma seleção cuidadosa de substituintes para esta região.

4.1.7. Análise dos Derivados Mais e Menos Potentes

Os compostos mais potentes dos conjuntos de treinamento e de teste, **59** ($pIC_{50Obs} = 7,52$ e resíduo = 0,08) e **74** ($pIC_{50Obs} = 7,30$ e resíduo = 0,06), respectivamente, foram preditos com grande exatidão usando o melhor modelo de CoMFA (Tabela 9).

Do mesmo modo, os compostos menos potentes dos conjuntos de treinamento e de teste, **1** ($pIC_{50Obs} = 4,23$ e resíduo = 0) e **60** ($pIC_{50Obs} = 4,35$ e resíduo = -0,05), respectivamente, também foram preditos com grande exatidão usando o melhor modelo de CoMFA. Estes compostos têm uma baixa potência de inibição enzimática, provavelmente devido ao volume do anel naftila (substituinte da posição **C6** do anel 4-oxo-pirimidina), posicionado em uma região de volume restrito no NNBS, como é mostrado no mapa estérico (poliedros amarelos ao redor do anel naftila) da Figura 24 para o composto **1**.

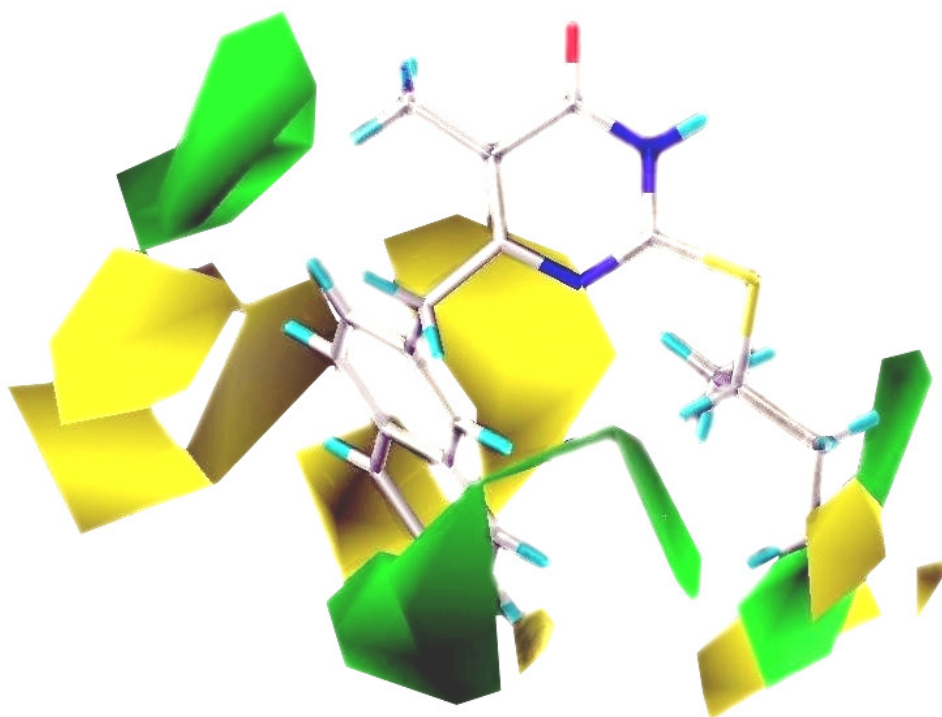


Figura 24. Mapa de contorno estérico do melhor modelo de CoMFA para o composto **1** (em modelo bastão e colorido por elemento), mostrando as áreas onde grupos volumosos aumentam (verde) ou diminuem (amarelo) a potência.

4.1.8. Análise dos Compostos *Outliers*

Neste trabalho, foram considerados como *outliers* aqueles compostos que apresentam resíduo maior do que o dobro do desvio padrão da estimativa ($SEE = 0,226$ e $2 \times SEE = 0,452$). Assim, o melhor modelo de CoMFA (Tabela 6) apresenta dois *outliers* no conjunto de treinamento, compostos **46** (resíduo = 0,48) e **50** (resíduo = 0,72) e três *outliers* no conjunto de teste, compostos **68** (resíduo = 0,51), **69** (resíduo = 0,54) e **70** (resíduo = 0,62).

Considerando os *outliers* do conjunto de treinamento, o composto **50** ($pIC_{50Obs} = 6,92$ M) é o pior predito pelo modelo de CoMFA ($pIC_{50Pred} = 6,20$ M), apresentando, inclusive, o maior valor residual (resíduo = 0,72) de ambos os conjuntos de treinamento e de teste. Analisando este composto (**50**) dentro do mapa

de CoMFA eletrostático, Figura 25, pode-se observar uma região azul próxima a um dos átomos de cloro do anel 2,6-dicloro-fenila, indicando que grupos com alta densidade eletrônica nesta posição diminuem a potência inibitória. Essa proximidade, não observada para o análogo fluorado **52** (resíduo = $-0,17$), implica em uma predição ruim da potência para este composto, provavelmente devido ao maior tamanho do átomo de cloro em relação ao átomo de flúor. Comportamento similar ocorreu com o *outlier* **46** (resíduo = $0,48$), que pode ser comparado ao análogo fluorado **55** (resíduo = $0,05$), e também com o composto **49** (resíduo = $0,43$) (não é *outlier*, mas tem um resíduo grande), que pode ser comparado ao análogo fluorado **51** (resíduo = $-0,01$).

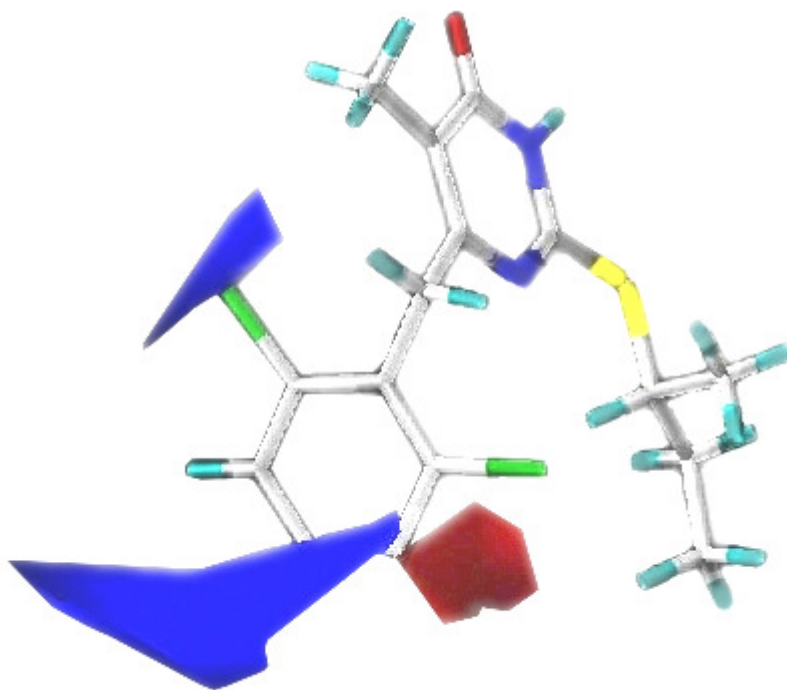


Figura 25. Mapa de contorno eletrostático do melhor modelo de CoMFA para o composto **50** (em modelo bastão e colorido por elemento), mostrando as áreas onde grupos com alta densidade eletrônica aumentam (vermelho) ou diminuem (azul) a potência.

Considerando os *outliers* do conjunto de teste, o composto **70** ($\text{pIC}_{50\text{Obs}} = 6,40$ M) é o pior predito pelo modelo de CoMFA ($\text{pIC}_{50\text{Pred}} = 5,78$ M), apresentando o maior valor residual (resíduo = 0,62) neste conjunto. Comparando **70** (Ar=2,6-dicloro-fenila) com o análogo fluorado **73** (Ar=2,6-difluor-fenila, resíduo = 0), o comportamento como *outlier* deste composto pode ser explicado pela mesma razão apresentada anteriormente para o *outlier* **50**.

Entretanto, para os demais *outliers* do conjunto de teste, compostos **68** (Ar = 3-flúor-fenila, resíduo = 0,51) e **69** (Ar = 2-nitro-fenila, resíduo = 0,54), ambos contendo um padrão de substituição assimétrico na posição **C6** no anel 4-oxo-pirimidina, não é clara a razão do comportamento como *outlier* desses compostos, pois outros compostos substituídos em *orto* ou *meta* foram bem preditos.

4.2. PARTE B – QSAR-3D DEPENDENTE DO RECEPTOR

4.2.1. Análise da Simulação por Dinâmica Molecular dos Complexos Ligantes-Enzima

A análise da curva de energia potencial (kJ/mol) em função do tempo de simulação (ps) de cada complexo ligante-enzima demonstra que o sistema simulado atingiu um equilíbrio em 1000 ps, visto que as flutuações de energia potencial com o decorrer do tempo são suaves e tendem a valores baixos. Na Figura 26 é apresentado como exemplo o gráfico da curva de energia potencial *versus* o tempo de simulação referente ao complexo da RT com o composto **59**.

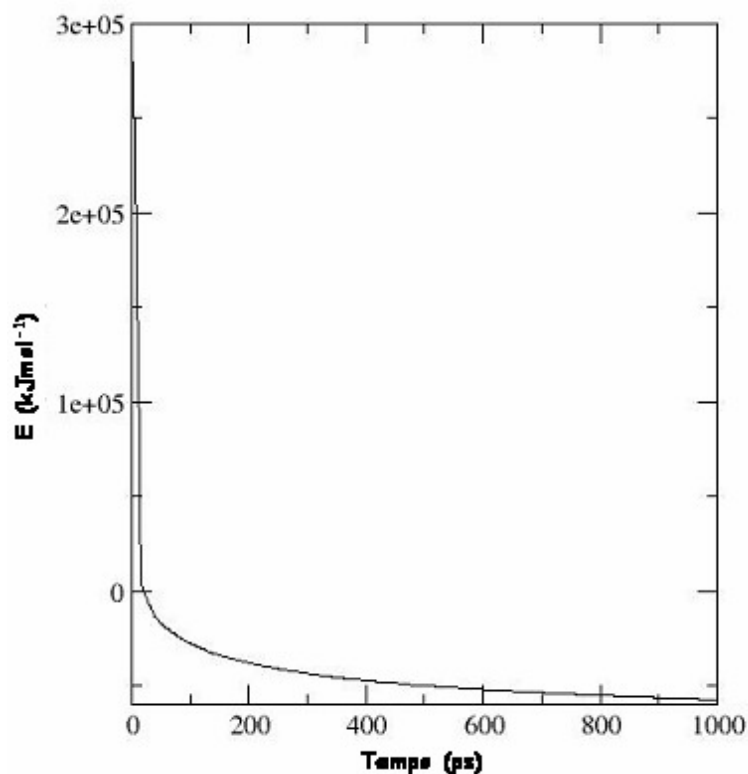


Figura 26. Gráfico da variação da energia potencial (kJ/mol) *versus* o tempo de simulação (ps), em temperatura constante ($T = 310\text{ K}$), do complexo referente ao composto **59**.

4.2.2. Análise do Recorte dos Complexos Ligantes-Enzima

Para o cálculo das energias de interação estéricas e eletrostáticas de cada um dos 74 inibidores com a RT, foram considerados apenas os 53 resíduos de aminoácidos contidos no raio de corte de 10 Å, definido a partir do ligante, listados a seguir de acordo com a subunidade a que pertencem (Figura 27.A).

a) subunidade p66: **Ile94, Pro95, His96, Pro97, Ala98, Gly99, Leu100, Lys101, Lys102, Lys103, Lys104, Ser105, Val106, Thr107, Val108, Ile178, Val179, Ile180, Tyr181, Gln182, Tyr183, Asp186** (catalítico), **Leu187, Tyr188, Val189, Gly190, Ser191, Asp192, His198, Lys223, Glu224, Pro225, Pro226, Phe227, Leu228, Trp229, Met230, Tyr232, Glu233, Leu234, His235, Pro236, Asp237, Lys238, Trp239, Thr240, Tyr317, Tyr318 e Asp319;**

b) subunidade p51: **Asn136, Asn137, Glu138 e Thr139.**

Dos três resíduos catalíticos da RT (*i.e.* **Asp110, Asp185 e Asp186**), apenas o resíduo **Asp186** está contido no recorte de 10 Å.

É interessante ressaltar a importância dos resíduos **Leu100, Lys101, Lys103, Val106, Val108, Tyr181, Tyr188, Gly190, Pro225 e Phe227**, contidos no recorte de 10 Å, que correspondem a posições de freqüente mutação frente aos NNRTIs (El-Brollosy *et al.*, 2002; De Clercq, 1998; Hopkins *et al.*, 1999; Hopkins *et al.*, 1996; Brennan *et al.*, 1995).

Para facilitar a discussão dos resultados quanto à localização espacial dos resíduos no NNBS, delimitamos um raio menor, de 5 Å a partir do ligante, que compreende apenas os seguintes resíduos (Figura 27.B): **Leu100, Lys103, Val106, Val179, Tyr181, Tyr188, Gly190, Phe227, Trp229, Leu234, Pro236 e Tyr318.**

4.2.3. Avaliação dos Bancos de Dados

As melhores equações de QSAR provenientes dos quatro Bancos de Dados (**BD-I**, **BD-II**, **BD-III** e **BD-IV**) estudados, Tabela 10, foram analisadas, considerando os índices estatísticos e o número de compostos *outliers*, com o objetivo de selecionar uma melhor equação por **BD**.

Assim, quatro equações (uma de cada **BD**) foram analisadas em detalhe, comparando os descritores selecionados (energias de interação por resíduos de aminoácidos), os valores de potência preditos e os respectivos valores residuais dos compostos dos conjuntos de treinamento e de teste ($pIC_{50obs} - pIC_{50pred}$) e do composto MKC-442, os compostos *outliers* e as matrizes de correlação cruzada.

Diferente do estudo de CoMFA (QSAR-3D independente do receptor), o estudo de QSAR-3D dependente do receptor fornece uma maior quantidade de informações a serem analisadas, como era esperado, visto que envolve a estrutura 3D da enzima.

Tabela 10. Resumo das características dos quatro bancos de dados (BD) (59 compostos do conjunto de treinamento) usados nas análises de QSAR-3D dependente do receptor.

BD	Características	Nº total de descritores
BD-I	Energias de Lennard-Jones (LJ) e Coulomb (C) calculadas individualmente por resíduo	106 (53 LJ + 53 C)
BD-II	Somatório dos descritores do BD-I por resíduo	53 (53 LJ+C)
BD-III	BD-I + BD-II	159 (53 LJ + 53 C + 53 LJ+C)
BD-IV	Pré-tratamento do BD-I por exclusão das colunas de energia (descritores) com variância < 0,0001	95 (42 LJ + 53 C)

4.2.3.1. Análise do Banco de Dados I

A) Análise dos Índices Estatísticos do BD-I

A Tabela 11 resume os índices estatísticos referentes às cinco melhores equações (**A-E**), contendo de seis a dez termos (descritores), geradas a partir do **BD-I**. As equações **D** (com 9 termos; $q^2_{\text{ajus}} = 0,647$; $SE_{CV} = 0,438$ e 4 *outliers*) e **E** (com 10 termos; $q^2_{\text{ajus}} = 0,660$; $SE_{CV} = 0,420$ e 3 *outliers*) apresentam maiores valores de q^2 ajustado e menores valores de SE_{CV} do que as equações **A-C**. Considerando estas duas equações, a **Eq.E** foi selecionada como a melhor equação do **BD-I**, porque apresenta maior valor de q^2 ajustado e menor número de *outliers* no conjunto de teste.

As equações **A-C** foram geradas com a opção de 10.000 operações de *crossover* (cruzamento), enquanto que as equações **D** e **E** foram geradas com a

opção de 50.000 operações de *crossover*. Como as equações **D** e **E** apresentam os melhores valores estatísticos ($q^2 > 0,6$), conclui-se que um maior número de operações de *crossover* propicia equações mais preditivas. Isso é justificado, uma vez que, a cada nova operação de *crossover*, dois bons modelos (equações) são combinados para a geração de um novo modelo, que carrega as boas “características” (descritores) dos seus antecessores, tornando-o mais robusto.

Tabela 11. Resultados estatísticos dos cinco melhores modelos (**A-E**) de QSAR-3D dependente do receptor obtidos a partir do Banco de Dados I (**BD-I**).

BD-I	N.T.E. ^a	$q^2_{\text{ajus.}}$ ^b	q^2 ^c	SE_{CV} ^d	PC ^e	r^2 ^f	SEE ^g	Outliers ^h
Eq.A	6	0,541	0,581	0,516	4	0,664	0,550	3
Eq.B	7	0,548	0,595	0,507	4	0,688	0,560	3
Eq.C	8	0,523	0,581	0,507	4	0,723	0,590	4
Eq.D	9	0,647	0,696	0,438	5	0,780	0,600	4
Eq.E	10	0,660	0,713	0,420	4	0,822	0,500	3

a) Número de termos da equação (N.T.E.). b) q^2 ajustado ao número de termos da equação. c) r^2 da validação cruzada (q^2). d) Desvio padrão da validação cruzada (SE_{CV}). e) Número ótimo de componentes principais (PC). f) Coeficiente de correlação linear quadrático (r^2). g) Desvio padrão da estimativa (SEE). h) Número de compostos *outliers* no conjunto de teste.

B) Análise da Melhor Equação do BD-I (Eq.E)

Na **Eq.E (BD-I)**, descrita a seguir, cada uma das 10 variáveis independentes (energia de interação estérica ou eletrostática por resíduo de aminoácido) está representada pelo código de três letras do respectivo aminoácido, seguido do código **LJ** (Lennard-Jones) ou **C** (Coulomb), que indica se a interação refere-se à contribuição estérica ou eletrostática, respectivamente. A Figura 28 mostra a representação gráfica 3D da **Eq.E** com o composto NH-DABO **59**.

Eq.E
(BD-I)

$$\begin{aligned}
 \text{pIC}_{50} = & 4,853 - 56,813 \text{ Asn137LJ} - 0,110 \text{ Tyr181LJ} - 0,791 \text{ Gln182C} \\
 & + 0,231 \text{ Pro97LJ} - 0,122 \text{ Tyr188LJ} + 22,417 \text{ Ile94LJ} + 0,087 \text{ His235LJ} \\
 & + 0,043 \text{ Pro226C} + 0,323 \text{ Ser191C} - 0,153 \text{ Lys101LJ}
 \end{aligned}$$

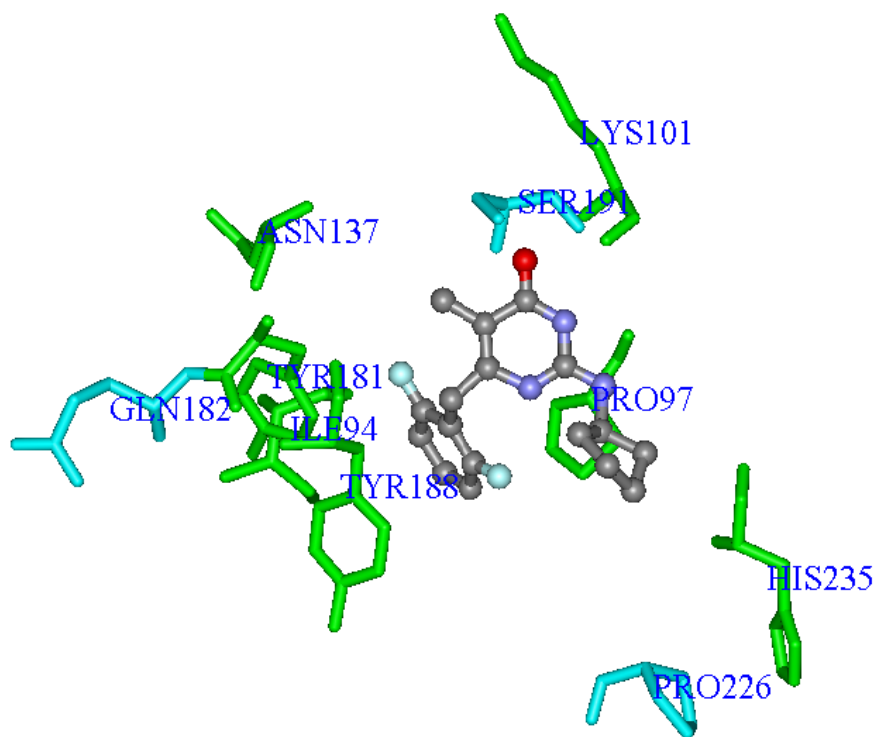


Figura 28. Representação gráfica 3D da **Eq.E (BD-I)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Ile94**, **Pro97**, **Lys101**, **Tyr181**, **Tyr188**, **His235** e **Asn137**) estão coloridos em verde e os que representam contribuições de Coulomb (**Gln182**, **Ser191** e **Pro226**), em azul claro. Os átomos de hidrogênio foram omitidos para melhor visualização.

É oportuno notar que os resíduos **Lys101**, **Tyr181**, **Tyr188**, **His235**, que são descritos na literatura como resíduos que freqüentemente fazem interações com diversos NNRTIs (Kuno *et al.*, 2006; Jorgensen *et al.*, 2006; Ren *et al.*, 2006; Ragno *et al.*, 2004; Parreira *et al.* 2001; Neto *et al.*, 1992a; Neto *et al.*, 1992b), foram selecionados nesta equação.

Na **Eq.E**, apenas três termos representam contribuições de Coulomb (**Gln182C**, **Ser191C** e **Pro226C**), enquanto os outros sete termos representam contribuições de Lennard-Jones (**Ile94LJ**, **Pro97LJ**, **Lys101LJ**, **Tyr181LJ**, **Tyr188LJ**, **His235LJ** e **Asn137LJ**), indicando um maior peso das interações estéricas na relação estrutura-atividade. Esse resultado ratifica a importância das interações estéricas na cavidade hidrofóbica do NNBS, como enfocado por diversos autores (Kuno *et al.*, 2006; Jorgensen *et al.*, 2006; Ren *et al.*, 2006; Janssen *et al.*, 2005). É interessante notar também que os três termos eletrostáticos selecionados nesta equação (**Gln182C**, **Ser191C** e **Pro226C**) estão relacionados a resíduos que se localizam fora do raio de 5 Å, definido anteriormente, o que deve ser justificado pelo fato das interações eletrostáticas serem de maior alcance do que as interações estéricas.

Os valores de pIC_{50} calculados (preditos) pela **Eq.E** sofrem influência da magnitude do coeficiente de cada termo de energia de interação na equação e também do sinal (+ ou -) do respectivo coeficiente. Se o sinal do coeficiente de determinado termo de energia for positivo (+), a energia de interação relativa a este termo será positiva, conseqüentemente, contribuindo para aumentar a potência do composto. Por outro lado, se o sinal deste mesmo coeficiente for negativo (-), a energia de interação relativa a este termo será negativa, conseqüentemente, contribuindo para aumentar a potência do composto.

Como exemplo, o termo **Tyr181LJ** apresenta coeficiente negativo ($-0,110$) na **Eq.E**, portanto, a energia de interação estérica entre o resíduo **Tyr181** e um determinado composto deverá ser negativa para que este termo de energia contribua para aumentar a potência do composto; se a energia de interação for positiva, o termo contribuirá para diminuir a potência do composto.

Em contrapartida, o termo **Ile94LJ** apresenta coeficiente positivo ($+22,417$) na **Eq.E**, portanto, a energia de interação estérica entre o resíduo **Ile94** e um determinado composto deverá ser positiva para que este termo de energia contribua para aumentar a potência do composto; se a energia de interação for negativa, o termo contribuirá para diminuir a potência do composto.

O gráfico da Figura 29 mostra a média dos valores de energia de interação (kcal.mol^{-1}) dos compostos (**1-74**) com os resíduos selecionados na **Eq.E**. Como pode ser observado neste gráfico, a maioria dos inibidores apresenta energias de interação negativas (ou próximas a zero) com a maioria dos resíduos da **Eq.E** (exceto com o resíduo **Ser191**), onde os valores de energia de maior magnitude negativa são referentes aos termos **Lys101LJ**, **Tyr181LJ**, **Tyr188LJ** e **Pro226C**, com média de energia de cerca de $-3,5 \text{ kcal.mol}^{-1}$. Os termos **Pro97LJ** e **His235LJ**, que também correspondem a valores de energia negativos, apresentam médias de energia de cerca de $-0,4 \text{ kcal.mol}^{-1}$. Considerando estes seis termos com valores de energia negativos, os termos **Lys101LJ**, **Tyr181LJ**, **Tyr188LJ**, que apresentam coeficiente de sinal negativo (**Eq.E**), contribuem para aumentar a potência, enquanto que os termos **Pro97LJ**, **Pro226C** e **His235LJ**, que apresentam coeficiente de sinal positivo (**Eq.E**), contribuem para diminuir a potência.

Por sua vez, os termos **Ile94LJ**, **Gln182C** e **Asn137LJ**, que apresentam médias de energia de interação próximas a zero, são os que têm os maiores valores

(modulares) de coeficientes, o que permite equilibrar o peso deles na **Eq.E** em relação aos demais termos. É interessante notar também que estes três termos estão relacionados a resíduos de aminoácidos que localizam-se fora do raio de 5 Å, o que pode justificar também a menor energia (modular) de interação calculada.

Finalmente, o termo **Ser191C (Eq.E)** é o único que apresenta média de energia de interação positiva, o que corresponde a uma interação eletrostática repulsiva, *i.e.*, desfavorável, com valor próximo a 0,3 kcal.mol⁻¹. Paradoxalmente, este termo contribui para aumentar a potência, visto que apresenta coeficiente positivo na **Eq.E**. Outros autores já relataram energias de interação estéricas e eletrostáticas positivas para outros sistemas (Aparna *et al.*, 2007; Kuno *et al.*, 2006).

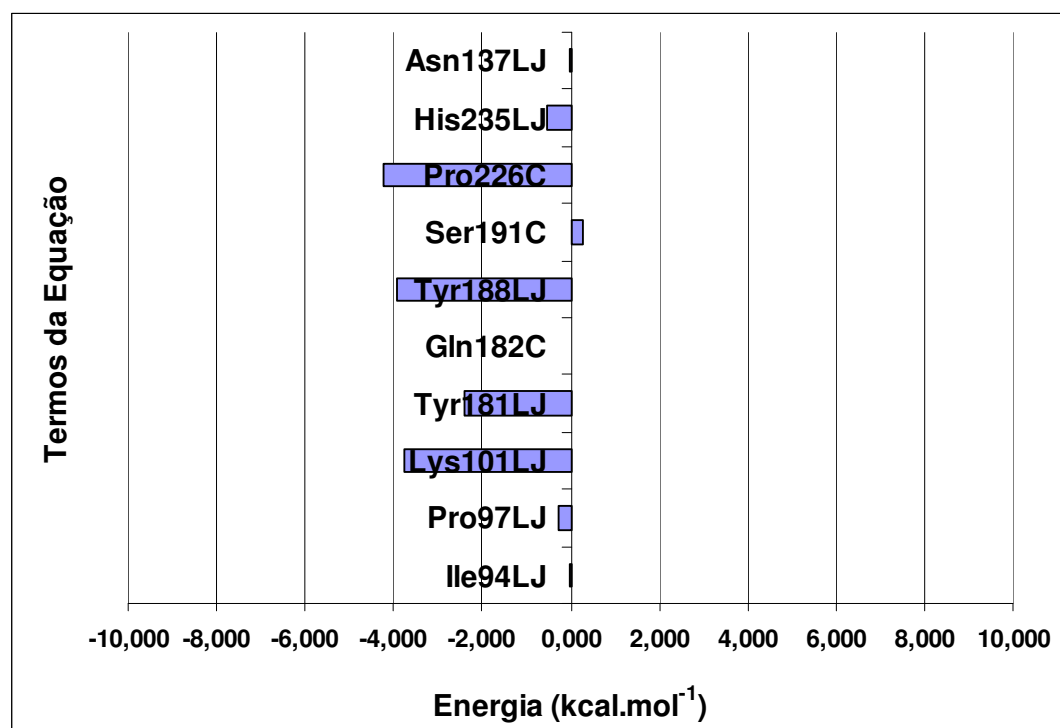


Figura 29. Médias das energias de interação (kcal.mol⁻¹) dos compostos **1-74** com cada um dos termos selecionados na **Eq.E (BD-I)**.

Em relação aos resíduos de aminoácidos na RT que sofrem freqüente mutação frente aos NNRTIs, três deles aparecem na **Eq.E** relacionados aos termos **Lys101LJ**, **Tyr181LJ** e **Tyr188LJ**. Como estes três termos têm valores médios de energia de interação estérica negativos e coeficientes também negativos (**Eq.E**), todos contribuem para aumentar a potência.

No caso da mutação de **Lys101** para **Gly101** (designada como **Lys101Gly**) ocorre um fato curioso. Na RT nativa, o átomo de oxigênio do grupo amida da cadeia principal de **Lys101** é capaz de fazer interação por ligação hidrogênio com o grupo NH do anel 4-oxo-pirimidina dos inibidores, como mostrado na Figura 30 para o composto **59**. Adicionalmente, o grupo amina protonado da cadeia lateral deste mesmo resíduo, pertencente a subunidade p66, é capaz de fazer interação iônica com o grupo carboxilato da cadeia lateral de **Glu28** (a cerca de 5 Å de distância), e **Glu138** (a cerca de 6,5 Å de distância), ambos pertencentes à subunidade p51 (é possível verificar na estrutura de raios-X do complexo do MKC-442 com a RT, dados não mostrados), sendo responsável, portanto, por parte das interações intercadeias (p66-p51). O curioso é que após mutação para **Gly101**, a interação direta via ligação hidrogênio entre o resíduo e o inibidor pode ser mantida, visto que envolve a cadeia principal do resíduo, que não sofre alteração. Porém, as interações intercadeias correspondentes, que dependem da cadeia lateral do resíduo, são perdidas, comprometendo a composição de parte do NNBS (referente aos resíduos de aminoácidos da subunidade p51).

Por outro lado, as mutações **Tyr181Cys** e **Tyr188Leu** correspondem a um caso mais comum, onde a troca entre os resíduos afeta diretamente a interação ligante-receptor, e não indiretamente, como no caso anterior. Em ambas as mutações, ocorre perda das interações do tipo π - π -stacking dos anéis aromáticos

das cadeias laterais de **Tyr181** e **Tyr188** com o anel aromático dos inibidores (Figura 23) (Kuno *et al.*, 2006; Janssen *et al.*, 2005), diminuindo a afinidade destes pelo NNBS.

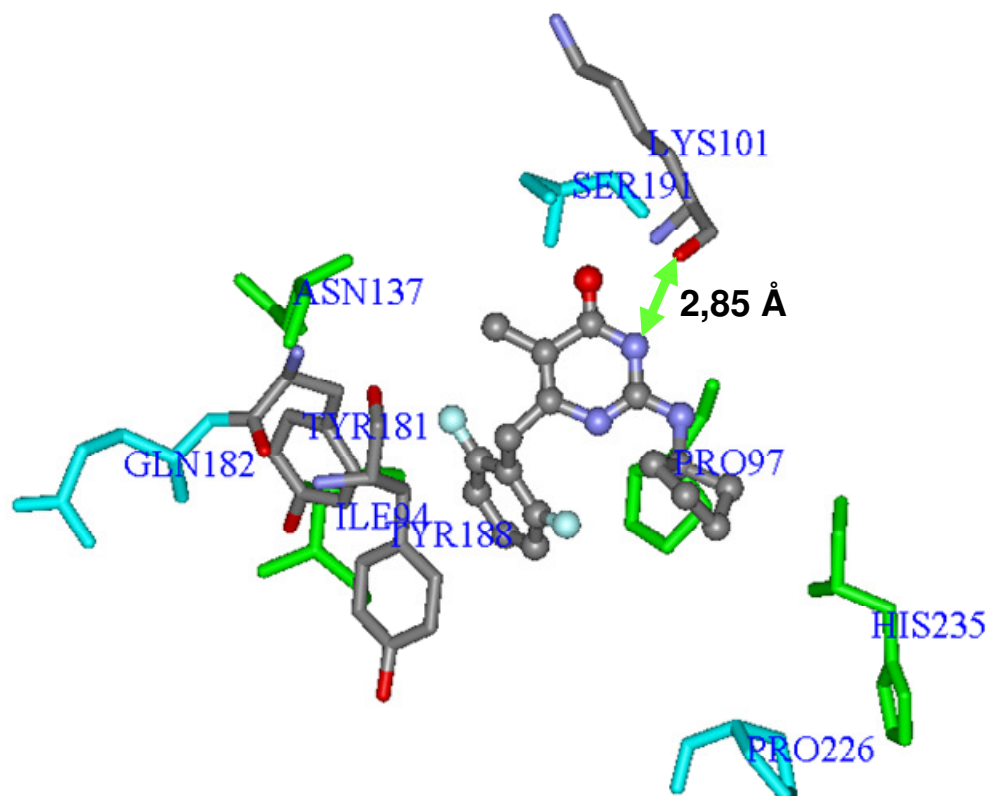


Figura 30. Representação gráfica 3D da **Eq.E (BD-I)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento), destacando os resíduos (em modelo bastão e colorido por elemento) que podem sofrer mutação **Lys101**, **Tyr181** e **Tyr188** (Lennard-Jones). A seta indica uma possível interação por ligação hidrogênio entre **59** e **Lys101**. Em verde (Lennard-Jones) e em azul (Coulomb) estão representados os demais resíduos (em modelo bastão) da equação. Os átomos de hidrogênio foram omitidos para melhor visualização.

C) Análise dos Valores Residuais da Melhor Equação do BD-I (Eq.E)

Na Tabela 12 são mostrados os valores de pIC_{50} (M) observados (experimentais), os valores de pIC_{50} (M) preditos pela **Eq.E** e os respectivos valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (1-59) e de teste (60-74).

A análise da Tabela 12 para o conjunto de treinamento (1-59) mostra que 81% dos compostos apresentam resíduos menores do que 0,50 (em valores modulares). Além disso, nenhum composto do conjunto de treinamento comporta-se como *outlier*, visto que nenhum apresenta valor residual maior do que o dobro do erro padrão da estimativa ($2 \times SEE = 1,00$). Isso mostra uma excelente capacidade preditiva interna do modelo. Na Figura 31.A, os resíduos para os compostos do conjunto de treinamento são mostrados na forma de gráfico de barras.

No caso dos compostos do conjunto de teste (60-74), 40% apresentam resíduos menores do que 0,50, enquanto três compostos são considerados *outliers* (67, 69 e 70). Na Figura 31.B, os resíduos para os compostos do conjunto de teste são mostrados na forma de gráfico de barras.

É importante ressaltar que os resíduos de ambos os conjuntos de treinamento e de teste apresentam variações aleatórias em relação à variação da potência. Isso significa que o modelo não é tendencioso para um valor de atividade mais alto ou mais baixo.

Tabela 12. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**) de acordo com a **Eq.E (BD-I)**.

#	pIC_{50Obs}	pIC_{50Pred}	Res.	#	pIC_{50Obs}	pIC_{50Pred}	Res.
1	4,23	4,23	0,00	38	6,10	5,95	0,15
2	4,31	4,34	-0,03	39	6,10	6,09	0,01
3	4,35	4,30	0,05	40	6,10	5,96	0,14
4	4,59	5,41	-0,82	41	6,10	6,66	-0,56
5	4,77	4,97	-0,20	42	6,22	6,27	-0,05
6	4,79	5,35	-0,56	43	6,22	6,47	-0,25
7	4,83	5,14	-0,31	44	6,22	5,87	0,35
8	4,83	4,94	-0,11	45	6,30	6,04	0,26
9	5,02	4,99	0,03	46	6,40	6,15	0,25
10	5,07	5,37	-0,30	47	6,70	6,73	-0,03
11	5,09	4,94	0,15	48	6,70	6,42	0,28
12	5,27	5,24	0,03	49	6,70	6,56	0,14
13	5,31	5,88	-0,57	50	6,92	6,48	0,44
14	5,31	5,03	0,28	51	7,00	6,89	0,11
15	5,32	5,79	-0,47	52	7,00	7,25	-0,25
16	5,34	5,19	0,15	53	7,05	6,88	0,17
17	5,37	5,67	-0,30	54	7,05	6,98	0,07
18	5,42	5,24	0,18	55	7,10	6,45	0,65
19	5,44	5,38	0,06	56	7,10	6,86	0,24
20	5,47	5,42	0,05	57	7,15	6,61	0,54
21	5,49	5,60	-0,11	58	7,30	7,51	-0,21
22	5,52	5,59	-0,07	59	7,52	6,68	0,84
23	5,52	6,04	-0,52	60	4,35	5,18	-0,83
24	5,52	5,57	-0,05	61	4,48	5,01	-0,53
25	5,54	5,53	0,01	62	5,27	6,03	-0,76
26	5,55	5,16	0,39	63	5,47	5,88	-0,41
27	5,59	5,61	-0,02	64	5,59	5,60	-0,01
28	5,60	6,00	-0,40	65	5,60	5,08	0,52
29	5,60	5,87	-0,27	66	5,62	6,02	-0,40
30	5,62	5,23	0,39	67	5,74	4,68	1,06
31	5,66	5,38	0,28	68	5,92	5,13	0,79
32	5,72	5,40	0,32	69	6,22	8,52	-2,30
33	5,80	6,37	-0,57	70	6,40	5,28	1,12
34	5,89	4,23	-0,34	71	6,70	7,14	-0,44
35	5,94	4,34	0,60	72	7,05	6,92	0,13
36	5,94	6,50	-0,56	73	7,15	6,35	0,80
37	5,96	5,75	0,21	74	7,30	7,06	0,24

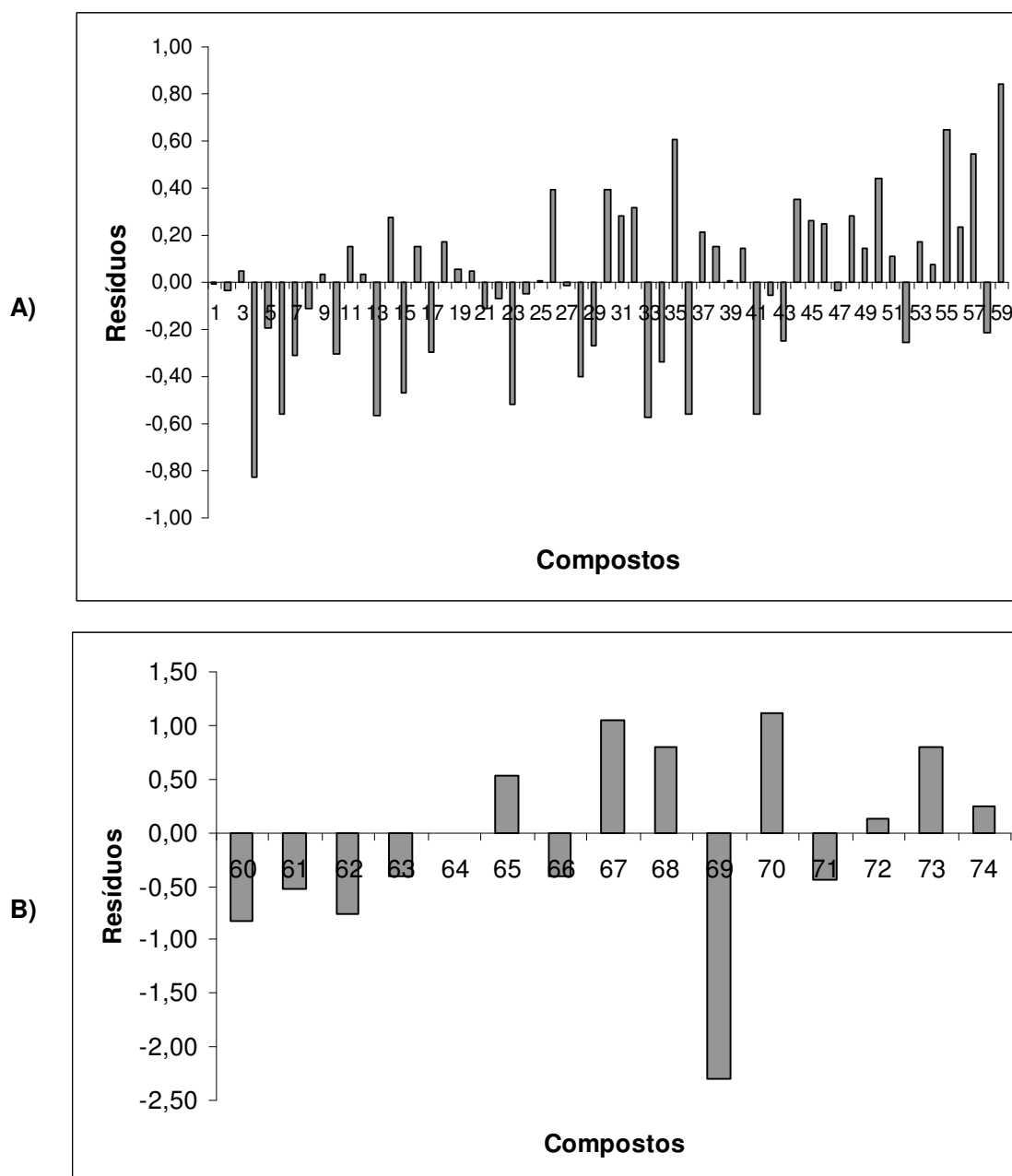


Figura 31. Gráfico de barras dos resíduos dos compostos dos conjuntos de (A) treinamento (1-59) e de (B) teste (60-74) de acordo com a Eq.E (BD-I).

D) Análise dos Compostos *Outliers* da Eq.E (BD-I)

Como mencionado anteriormente, identificou-se apenas três compostos *outliers*: **67**, **69** e **70**, todos do conjunto de teste, cujas estruturas, valores de atividade biológica (pIC_{50}) observados e preditos e respectivos valores residuais, encontram-se na Figura 32, onde se observa que os compostos **67** e **70** apresentam potências preditas inferiores às experimentais, enquanto o composto **69** apresenta potência predita superior à experimental.

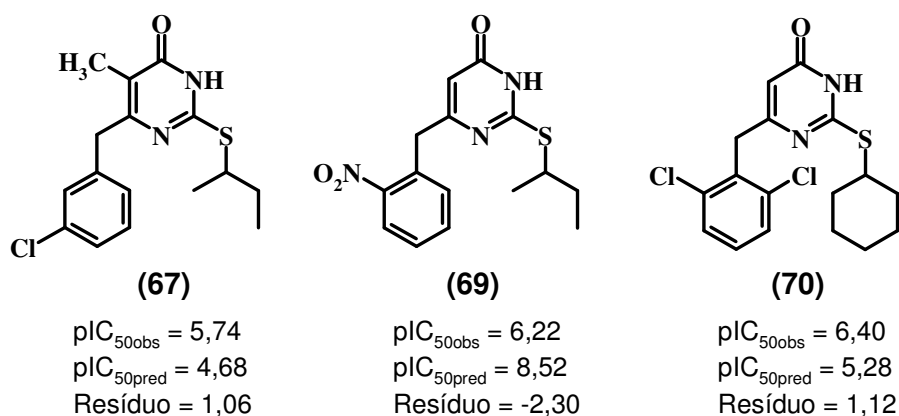


Figura 32. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da Eq.E (BD-I).

Considerando os descritores (valores de energia de interação) selecionados na Eq.E, observamos que, para os três compostos, os termos que apresentam maior variação em relação aos demais termos são **Tyr181LJ** e **Gln182C**, que correspondem a interações estérica e eletrostática, respectivamente.

No caso de **Tyr181LJ**, os valores de energia obtidos são os seguintes: $-2,540$ kcal.mol⁻¹ para o composto **67**, $-7,517$ kcal.mol⁻¹ para o composto **69** e $1,620$ kcal.mol⁻¹ para o composto **70**. Este termo (**Tyr181LJ**), que tem coeficiente negativo

(-0,110) na **Eq.E**, contribui para aumentar a potência, podendo justificar a maior potência predita para **69**, visto que o resíduo **Tyr181** forma interações do tipo π - π -*stacking* com anéis aromáticos dos NNRTIs, como descrito anteriormente. No entanto, não fica claro porque análogos contendo grupo nitro em posição *para* (**30**) ou *meta* (**43**) do substituinte Ar (Tabela 5, Metodologia) não são *outliers*, enquanto que **69** (*orto*-nitro) é *outlier*.

No caso do termo **Gln182C**, que também apresenta coeficiente negativo na **Eq.E** (-0,791) (contribuindo para aumentar a potência), o comportamento parece similar ao do termo anterior, porém, com menor intensidade, visto que os valores de energia obtidos são os seguintes: 0,188 kcal.mol⁻¹ (**67**), -0,606 kcal.mol⁻¹ (**69**) e 0,097 kcal.mol⁻¹ (**70**).

E) Análise do MKC-442 Segundo a Eq.E (BD-I)

A predição da potência para o composto MKC-442, que pertence à classe dos HEPTs, pelos melhores modelos de QSAR-3D dependente do receptor foi empregada como um teste adicional de validação externa, visto que este composto não faz parte do banco de dados dos DABOs; e também porque utilizou-se a conformação e a orientação deste composto complexado com a RT do HIV-1 (estrutura de raios-X 1RT1) como dados de partida para a construção e o posicionamento dos DABOs no NNBS.

Desta forma, de acordo com a **Eq.E (BD-I)**, o composto MKC-442 têm potência predita ($pIC_{50} = 6,68$ M) inferior à experimental, visto que o valor de IC_{50obs} deste composto é igual à 0,04 μ M (Mai *et al.*, 1999), *i.e.* $pIC_{50obs} = 7,40$ M. Logo, o valor

residual é igual a $-0,72$, que é inferior ao dobro do SEE da **Eq.E**, indicando que o composto não é um *outlier*.

F) Análise da Matriz de Correlação Cruzada entre os Descritores da Eq.E (BD-I)

A análise da matriz de correlação cruzada entre os termos da **Eq.E** (Tabela 13) mostra que não há correlação significativa entre os diferentes descritores (energias de interação estérica e eletrostática), uma vez que não há nenhum valor de r (coeficiente de correlação linear simples) maior do que $0,70$ (em valor modular) (Livingtone, 1996). Isso demonstra que cada descritor traz informação única para o modelo, não havendo, portanto, informação redundante. A maior correlação encontrada ocorre entre os termos **Tyr181LJ** e **Tyr188LJ** ($r = 0,543$), provavelmente devido a proximidade espacial entre os resíduos de aminoácidos correspondentes no NNBS e ao fato do mesmo tipo de aminoácido (Tyr) estar envolvido.

Tabela 13. Matriz de correlação cruzada entre os descritores da **Eq.E (BD-I)**.

	Ile94 LJ	Pro97 LJ	Lys101 LJ	Tyr181 LJ	Gln182 C	Tyr188 LJ	Ser191 C	Pro226 C	His235 LJ	Asn137 LJ
Ile94 LJ	1,000									
Pro97 LJ	-0,126	1,000								
Lys101 LJ	0,155	-0,122	1,000							
Tyr181 LJ	0,354	0,039	-0,131	1,000						
Gln182 C	0,238	-0,107	0,116	-0,073	1,000					
Tyr188 LJ	0,176	0,169	-0,212	0,543	-0,243	1,000				
Ser191 C	0,086	0,001	0,260	0,194	0,043	0,166	1,000			
Pro226 C	0,204	-0,083	0,214	0,200	0,056	0,029	0,313	1,000		
His235 LJ	0,080	0,131	-0,087	-0,030	-0,203	-0,037	-0,320	-0,214	1,000	
Asn137 LJ	0,435	-0,011	-0,099	0,036	0,216	0,012	-0,055	-0,012	0,135	1,000

4.2.3.2. Análise do Banco de Dados II

A) Análise dos Índices Estatísticos do BD-II

A Tabela 14 resume os índices estatísticos referentes às cinco melhores equações (**F-J**), contendo de seis a dez termos (descritores), geradas a partir do **BD-II**. As equações **I** (com 9 termos; $q^2_{\text{ajus}} = 0,592$; $SE_{CV} = 0,473$ e 3 *outliers*) e **J** (com 10 termos; $q^2_{\text{ajus}} = 0,606$; $SE_{CV} = 0,460$ e 3 *outliers*) apresentam maiores valores de q^2 ajustado e menores valores de SE_{CV} do que as equações **F-H**. Considerando estas duas equações, a **Eq.J** foi selecionada como a melhor equação

do **BD-II**, porque apresenta maior valor de q^2 ajustado. Novamente, em analogia ao que foi observado para o **BD-I**, verifica-se que as equações geradas a partir de um número maior de operações de *crossover* (50.000), **Eq.I** e **Eq.J**, apresentam melhores resultados.

Tabela 14. Resultados estatísticos dos cinco melhores modelos (**F-J**) de QSAR-3D dependente do receptor obtidos a partir do Banco de Dados II (**BD-II**).

BD-II	N.T.E.^a	$q^2_{\text{ajus.}}^b$	q^2^c	SE_{cv}^d	PC^e	r^2^f	SEE^g	Outliers^h
Eq.F	6	0,517	0,559	0,530	5	0,641	0,720	3
Eq.G	7	0,507	0,558	0,532	3	0,658	0,710	3
Eq.H	8	0,527	0,584	0,515	6	0,703	1,520	1
Eq.I	9	0,592	0,648	0,473	4	0,743	0,570	3
Eq.J	10	0,606	0,667	0,460	6	0,766	0,600	3

a) Número de termos da equação (N.T.E). b) q^2 ajustado ao número de termos da equação. c) r^2 da validação cruzada (q^2). d) Desvio padrão da validação cruzada (SE_{cv}). e) Número ótimo de componentes principais (PC). f) Coeficiente de correlação linear quadrático (r^2). g) Desvio padrão da estimativa (SEE). h) Número de compostos *outliers* no conjunto de teste.

B) Análise da Melhor Equação do BD-II (Eq.J)

Na **Eq.J (BD-II)**, descrita a seguir, cada uma das 10 variáveis independentes (energias de interação estérica e eletrostática somadas por resíduo de aminoácido) está representada apenas pelo código de três letras do próprio aminoácido. A Figura 33 mostra a representação gráfica 3D da **Eq.J** com o composto NH-DABO **59**.

$$\begin{aligned}
 \text{Eq.J} \quad \text{pIC}_{50} = & 6,802 + 1,202 \text{ Leu187} + 0,061 \text{ Asp237} - 0,026 \text{ Trp229} \\
 \text{(BD-II)} \quad & + 0,101 \text{ Gly99} + 1,437 \text{ Thr139} - 0,059 \text{ Tyr188} + 0,360 \text{ Ser191} \\
 & + 0,822 \text{ Glu224} - 0,028 \text{ Phe227} - 0,244 \text{ Tyr183}
 \end{aligned}$$

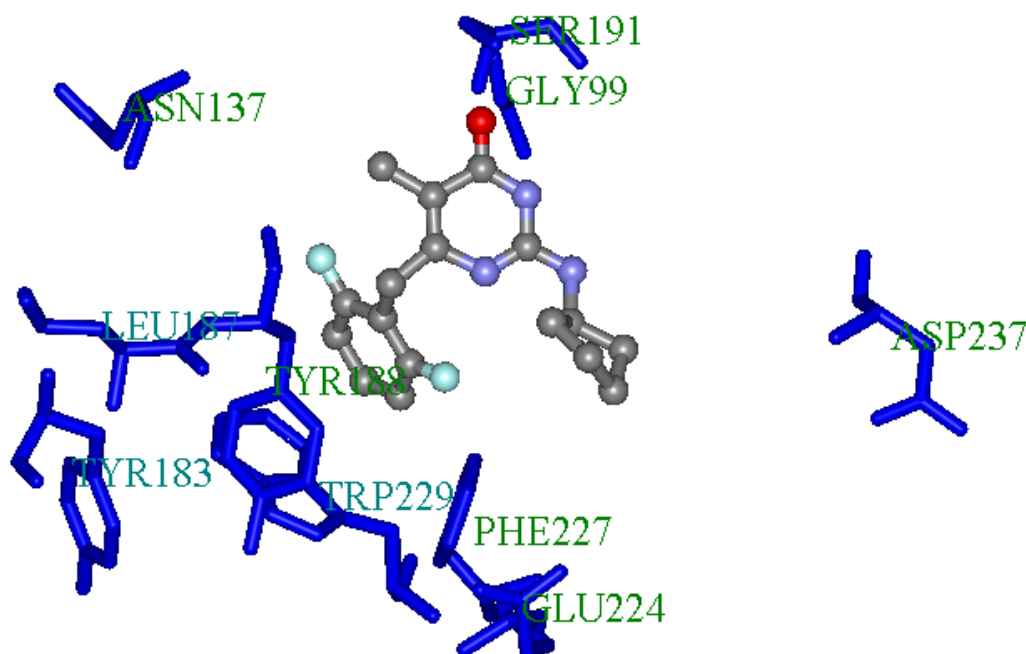


Figura 33. Representação gráfica 3D da **Eq.J (BD-II)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) selecionados nesta equação (**Leu187**, **Asp237**, **Trp229**, **Glu224**, **Gly99**, **Phe227**, **Ser191**, **Tyr188**, **Thr139** e **Tyr183**) representam as interações de Lennard-Jones e Coulomb somadas e estão coloridos em azul escuro. Os átomos de hidrogênio foram omitidos para melhor visualização.

Nesta equação (**Eq.J**), também foram selecionados termos que correspondem a resíduos descritos na literatura por fazer interações com os NNRTIs: **Tyr188**, **Phe227** e **Trp229** (Jorgensen *et al.*, 2006; Ragno *et al.*, 2004). É interessante notar que os três estão contidos no raio de 5 Å, definido anteriormente, e que dos três, apenas **Tyr188** corresponde a um termo também selecionado na **Eq.E** do **BD-I**.

Comparando as equações **E (BD-I)** e **J (BD-II)**, apenas os termos referentes aos resíduos **Tyr188** e **Ser191** aparecem nas duas equações. A análise da matriz de correlação cruzada entre os descritores (ver Tabela A do Anexo) mostra que os

termos **Tyr188LJ (Eq.E)** e **Tyr188 (Eq.J)** não estão fortemente correlacionados ($r = 0,467$). Os termos **Ser191C (Eq.E)** e **Ser191 (Eq.J)** estão perfeitamente correlacionados ($r = 1,0$), indicando que houve apenas uma substituição entre estes termos de uma equação para a outra. Esta correlação se estende também a **Ser191** e **Ser191LJ**.

Os resíduos que correspondem a termos que aparecem na **Eq.E (BD-I)** e que não aparecem na **Eq.J (BD-II)** são: **Ile94**, **Pro97**, **Lys101**, **Tyr181**, **Gln182**, **Pro226**, **His235** e **Asn137**. A ausência do termo que corresponde ao resíduo **Tyr181** na **Eq.J (BD-II)** pode comprometer os resultados obtidos com esta equação, visto que este resíduo, como dito anteriormente, é importante para a interação com os NNRTIs. Por outro lado, analisando a matriz de correlação cruzada entre os descritores, a ausência do termo correspondente ao resíduo **Tyr181LJ** na **Eq.J** pode estar sendo compensada, parcialmente, pela presença do termo correspondente ao resíduo **Tyr183** ($r = -0,505$). No caso dos pares de termos **Pro226C (Eq.E)** e **Phe227 (Eq.J)** ($r = 0,759$) e **Asn137LJ (Eq.E)** e **Thr139 (Eq.J)** ($r = -0,619$), a análise da matriz de correlação cruzada entre os descritores (ver Tabela A do Anexo) mostra que há correlação, indicando que houve uma substituição entre estes pares de termos, provavelmente devido a proximidade espacial. Em relação aos demais termos das equações **E** e **J**, não se observa correlação significativa.

O gráfico da Figura 34 mostra a média das energias de interação (kcal.mol^{-1}) dos compostos (**1-74**) com os resíduos selecionados na **Eq. J**, onde o termo **Phe227** se destaca dos demais por ser o mais estabilizante, *c.a.* $-29 \text{ kcal.mol}^{-1}$. Os termos **Gly99**, **Tyr183**, **Ser191**, **Glu224** e **Thr139** apresentam médias de energia próximas a zero, enquanto os termos **Tyr188**, **Trp229** e **Asp237** apresentam as médias mais altas, *c.a.* 7 kcal.mol^{-1} . Com exceção do termo **Asp237**, os termos que apresentam

maiores médias (em valor modular) de energia de interação estão relacionados a resíduos de aminoácidos que localizam-se dentro do raio de 5 Å (**Tyr188**, **Phe227** e **Trp229**), mais próximos ao inibidor.

Considerando os valores médios de energia (Figura 34) e os sinais dos coeficientes dos termos na **Eq.J (BD-II)**, os termos **Tyr183**, **Ser191**, **Glu224**, **Asp237** e **Thr139** contribuem para aumentar a potência, enquanto que os demais termos contribuem para diminuir a potência.

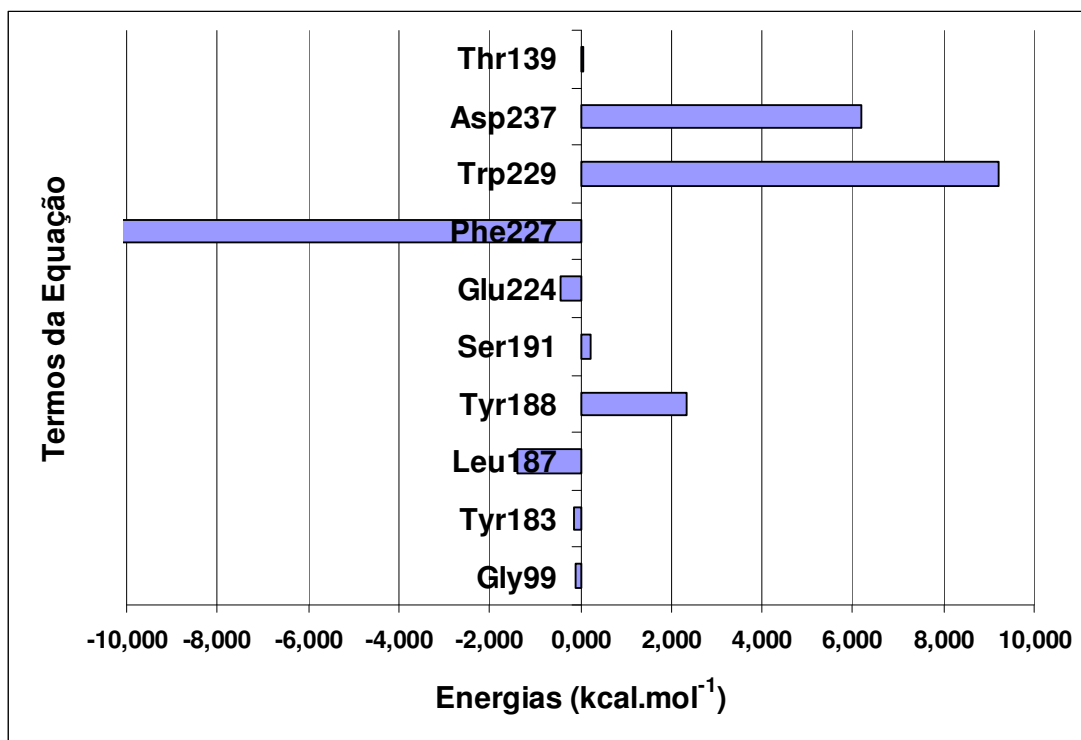


Figura 34. Médias das energias de interação (kcal.mol⁻¹) dos compostos **1-74** para cada um dos termos selecionados na **Eq.J (BD-II)**. O gráfico foi truncado em -10 kcal.mol⁻¹. A média de energia referente ao termo **Phe227**= -29,037 kcal.mol⁻¹.

Em relação aos resíduos de aminoácidos na RT que sofrem freqüente mutação frente aos NNRTIs, dois deles aparecem na **Eq.J** relacionados aos termos **Tyr188** e **Phe227** (ambos referentes a resíduos de aminoácidos contidos no raio de 5 Å), sendo que apenas o termo **Phe227** contribui para aumentar a potência, porque tem coeficiente e valor médio de energia de interação negativos. Estes dois resíduos de aminoácidos sofrem mutação para o mesmo aminoácido (**Leu**), gerando os respectivos mutantes: **Tyr188Leu** (discutido anteriormente na **Eq.E** do **BD-I**) e **Phe227Leu**. Novamente, em ambas as mutações ocorre perda das interações do tipo π - π -*stacking* dos anéis aromáticos das cadeias laterais de **Tyr188** e **Phe227** com o anel aromático dos inibidores, diminuindo a afinidade destes pelo NNBS.

C) Análise dos Valores Residuais da Eq.J (BD-II)

Na Tabela 15 são mostrados os valores de pIC_{50} (M) observados (experimentais), os valores de pIC_{50} (M) preditos pela **Eq.J (BD-II)** e os respectivos valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**).

Tabela 15. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**) de acordo com a **Eq.J (BD-II)**.

#	pIC_{50Obs}	pIC_{50Pred}	Res.	#	pIC_{50Obs}	pIC_{50Pred}	Res.
1	4,23	4,40	-0,17	38	6,10	5,98	0,12
2	4,31	4,45	-0,14	39	6,10	5,29	0,81
3	4,35	4,29	0,06	40	6,10	6,12	-0,02
4	4,59	5,35	-0,76	41	6,10	6,57	-0,47
5	4,77	4,87	-0,10	42	6,22	6,68	-0,46
6	4,79	4,57	0,22	43	6,22	6,77	-0,55
7	4,83	5,21	-0,38	44	6,22	6,16	0,06
8	4,83	5,07	-0,24	45	6,30	6,78	-0,48
9	5,02	5,39	-0,37	46	6,40	6,10	0,30
10	5,07	5,17	-0,10	47	6,70	6,83	-0,13
11	5,09	5,02	0,07	48	6,70	6,53	0,17
12	5,27	5,37	-0,10	49	6,70	6,10	0,60
13	5,31	6,06	-0,75	50	6,92	5,93	0,99
14	5,31	5,48	-0,17	51	7,00	6,76	0,24
15	5,32	5,63	-0,31	52	7,00	6,41	0,59
16	5,34	5,12	0,22	53	7,05	6,45	0,60
17	5,37	6,32	-0,95	54	7,05	7,14	-0,09
18	5,42	5,33	0,09	55	7,10	6,66	0,44
19	5,44	5,68	-0,24	56	7,10	6,91	0,19
20	5,47	5,33	0,14	57	7,15	6,81	0,34
21	5,49	4,83	0,66	58	7,30	7,11	0,19
22	5,52	5,10	0,42	59	7,52	6,71	0,81
23	5,52	5,57	-0,05	60	4,35	5,57	-1,22
24	5,52	5,65	-0,13	61	4,48	3,72	0,76
25	5,54	5,43	0,11	62	5,27	6,94	-1,67
26	5,55	5,69	-0,14	63	5,47	5,97	-0,50
27	5,59	5,60	-0,01	64	5,59	5,62	-0,03
28	5,60	5,83	-0,23	65	5,60	4,46	1,14
29	5,60	6,03	-0,43	66	5,62	8,19	-2,57
30	5,62	5,62	-0,00	67	5,74	5,12	0,62
31	5,66	5,88	-0,22	68	5,92	5,05	0,87
32	5,72	5,69	0,03	69	6,22	6,68	-0,46
33	5,80	5,97	-0,17	70	6,40	5,61	0,79
34	5,89	6,00	-0,11	71	6,70	6,40	0,30
35	5,94	5,63	0,31	72	7,05	7,04	0,01
36	5,94	6,22	-0,28	73	7,15	6,05	1,10
37	5,96	5,98	-0,02	74	7,30	6,43	0,87

A análise da Tabela 15 para o conjunto de treinamento (**1-59**) mostra que 81% dos compostos apresentam resíduos menores do que 0,50 (em valores modulares). Além disso, nenhum composto do conjunto de treinamento comporta-se como *outlier*, visto que nenhum apresenta valor residual maior do que o dobro do erro padrão da estimativa ($2 \times \text{SEE} = 1,20$). Isso mostra uma excelente capacidade preditiva interna do modelo. Na Figura 35.A, os resíduos para os compostos do conjunto de treinamento são mostrados na forma de gráfico de barras.

No caso dos compostos do conjunto de teste (**60-74**), 26% apresentam resíduos menores do que 0,50 (Tabela X), enquanto três compostos são considerados *outliers* (**60**, **62** e **66**). Na Figura 35.B, os resíduos para os compostos do conjunto de teste são mostrados na forma de gráfico de barras.

Em analogia a **Eq.E (BD-I)**, os resíduos de ambos os conjuntos (treinamento e teste) da **Eq.J (BD-II)** também apresentam variações aleatórias em relação à variação da potência. Isso significa que este modelo também não é tendencioso para um valor de atividade mais alto ou mais baixo.

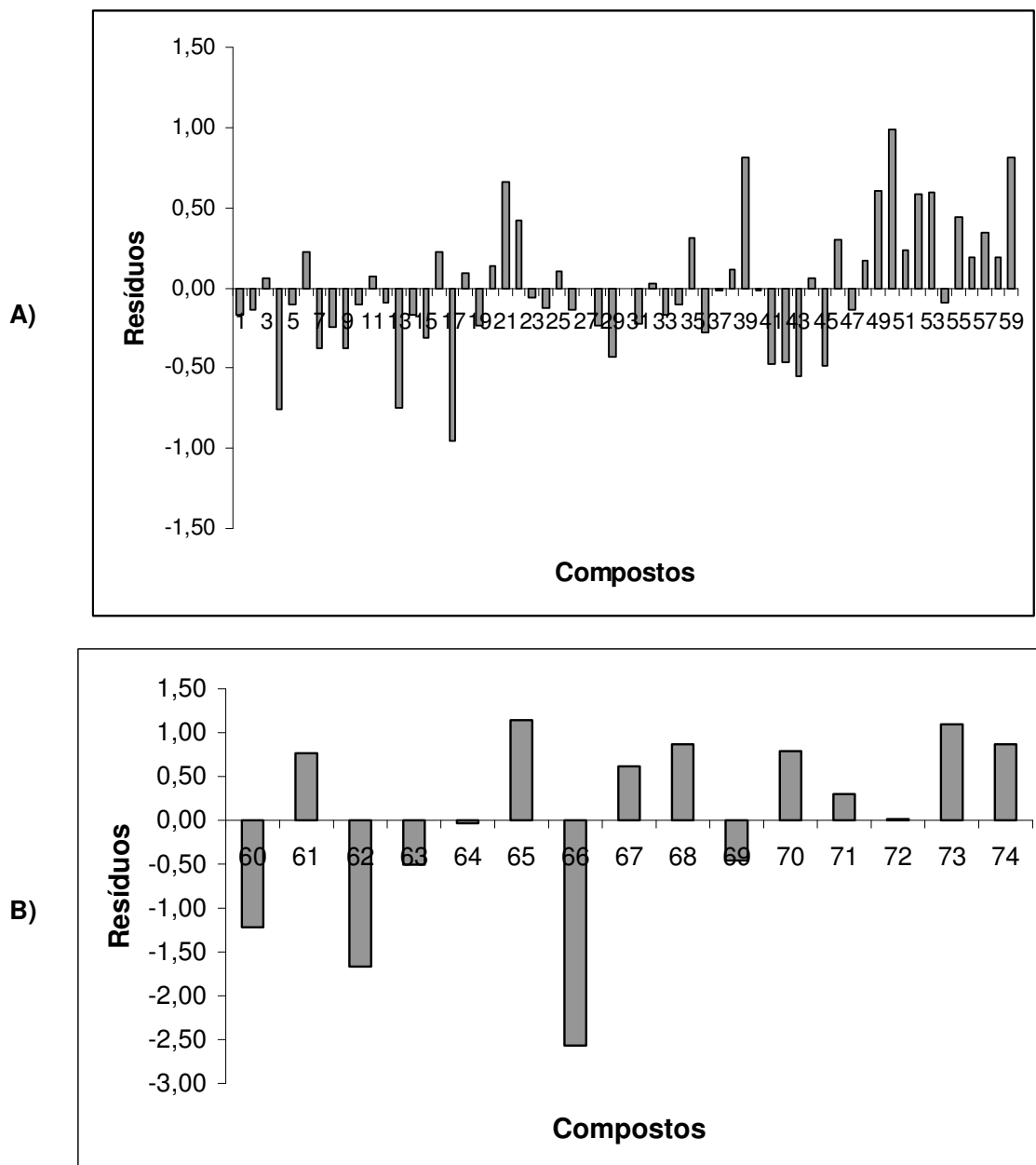


Figura 35. Gráfico de barras dos resíduos ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos dos conjuntos de (A) treinamento (1-59) e de (B) teste (60-74) de acordo com a **Eq.J (BD-II)**.

D) Análise dos Compostos *Outliers* da Eq.J (BD-II)

Na **Eq.J (BD-II)**, como dito anteriormente, identificou-se apenas três compostos *outliers*: **60**, **62** e **66**, todos do conjunto de teste, porém, diferentes daqueles da **Eq.E (BD-I)**. A Figura 36 mostra as estruturas, os valores de pIC_{50} observados e preditos e os respectivos valores residuais para estes *outliers*, onde se observa que todos apresentam potências preditas superiores às observadas experimentalmente, indicando que o modelo superestimou a potência desses compostos, principalmente em relação ao composto **66**.

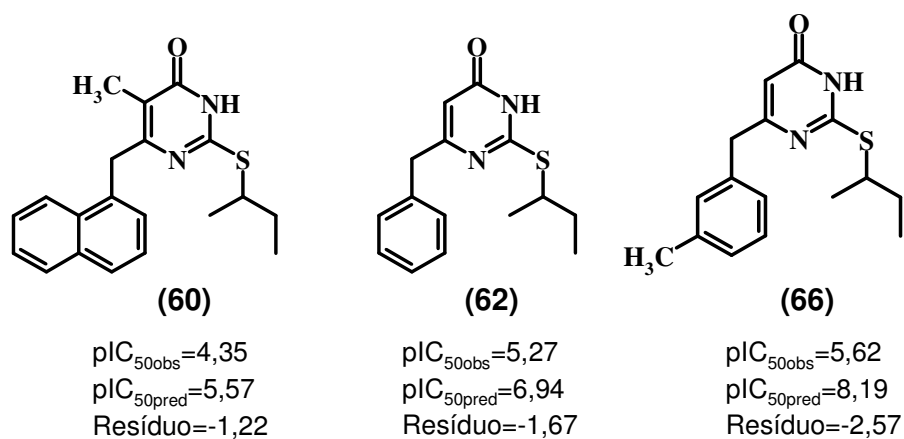


Figura 36. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.J (BD-II)**.

Considerando os descritores (*i.e.* energias de interação) selecionados na **Eq.J (BD-II)**, observa-se que, para o composto **66**, os termos que apresentam maior variação em relação aos demais termos são referentes aos resíduos **Tyr188** e **Trp229**, o que pode explicar o comportamento *outlier* deste composto. Como os valores de energia obtidos para o composto **66** em relação aos termos **Tyr188** e **Trp229** são negativos ($-8,083 \text{ kcal.mol}^{-1}$ e $-6,303 \text{ kcal.mol}^{-1}$, respectivamente) e como ambos os termos têm coeficientes também negativos na **Eq.J** ($-0,059$ e -

0,026, respectivamente), estes termos contribuem para aumentar a potência, justificando a maior potência predita para **66**.

No caso dos *outliers* **60** e **62**, o termo **Phe227** é o que apresenta maior variação em relação aos demais termos. Como os valores de energia de interação obtidos entre o resíduo **Phe227** e os compostos **60** e **62** são negativos ($-3,287 \text{ kcal.mol}^{-1}$ e $-20,990 \text{ kcal.mol}^{-1}$, respectivamente) e como o coeficiente do termo **Phe227** na **Eq.J** também é negativo ($-0,028$), este termo contribui para aumentar a potência, justificando a maior potência predita para **60** e **62**.

É oportuno lembrar que estes três resíduos (**Tyr188**, **Phe227** e **Trp229**) estão dentro do raio de 5 Å e fazem interações do tipo π - π -stacking com o anel aromático dos inibidores.

E) Análise do MKC-442 Segundo a Eq.J (BD-II)

De acordo com a **Eq.J (BD-II)**, o composto MKC-442 têm potência predita ($\text{pIC}_{50} = 9,50 \text{ M}$) superior à experimental ($\text{pIC}_{50\text{Pred}} = 7,40 \text{ M}$) e, conseqüentemente, valor residual (resíduo = 2,10) superior ao dobro do SEE da **Eq.J**, indicando que o composto é um *outlier*.

Considerando que os valores de r^2 e q^2 das equações do **BD-I** (equações **A** a **E**) foram maiores do que os valores das respectivas equações do **BD-II** (equações **F** a **J**), contendo o mesmo número de termos, as equações provenientes do **BD-I** apresentam, em conjunto, melhor ajuste dos dados e maior capacidade preditiva do que as equações provenientes do **BD-II**. A melhor predição do MKC-442 de acordo com a **Eq.E (BD-I)** confirma este dado.

F) Análise da Matriz de Correlação Cruzada entre os Descritores da Eq.J (BD-II)

A análise da matriz de correlação cruzada entre os termos da **Eq.J** (Tabela 16) mostra que não há correlação significativa entre os diferentes descritores (energias de interação estérica e eletrostática somadas por resíduo), uma vez que nenhum valor de $r > 0,70$ foi encontrado. Isso demonstra que cada descritor traz informação única para o modelo, não havendo informação redundante. A maior correlação encontrada ocorre entre os termos **Glu224** e **Phe227** ($r = 0,587$), provavelmente devido a proximidade espacial entre os resíduos de aminoácidos correspondentes no NNBS.

Tabela 16. Matriz de correlação cruzada entre os descritores da **Eq.J (BD-II)**.

	Gly99	Tyr183	Leu187	Tyr188	Ser191	Glu224	Phe227	Trp229	Asp237	Thr139
Gly99	1,000									
Tyr183	0,029	1,000								
Leu187	-0,232	0,383	1,000							
Tyr188	-0,091	-0,297	-0,136	1,000						
Ser191	-0,141	0,007	-0,052	0,360	1,000					
Glu224	-0,227	0,151	0,120	0,228	0,232	1,000				
Phe227	-0,381	0,053	0,439	0,248	0,333	0,587	1,000			
Trp229	0,141	0,160	-0,046	0,100	-0,172	-0,231	-0,348	1,000		
Asp237	0,242	0,152	-0,146	-0,116	-0,102	-0,020	-0,019	-0,123	1,000	
Thr139	0,094	0,136	-0,100	-0,062	0,025	-0,086	-0,140	-0,189	0,195	1,000

4.2.3.3. Análise do Banco de Dados III

A) Análise dos Índices Estatísticos do BD-III

A Tabela 17 resume os índices estatísticos referentes às quatro melhores equações (**K-N**), contendo de seis a nove termos, geradas a partir do **BD-III**. Não foram obtidas equações com dez termos a partir deste banco de dados. Em geral, comparando com as equações obtidas a partir dos **BD-I** e **BD-II**, as equações obtidas a partir do **BD-III** apresentam maiores valores de SEE e maior número de compostos *outliers*.

Considerando apenas as equações do **BD-III**, as equações **L** (com 7 termos; $q^2_{\text{ajus}} = 0,594$; $SE_{CV} = 0,480$ e 4 *outliers*), **M** (com 8 termos; $q^2_{\text{ajus}} = 0,584$; $SE_{CV} = 0,480$ e 7 *outliers*) e **N** (com 9 termos; $q^2_{\text{ajus}} = 0,578$; $SE_{CV} = 0,479$ e 6 *outliers*) apresentam maiores valores de q^2 ajustado e menores valores de SE_{CV} do que a equação **K**.

Dentre estas três equações, a **Eq. L** foi selecionada como a melhor equação do **BD-III**, porque apresenta maior valor de q^2 ajustado e menor número de compostos *outliers*. Novamente, em analogia ao que foi observado para os bancos de dados **I** e **II**, verifica-se que as **Eq. L, M e N**, geradas a partir de um número maior de operações de *crossover* (50.000), apresentam melhores resultados.

Tabela 17. Resultados estatísticos dos quatro melhores modelos (**K-N**) de QSAR-3D dependente do receptor obtidos a partir do Banco de Dados III (**BD-III**).

BD-III	N.T.E.^a	q²_{ajus.}^b	q²^c	SE_{cv}^d	PC^e	r²^f	SEE^g	Outliers^h
Eq.K	6	0,506	0,549	0,537	3	0,630	1,35	6
Eq.L	7	0,594	0,636	0,480	3	0,723	1,18	4
Eq.M	8	0,584	0,634	0,480	5	0,737	0,85	7
Eq.N	9	0,578	0,636	0,479	4	0,753	0,83	6

a) Número de termos da equação (N.T.E). b) q² ajustado ao número de termos da equação. c) r² da validação cruzada (q²). d) Desvio padrão da validação cruzada (SE_{cv}). e) Número ótimo de componentes principais (PC). f) Coeficiente de correlação linear quadrático (r²). g) Desvio padrão da estimativa (SEE). h) Número de compostos *outliers* no conjunto de teste.

B) Análise da Melhor Equação do BD-III (Eq.L)

Como o **BD-III** é formado pela união dos **BD-I** e **BD-II**, na **Eq.L (BD-III)**, descrita a seguir, seis variáveis independentes são provenientes do **BD-I (Ile94LJ, Tyr181LJ, Pro225C, His235LJ, Asn137LJ e Glu138C)** e uma é proveniente do **BD-II (Glu224)**.

Os resíduos **Ile94, Tyr181, His235 e Asn137** representam contribuições de Lennard-Jones (**LJ**), os resíduos **Glu138 e Pro225** representam contribuições de Coulomb (**C**), enquanto que o resíduo **Glu224** representa o somatório das contribuições de Lennard-Jones e Coulomb. A Figura 37 mostra a representação gráfica 3D da **Eq.L** com o composto NH-DABO **59**.

$$\begin{aligned}
 \text{Eq.L} \quad \text{pIC}_{50} &= 6,257 + 0,062 \text{ His235LJ} + 15,851 \text{ Ile94LJ} - 52,568 \text{ Asn137LJ} \\
 (\text{BD-III}) \quad &- 0,118 \text{ Tyr181LJ} - 0,008 \text{ Glu138C} + 0,101 \text{ Pro225C} + 3,525 \text{ Glu224}
 \end{aligned}$$

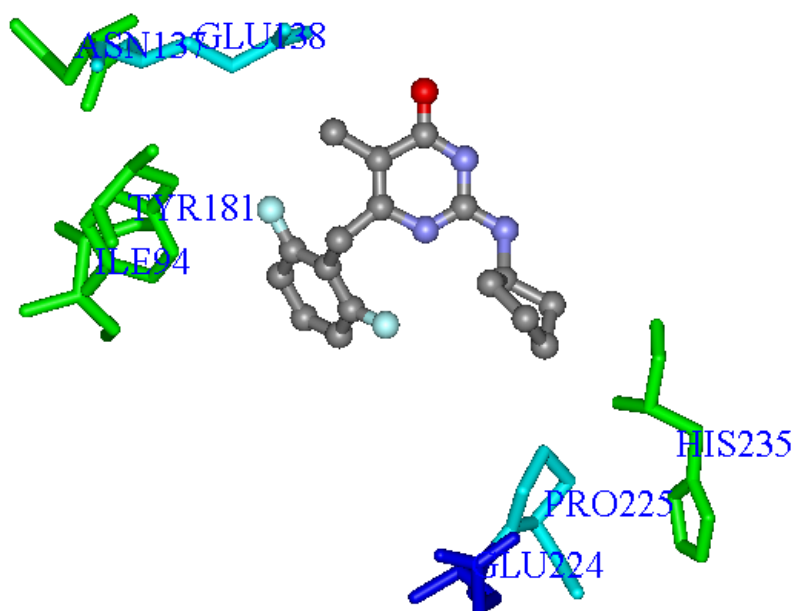


Figura 37. Representação gráfica 3D da **Eq.L (BD-III)** com o composto **NH-DABO 59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Ile94**, **Tyr181**, **His235** e **Asn137**) estão coloridos em verde e os que representam contribuições de Coulomb (**Glu138** e **Pro225**), em azul claro. O resíduo **Glu224**, colorido em azul escuro, representa o somatório das contribuições de Lennard-Jones e Coulomb. Os átomos de hidrogênio foram omitidos para melhor visualização.

Em analogia ao que ocorreu na **Eq.E**, apenas dois termos da **Eq.L** representam contribuições de Coulomb (**Glu138C** e **Pro225C**), enquanto quatro termos representam contribuições de Lennard-Jones (**Ile94LJ**, **Tyr181LJ**, **His235LJ** e **Asn137LJ**), indicando um maior peso das interações estéricas na relação estrutura-atividade.

Na **Eq.L** também foram selecionados termos que correspondem a resíduos descritos na literatura por fazer interações com os NNRTIs: **Tyr181**, **Pro225** e

His235. Destes, apenas **Pro225** não havia sido citado anteriormente, pois aparece somente nesta equação.

Comparando os termos que aparecem na **Eq.L (BD-III)** com os que aparecem nas **Eq.E (BD-I)** e **Eq.J (BD-II)**, observa-se que quatro (**Ile94LJ**, **Tyr181LJ**, **His235LJ**, **Asn137LJ**) são idênticos aos da **Eq.E** e um (**Glu224**) é idêntico ao da **Eq.J**. Apenas dois termos da **Eq.L** não aparecem nas equações anteriores, **Pro225C** e **Glu138C**, ambos representando interações de Coulomb. Desta forma, apenas estes dois termos serão discutidos, visto que os demais termos foram discutidos anteriormente nas equações **E** e **J**.

Apesar do termo **Pro225C** aparecer apenas na **Eq.L**, a informação contida nele não é exclusiva desta equação, visto que a análise da matriz de correlação cruzada entre os descritores (ver Tabela A do Anexo) mostra forte correlação ($r = 0,922$) entre os termos **Pro225C (Eq.L)** e **Pro226C (Eq.E)**, indicando que houve apenas uma substituição entre estes termos. Esta correlação elevada é devida a proximidade espacial entre os resíduos de aminoácidos correspondentes no NNBS e ao fato do mesmo tipo de aminoácido (**Pro**) estar envolvido. Portanto, este termo não será discutido porque ele é representado pelo termo **Pro226C** da **Eq.E**.

Será dado destaque ao termo **Glu138C (Eq.L)**, visto que a análise da matriz de correlação cruzada entre os descritores (ver na Tabela A do Anexo) não mostra correlação significativa deste termo com qualquer outro termo, indicando que a informação contida no termo **Glu138C** é exclusiva da **Eq.L**.

De fato, observando-se o gráfico da Figura 38, que mostra a média das energias de interação (kcal.mol^{-1}) dos compostos (**1-74**) com os resíduos selecionados na **Eq.L**, o termo **Glu138C** se destaca nitidamente dos demais por ser o mais desestabilizante (c.a. 48 kcal.mol^{-1}). Isso ocorre provavelmente devido a

proximidade entre o grupo carboxilato do resíduo **Glu138** e o átomo de oxigênio do grupo carbonila do ligante, gerando uma interação eletrostática repulsiva. Na estrutura de raios-X do complexo da RT com o MKC-442, observa-se uma interação estabilizante indireta entre o ligante e a enzima via a formação de ligação hidrogênio mediada por uma molécula de água que está localizada entre estes dois grupos. Esta estabilização não foi possível na simulação por DM porque todas as moléculas de água foram previamente removidas.

Poderia-se argumentar que a explicação dada para o comportamento anômalo do termo **Glu138C** não seria pertinente, porque para o termo **Glu224**, que também corresponde a um resíduo ácido, não se observa discrepância dos valores de média de energia em relação aos demais termos, onde a média de energia para o termo **Glu224** é próxima a zero. Entretanto, neste caso, o resíduo **Glu224** encontra-se na periferia do raio de 10 Å e não há interação direta ou indireta com o ligante.

Considerando-se que o termo **Glu138C** apresenta valor médio de energia positivo e coeficiente negativo na **Eq.L (BD-III)**, este termo contribui para diminuir a potência.

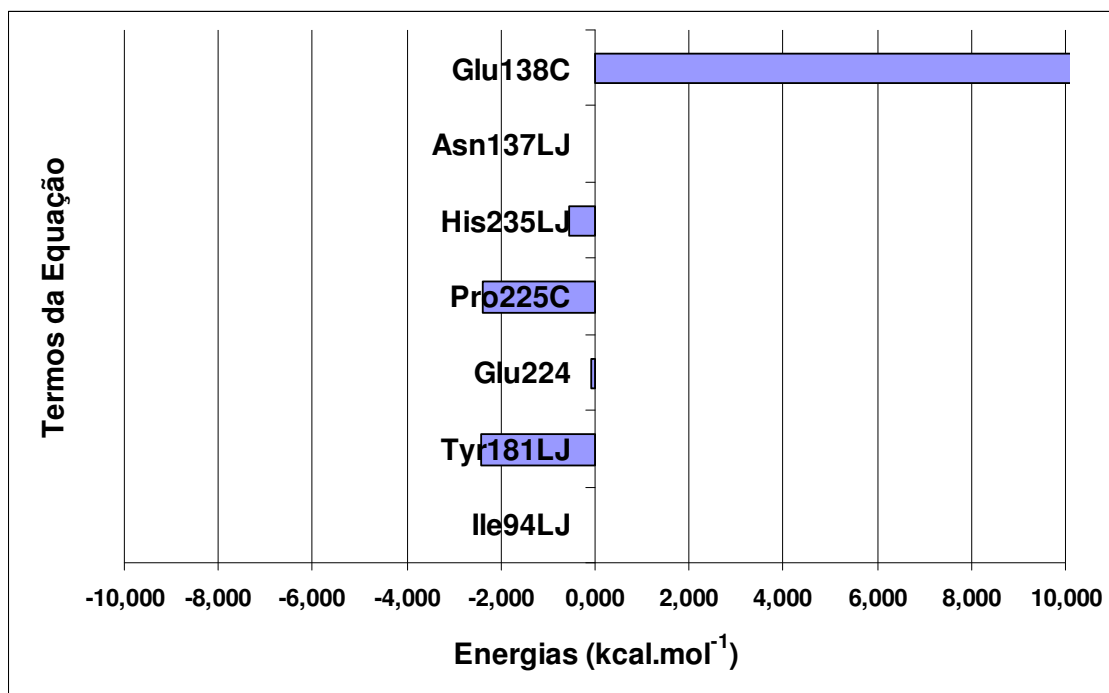


Figura 38. Médias das energias de interação (kcal.mol⁻¹) dos compostos **1-74** para cada um dos termos selecionados na **Eq.L (BD-III)**. O gráfico foi truncado em 10 kcal.mol⁻¹. A média de energia referente ao termo **Glu138C**= 47,676 kcal.mol⁻¹.

Em relação aos resíduos de aminoácidos na RT que sofrem freqüente mutação frente aos NNRTIs, dois aparecem na **Eq.L** relacionados aos termos **Tyr181LJ** e **Pro225C**. A mutação **Tyr181Cys** foi relatada anteriormente na discussão da **Eq.E (BD-I)**. No caso da mutação **Pro225His**, a cadeia lateral apolar do resíduo **Pro225**, que está voltada para o grupo alquila do ligante, é substituída pela cadeia lateral polar do resíduo **His225**, dificultando a interação ligante-enzima.

C) Análise dos Valores Residuais da Eq.L (BD-III)

Na Tabela 18 são mostrados os valores de pIC_{50} (M) observados (experimentais), os valores de pIC_{50} (M) preditos pela **Eq.L (BD-III)** e os respectivos valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**).

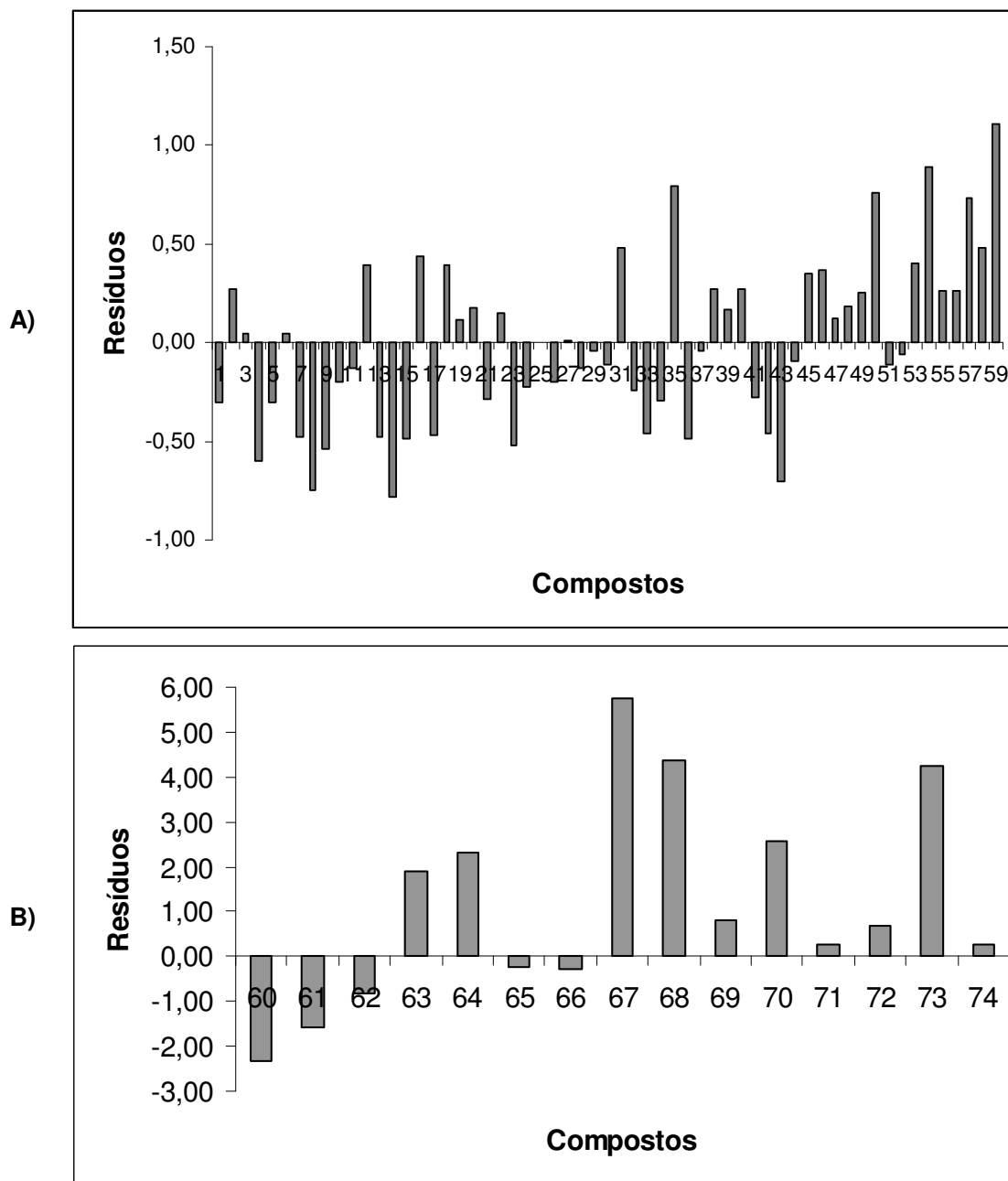
A análise da Tabela 18 para o conjunto de treinamento (**1-59**) mostra que 83 % dos compostos apresentam resíduos menores do que 0,50 (em valores modulares). Além disso, nenhum composto do conjunto de treinamento comporta-se como *outlier*, visto que nenhum apresenta valor residual maior do que o dobro do erro padrão da estimativa ($2 \times SEE = 2,36$). Isso mostra uma excelente capacidade preditiva interna do modelo. Na Figura 39.A, os resíduos para os compostos do conjunto de treinamento são mostrados na forma de gráfico de barras.

No caso dos compostos do conjunto de teste (**60-74**), 26 % apresentam resíduos menores do que 0,50, enquanto quatro compostos são considerados *outliers* (**67**, **68**, **70** e **73**). Na Figura 39.B, os resíduos para os compostos do conjunto de teste são mostrados na forma de gráfico de barras.

Os resíduos do conjunto de treinamento apresentam variações aleatórias em relação à variação da potência. Isso significa que o modelo não é tendencioso para um valor de atividade mais alto ou mais baixo. Porém, no caso do conjunto de teste, para dez dos quinze compostos, os resíduos calculados apresentam sinais positivos, ou seja, o modelo mostra-se tendencioso no momento da predição, onde os valores de atividade preditos são geralmente menores do que os observados.

Tabela 18. Valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**) de acordo com a **Eq.L (BD-III)**.

#	pIC_{50Obs}	pIC_{50Pred}	Res.	#	pIC_{50Obs}	pIC_{50Pred}	Res.
1	4,23	4,53	-0,30	38	6,10	5,83	0,27
2	4,31	4,04	0,27	39	6,10	5,93	0,17
3	4,35	4,30	0,05	40	6,10	5,83	0,27
4	4,59	5,19	-0,60	41	6,10	6,38	-0,28
5	4,77	5,07	-0,30	42	6,22	6,68	-0,46
6	4,79	4,75	0,04	43	6,22	6,93	-0,71
7	4,83	5,31	-0,48	44	6,22	6,31	-0,09
8	4,83	5,58	-0,75	45	6,30	5,95	0,35
9	5,02	5,56	-0,54	46	6,40	6,03	0,37
10	5,07	5,27	-0,20	47	6,70	6,57	0,13
11	5,09	5,22	-0,13	48	6,70	6,52	0,18
12	5,27	4,88	0,39	49	6,70	6,44	0,26
13	5,31	5,79	-0,48	50	6,92	6,16	0,76
14	5,31	6,10	-0,79	51	7,00	7,11	-0,11
15	5,32	5,81	-0,49	52	7,00	7,06	-0,06
16	5,34	4,90	0,44	53	7,05	6,65	0,40
17	5,37	5,84	-0,47	54	7,05	6,16	0,89
18	5,42	5,02	0,40	55	7,10	6,84	0,26
19	5,44	5,33	0,11	56	7,10	6,84	0,26
20	5,47	5,29	0,18	57	7,15	6,42	0,73
21	5,49	5,78	-0,29	58	7,30	6,82	0,48
22	5,52	5,37	0,15	59	7,52	6,41	1,11
23	5,52	6,04	-0,52	60	4,35	6,69	-2,34
24	5,52	5,75	-0,23	61	4,48	6,06	-1,58
25	5,54	5,54	-0,00	62	5,27	6,11	-0,84
26	5,55	5,75	-0,20	63	5,47	3,55	1,92
27	5,59	5,58	0,01	64	5,59	3,28	2,31
28	5,60	5,73	-0,13	65	5,60	5,84	-0,24
29	5,60	5,64	-0,04	66	5,62	5,91	-0,29
30	5,62	5,73	-0,11	67	5,74	-0,01	5,75
31	5,66	5,18	0,48	68	5,92	1,54	4,38
32	5,72	5,96	-0,24	69	6,22	5,39	0,83
33	5,80	6,26	-0,46	70	6,40	3,85	2,55
34	5,89	6,19	-0,30	71	6,70	6,42	0,28
35	5,94	5,15	0,79	72	7,05	6,35	0,70
36	5,94	6,43	-0,49	73	7,15	2,89	4,26
37	5,96	6,00	-0,04	74	7,30	7,03	0,27



D) Análise dos Compostos *Outliers* da Eq.L (BD-III)

Na **Eq. L (BD-III)**, como dito anteriormente, identificou-se quatro compostos *outliers*: **67**, **68**, **70** e **73**, todos do conjunto de teste, com resíduos muito altos e diferentes daqueles da **Eq.J (BD-II)**.

A Figura 40 mostra as estruturas, os valores de pIC_{50} observados e preditos e os respectivos valores residuais para estes *outliers*, onde se observa que todos apresentam potências preditas muito inferiores (valores residuais variando de 2,55 a 5,75) às observadas experimentalmente, indicando que o modelo subestimou a potência desses compostos.

No caso dos compostos **67** e **70**, eles também foram previstos como *outliers* de acordo com a **Eq.E (BD-I)** e também foram preditos com potências inferiores às observadas. Na **Eq.L**, estes compostos tiveram potências preditas ainda menores do que as observadas. O termo que contribui significativamente para a previsão com potências inferiores às observadas é **Glu138C**, com valores de energia de interação da ordem de 50 kcal.mol^{-1} . Assim, devido a esta baixa capacidade preditiva, o uso desta equação está comprometida.

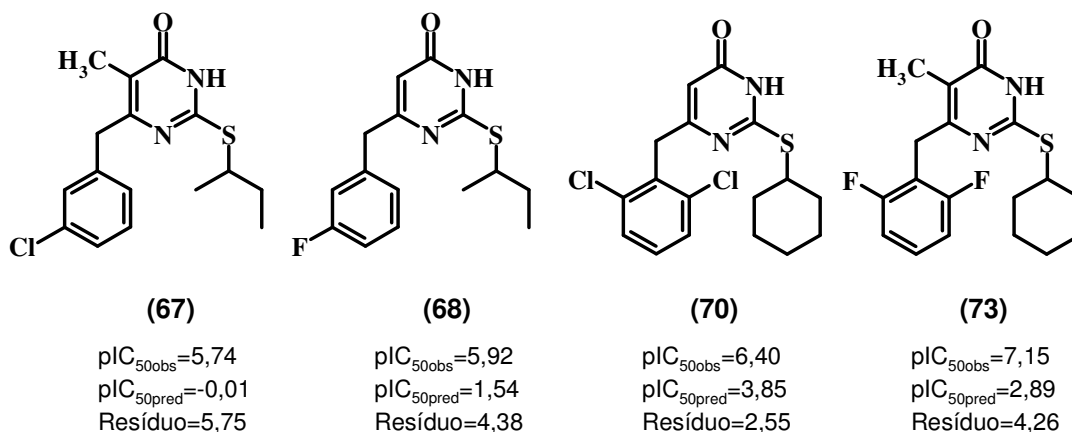


Figura 40. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.L (BD-III)**.

E) Análise do MKC-442 Segundo a Eq.L (BD-III)

De acordo com a **Eq.L (BD-III)**, o composto MKC-442 tem potência predita ($pIC_{50} = 9,08$ M) superior à experimental ($pIC_{50Pred} = 7,40$ M) e, conseqüentemente, valor residual (resíduo = 1,68) inferior ao dobro do SEE da **Eq.L**, indicando que o composto não é um *outlier*.

Este resultado não era esperado, visto que os compostos do próprio conjunto de teste, que apresentam maior similaridade estrutural com o conjunto de treinamento do que o MKC-442, foram previstos com resíduos muito altos.

F) Análise da Matriz de Correlação Cruzada dos Descritores da Eq.L (BD-III)

A análise da matriz de correlação cruzada entre os termos da **Eq.L** (Tabela 19) mostra que há forte correlação entre os descritores **Glu224** e **Pro225C** ($r = 0,955$). A correlação de **Glu224** com o termo **Glu224LJ** é maior e corresponde ao

valor de r acima, entretanto também há correlação grande com o termo **Glu224C** ($r = 0,893$). Isso pode acontecer provavelmente porque os resíduos de aminoácidos correspondentes são adjacentes. Este dado indica que a informação contida nestes termos pode ser redundante, o que não é óbvio, pois as características dos aminoácidos envolvidos são diferentes.

Tabela 19. Matriz de correlação cruzada entre os descritores da **Eq.L (BD-III)**.

	Ile94 LJ	Tyr181 LJ	Glu224 LJ+C	Pro225 C	His235 LJ	Asn137 LJ	Glu138 C
Ile94 LJ	1,000						
Tyr181 LJ	0,354	1,000					
Glu224 LJ+C	0,085	0,112	1,000				
Pro225 C	0,313	0,145	0,955	1,000			
His235 LJ	0,080	-0,030	-0,028	-0,084	1,000		
Asn137 LJ	0,435	0,036	0,059	0,028	0,135	1,000	
Glu138 C	-0,141	-0,202	-0,184	-0,179	-0,238	-0,238	1,000

4.2.3.4. Análise do Banco de Dados IV

O pré-tratamento do **BD-I** por exclusão dos descritores com variância de energia menor do que 0,0001 deu origem ao **BD-IV**. Essa exclusão foi feita no sentido de se diminuir o número de dados no **BD** (Albuquerque *et al.*, 1998). Esse tratamento é bastante usual em análises de QSAR com grande número de dados (Sodero, 2007; Pita, 2006; Cunha, 2006; Romeiro, 2002; Albuquerque, 1997).

Segundo este critério, os onze descritores a seguir foram excluídos: **Thr107LJ**, **Ile178LJ**, **Leu187LJ**, **Asp192LJ**, **His198LJ**, **Lys223LJ**, **Glu224LJ**, **Tyr232LJ**, **Thr240LJ**, **Asn137LJ** e **Thr139LJ**. Todos os descritores excluídos estão relacionados a resíduos que localizam-se fora do raio de 5 Å e correspondem exclusivamente a contribuições de Lennard-Jones. Portanto, parece razoável supor que estes descritores tenham sido excluídos porque a variação de energia estérica é praticamente nula. Além disso, segundo a literatura, esses resíduos não estão relacionados entre os resíduos que são essenciais para a interação com os NNRTIs, o que pode nos levar a supor uma menor importância destes termos na relação estrutura-atividade.

De fato, nenhum destes descritores foi selecionado nas equações discutidas anteriormente, com exceção do termo **Asn137LJ**, que aparece em duas equações: **Eq.E (BD-I)** e **Eq.L (BD-III)**. É oportuno lembrar que o termo **Asn137LJ** apresenta média de energia de interação com os ligantes próxima a zero (Figuras 29 e 38) e, nas duas equações (**E** e **L**), têm o maior valor modular de coeficiente, o que permite equilibrar o peso dele em relação aos demais termos nas respectivas equações.

Este resultado era realmente inesperado. Como justificar o fato intrigante da seleção deste termo, que aparentemente têm pouca importância matemática na correlação estrutura-atividade, em ao menos duas melhores equações? O que nos conduz a duas hipóteses: ou as equações não são de fato as melhores ou o termo tem maior importância do que aparenta. Ficamos com a segunda hipótese, lembrando que **Asn136**, **Asn137**, **Glu138** e **Thr139** são os únicos resíduos, dentro do raio de corte de 10 Å, que pertencem a subunidade p51, sendo responsáveis por parte das interações intercadeia (p66-p51) próximas ao NNBS.

Nas equações descritas até o momento, ao menos um termo está relacionado a um destes quatro resíduos da subunidade p51. De fato, dos dez termos da **Eq.E**, um, **Asn137LJ**, é referente ao resíduo **Asn137**; dos dez termos da **Eq.J**, um, **Thr139**, é referente ao resíduo Thr139, enquanto dos sete termos da **Eq.L**, dois (**Asn137LJ** e **Glu138C**) são referentes aos resíduos **Asn137** e **Glu138**. Este fato implica na importância da incorporação de ao menos um termo relacionado a um resíduo da subunidade p51 e, conseqüentemente, da interação intercadeia (próxima ao NNBS) nos modelos (equações) que melhor descrevem a relação estrutural-atividade para esta classe de compostos.

Outros descritores, também excluídos no pré-tratamento, **Leu187LJ**, **Glu224LJ** e **Thr139LJ**, apesar de não terem sido selecionados explicitamente nas equações discutidas anteriormente, estão incorporados nos termos **Leu187 (Eq.J)**, **Glu224 (Eq.J e Eq.L)** e **Thr139 (Eq.J)**, respectivamente, visto que estes termos representam o somatório das contribuições de Lennard-Jones e Coulomb.

A) Análise dos Índices Estatísticos do BD-IV

A Tabela 20 resume os índices estatísticos referentes às cinco melhores equações (**O-S**), contendo de sete a doze termos, geradas a partir do **BD-IV**. Todas as equações foram geradas a partir do maior número de operações de *crossover* (50.000) e não foram obtidas equações com dez termos a partir deste banco de dados.

Em geral, as equações obtidas a partir do **BD-IV** apresentam valores estatísticos similares aos das equações obtidas a partir do **BD-I**, do qual o **BD-IV** é proveniente, com exceção dos valores de SEE e do número de compostos *outliers* que são, em média, maiores nas equações do **BD-IV** (Tabela 20).

Tabela 20. Resultados estatísticos dos cinco melhores modelos (**O-S**) de QSAR-3D dependente do receptor obtidos a partir do Banco de Dados IV (**BD-IV**).

BD-IV	N.T.E.^a	q²_{ajus.}^b	q²^c	SE_{CV}^d	PC^e	r²^f	SEE^g	Outliers^h
Eq. O	7	0,568	0,613	0,498	4	0,691	0,710	5
Eq. P	8	0,572	0,624	0,485	5	0,721	0,750	6
Eq. Q	9	0,616	0,669	0,458	6	0,764	0,700	5
Eq. R	11	0,549	0,627	0,488	6	0,777	0,590	3
Eq.S	12	0,595	0,672	0,457	6	0,793	0,480	2

a) Número de termos da equação (N.T.E). b) q² ajustado ao número de termos da equação. c) r² da validação cruzada (q²). d) Desvio padrão da validação cruzada (SE_{CV}). e) Número ótimo de componentes principais (PC). f) Coeficiente de correlação linear quadrático (r²). g) Desvio padrão da estimativa (SEE). h) Número de compostos *outliers* no conjunto de teste.

Assim, considerando apenas as equações do **BD-IV**, as equações **Q** (com 9 termos; q²_{ajus} = 0,616; SE_{CV} = 0,458 e 5 *outliers*) e **S** (com 12 termos; q²_{ajus} = 0,595; SE_{CV} = 0,457 e 2 *outliers*) se destacam das demais por apresentarem maiores valores de q² ajustado e menores valores de SE_{CV}. Como as **Eq.Q** e **Eq.S** apresentam resultados semelhantes, mas a **Eq.S** apresenta menor valor de SEE (Tabela 20) e menor número de compostos *outliers*, ela deveria ter sido escolhida como a melhor equação do **BD-IV**, devido a sua maior capacidade preditiva.

Entretanto, a **Eq.S** possui 12 termos, valor que está no limite máximo do número de descritores permitido (por equação) para evitar super-ajuste, visto que o conjunto de treinamento é formado por 59 compostos. Além do mais, de acordo com Livingstone (1995), a probabilidade de encontrar uma correlação ao acaso, ou seja, não uma correlação, e sim uma coincidência, aumenta à medida que o número de descritores que aparecem nas equações de QSAR aumenta. Aplicando este conceito e o princípio da parcimônia de Occam (Kubinyi, 1997), a equação considerada para análise dos descritores será a **Eq.Q**, que contém apenas 9 termos.

B) Análise da Melhor Equação do BD-IV (Eq.Q)

Na **Eq.Q (BD-IV)**, descrita a seguir, cada um dos nove termos está representado pelo código de três letras do respectivo aminoácido, seguido do código **LJ** (Lennard-Jones) ou **C** (Coulomb), que indica se a interação refere-se a contribuição estérica ou eletrostática, respectivamente. A Figura 41 mostra a representação gráfica 3D da **Eq.Q** com o composto NH-DABO **59**.

Diferente do que ocorreu nas equações **Eq.E (BD-I)** e **Eq.L (BD-III)**, nesta Equação os termos referentes às interações estéricas (Lennard-Jones) não predominam, visto que cinco termos representam contribuições de Coulomb (**Gly99C**, **Leu187C**, **Tyr188C**, **Pro225C** e **Pro226C**), enquanto que quatro termos representam contribuições de Lennard-Jones (**Val179LJ**, **Lys103LJ**, **Asn136LJ** e **Phe227LJ**), indicando um peso mais equilibrado entre as interações estéricas e eletrostáticas na relação estrutura-atividade.

Também foram selecionados termos que correspondem a resíduos descritos na literatura por fazer interações com os NNRTIs: **Lys103** e **Tyr188**. Ambos estão contidos no raio de 5 Å e apenas **Lys103** não havia sido citado anteriormente, pois aparece somente nesta equação.

Eq.Q
(BD-IV)

$$\begin{aligned}
 \text{pIC}_{50} = & 7,706 + 1,162 \text{ Leu187C} - 0,050 \text{ Tyr188C} \\
 & + 0,403 \text{ Pro225C} + 0,151 \text{ Gly99C} - 0,238 \text{ Pro226C} - 0,282 \text{ Val179LJ} \\
 & + 0,124 \text{ Lys103LJ} - 9,989 \text{ Asn136LJ} - 0,074 \text{ Phe227LJ}
 \end{aligned}$$

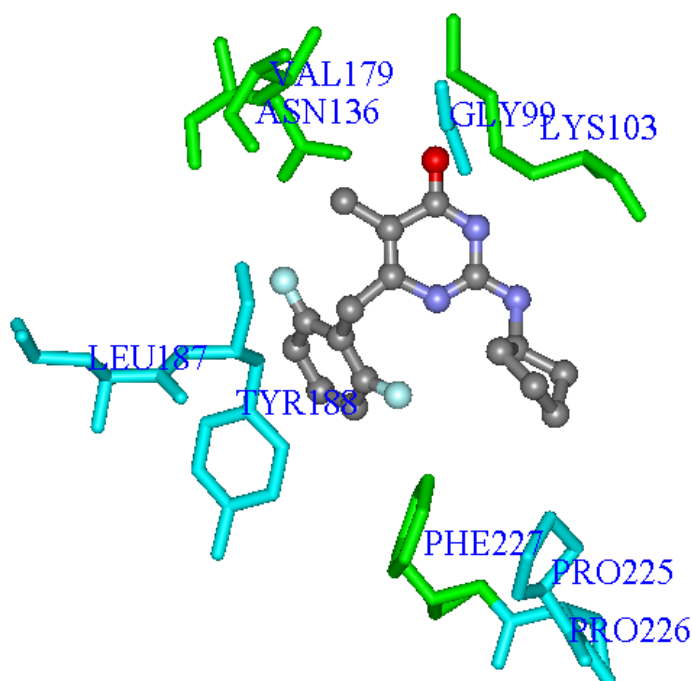


Figura 41. Representação gráfica 3D da **Eq.Q (BD-IV)** com o composto NH-DABO **59** (em modelo bastão-e-bola e colorido por elemento). Os resíduos de aminoácidos (em modelo bastão) que representam contribuições de Lennard-Jones (**Lys103**, **Val179**, **Phe227** e **Asn136**) estão coloridos em verde e os que representam contribuições de Coulomb (**Gly99**, **Leu187**, **Tyr188**, **Pro225** e **Pro226**), em azul claro. Os átomos de hidrogênio foram omitidos para melhor visualização.

Comparando os termos que aparecem na equação **Q (BD-IV)** com os que aparecem nas equações **E (BD-I)** e **L (BD-III)**, observa-se que apenas dois são repetidos: **Pro226C (Eq.E)** e **Pro225C (Eq.L)**. Além disso, como discutido anteriormente, a análise da matriz de correlação cruzada entre os descritores (ver

Tabela A do Anexo) mostra forte correlação ($r = 0,922$) entre estes dois termos, indicando que eles contribuem com o mesmo tipo de informação para a relação estrutura-atividade. Desta forma, serão discutidos somente os termos da **Eq.Q** que não se repetem nas equações anteriores, *i.e.* **Gly99C**, **Lys103LJ**, **Val179LJ**, **Leu187C**, **Tyr188C**, **Phe227LJ** e **Asn136LJ**.

Apesar de serem exclusivos da **Eq.Q**, os termos **Gly99C**, **Leu187C**, **Tyr188C** e **Phe227LJ** estão contidos, em diferentes graus, nos respectivos termos **Gly99**, **Leu187**, **Tyr188** e **Phe227** da **Eq.J (BD-II)**, como pode ser observado na matriz de correlação cruzada entre os descritores (ver Tabela A do Anexo): **Gly99C** x **Gly99** ($r = 0,994$), **Leu187C** x **Leu187** ($r = 1,0$), **Tyr188C** x **Tyr188** ($r = 0,832$) e **Phe227LJ** x **Phe227** ($r = 0,668$). Isto é possível porque os descritores da **Eq.J** são provenientes do **BD-II**, onde a energia de interação corresponde à soma das contribuições de Lennard-Jones e Coulomb para um mesmo resíduo de aminoácido.

Os resíduos relacionados aos termos **Lys103LJ**, **Asn136LJ** e **Val179LJ** da **Eq.Q** não haviam aparecido em qualquer um dos termos das equações anteriores. De acordo com a matriz de correlação cruzada entre os descritores (ver Tabela A do Anexo), o termo **Lys103LJ** apresenta maior correlação com os termos **Ser191C** ($r = 0,694$) e **Ser191** ($r = 0,688$) das **Eq.E** e **Eq.J**, respectivamente. Na estrutura 3D da RT em complexo com o inibidor MKC-442, o resíduo **Lys103** encontra-se em contato próximo e acomodado justamente entre o ligante e o resíduo **Ser191**. Portanto, a correlação observada é devido a proximidade espacial entre estes resíduos. De modo análogo, o termo **Asn136LJ** apresenta correlação forte ($r = 0,830$) com o termo **Asn137LJ** da **Eq.E**, devido a proximidade espacial entre os resíduos e ao mesmo tipo de aminoácido estar envolvido. No caso do termo **Val179LJ** não se

observa correlação significativa com qualquer outro termo, indicando que ele é um termo característico da **Eq.Q**.

O gráfico da Figura 42 mostra a média dos valores de energia de interação (kcal.mol^{-1}) dos compostos (**1-74**) com os resíduos selecionados na **Eq.Q**. Em analogia ao observado nas **Eq.E** e **Eq.L**, a maioria dos termos (exceto **Tyr188C**) têm médias de energia negativas (**Lys103LJ**, **Val179LJ**, **Leu187C**, **Pro225C**, **Pro226C** e **Phe227LJ**) ou próximas a zero (**Gly99C** e **Asn136LJ**). O valor de energia de maior magnitude negativa (*i.e.* mais estabilizante) é referente ao termo **Lys103LJ**, com média de energia próxima a $-6,0 \text{ kcal.mol}^{-1}$, cujo resíduo de aminoácido, como dito anteriormente, está em contato próximo com o ligante MKC-442.

É interessante notar que as contribuições de Lennard-Jones e Coulomb, referentes ao resíduo de aminoácido **Tyr188**, são opostas, visto que na **Eq.E**, o termo **Tyr188LJ** apresenta média de energia negativa (c.a. $-4,0 \text{ kcal.mol}^{-1}$), enquanto que na **Eq.Q**, o termo **Tyr188C** apresenta média de energia positiva (c.a. $6,5 \text{ kcal.mol}^{-1}$). Isso significa que, para este resíduo em particular, as interações estéricas (van der Waals) são predominantemente atrativas, enquanto que as interações eletrostáticas são predominantemente repulsivas. É oportuno lembrar que o anel aromático de **Tyr188** faz interação do tipo π - π -*stacking* com o anel aromático dos inibidores (Figura 23).

Considerando os oito termos com valores de energia negativos, os termos **Val179LJ**, **Pro226C**, **Phe227LJ** e **Asn136LJ**, que também apresentam coeficiente de sinal negativo na **Eq.Q**, contribuem para aumentar a potência, enquanto todos os demais termos (incluindo **Tyr188C**) contribuem para diminuir a potência.

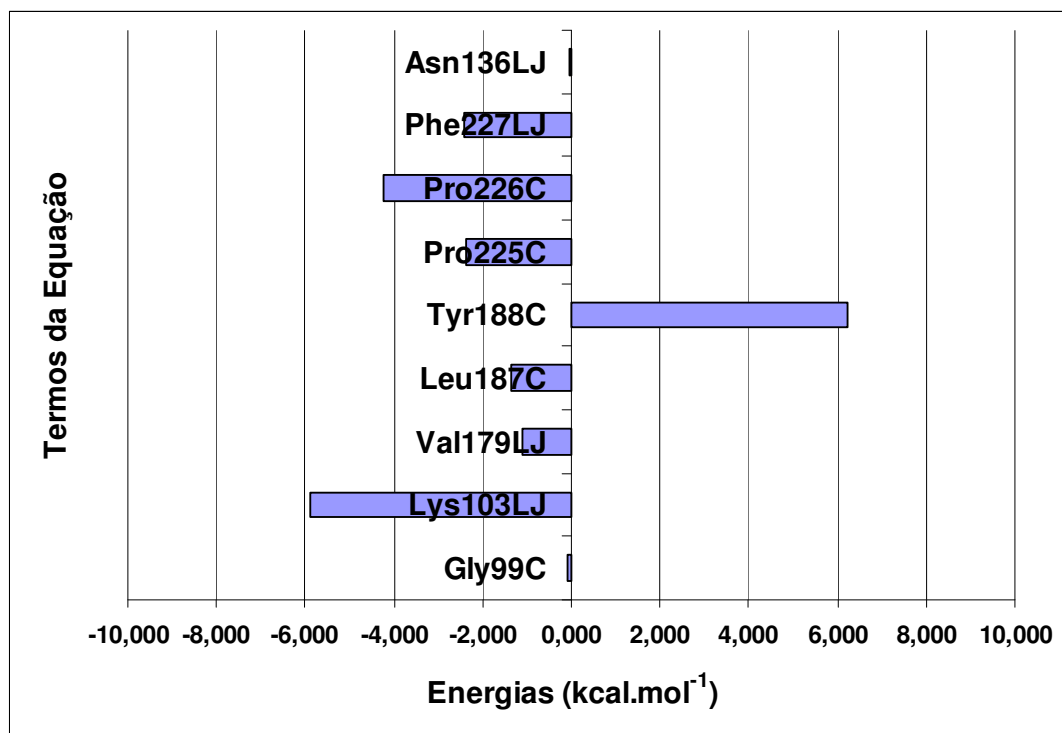


Figura 42. Médias das energias de interação (kcal.mol^{-1}) dos compostos **1-74** para cada um dos termos selecionados na **Eq.Q** (BD-IV).

Em relação aos resíduos de aminoácidos na RT que sofrem freqüente mutação frente aos NNRTIs, quatro aparecem na **Eq.Q** relacionados aos termos **Lys103LJ**, **Tyr188C**, **Pro225C** e **Phe227LJ**. As mutações **Tyr188Leu**, **Pro225His** e **Phe227Leu** foram discutidas anteriormente, pois os resíduos de aminoácidos correspondentes aparecem relacionados a termos das equações **E** (**Tyr188LJ**), **J** (**Tyr188** e **Phe227**) e **L** (**Pro225C**). O resíduo **Lys103** é um dos que mais sofre mutação sendo conhecidas, além da mutação simples **Lys103Asn**, quatro mutações duplas que combinam esta com mais uma das seguintes mutações: **Tyr181Cys**, **Val108Ile**, **Pro225His** e **Leu100Ile**. Nessa mutação há a substituição de um aminoácido com característica polar e básica (**Lys**) por um com característica polar, porém neutra (**Asn**). Em termos de interações com fármacos, a troca desprivilegiaria uma interação iônica.

C) Análise dos Valores Residuais da Eq.Q (BD-IV)

Na Tabela 21 são mostrados os valores de pIC_{50} (M) observados (experimentais), os valores de pIC_{50} (M) preditos pela **Eq.Q (BD-IV)** e os respectivos valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (**1-59**) e de teste (**60-74**).

A análise da Tabela 21 para o conjunto de treinamento (**1-59**) mostra que 74% dos compostos apresentam resíduos menores do que 0,50 (em valores modulares). Além disso, nenhum composto do conjunto de treinamento comporta-se como *outlier*, visto que nenhum apresenta valor residual maior do que o dobro do erro padrão da estimativa ($2 \times SEE = 1,400$). Isso mostra uma excelente capacidade preditiva interna do modelo. Na Figura 43.A, os resíduos para os compostos do conjunto de treinamento são mostrados na forma de gráfico de barras.

Tabela 21. Valores de pIC_{50} observados, e preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os conjuntos de treinamento (1-59) e de teste (60-74) de acordo com a Eq.Q (BD-IV).

#	pIC_{50Obs}	pIC_{50Pred}	Res.	#	pIC_{50Obs}	pIC_{50Pred}	Res.
1	4,23	4,48	0,25	38	6,10	5,91	-0,19
2	4,31	4,41	0,10	39	6,10	5,41	-0,69
3	4,35	4,13	-0,22	40	6,10	6,02	-0,08
4	4,59	5,28	0,69	41	6,10	6,52	0,42
5	4,77	4,59	-0,18	42	6,22	6,24	0,02
6	4,79	4,84	0,05	43	6,22	6,85	0,63
7	4,83	5,51	0,68	44	6,22	6,29	0,07
8	4,83	5,66	0,83	45	6,30	5,72	-0,58
9	5,02	5,55	0,53	46	6,40	6,40	0,00
10	5,07	5,20	0,13	47	6,70	6,35	-0,35
11	5,09	5,23	0,14	48	6,70	6,45	-0,25
12	5,27	5,31	0,04	49	6,70	6,60	-0,10
13	5,31	5,85	0,54	50	6,92	6,07	-0,85
14	5,31	5,26	-0,05	51	7,00	6,97	-0,03
15	5,32	4,89	-0,43	52	7,00	6,76	-0,24
16	5,34	5,29	-0,05	53	7,05	6,49	-0,56
17	5,37	6,24	0,87	54	7,05	6,78	-0,27
18	5,42	5,51	0,09	55	7,10	6,64	-0,46
19	5,44	4,71	-0,73	56	7,10	6,87	-0,23
20	5,47	5,55	0,08	57	7,15	7,05	-0,10
21	5,49	5,16	-0,33	58	7,30	7,17	-0,13
22	5,52	5,31	-0,21	59	7,52	6,85	-0,67
23	5,52	5,92	0,40	60	4,35	6,91	-2,56
24	5,52	5,74	0,22	61	4,48	4,52	-0,04
25	5,54	5,92	0,38	62	5,27	7,14	-0,47
26	5,55	5,58	0,03	63	5,47	5,74	-1,67
27	5,59	5,67	0,08	64	5,59	6,42	-0,83
28	5,60	5,97	0,37	65	5,60	3,55	2,05
29	5,60	5,75	0,15	66	5,62	8,74	-3,12
30	5,62	5,47	-0,15	67	5,74	5,16	0,58
31	5,66	5,93	0,27	68	5,92	8,74	-2,82
32	5,72	5,22	-0,50	69	6,22	5,60	-0,18
33	5,80	6,40	0,60	70	6,40	6,40	0,00
34	5,89	5,94	0,05	71	6,70	6,47	0,23
35	5,94	5,79	-0,15	72	7,05	5,90	1,15
36	5,94	6,17	0,23	73	7,15	5,53	1,62
37	5,96	5,76	-0,20	74	7,30	6,87	0,43

No caso dos compostos do conjunto de teste (**60-74**), 40% apresentam resíduos menores do que 0,50 (Tabela 21), enquanto seis compostos são considerados *outliers* (**60**, **63**, **65**, **66**, **68** e **73**). Na Figura 43.B, os resíduos para os compostos do conjunto de teste são mostrados na forma de gráfico de barras.

Além disso, os resíduos de ambos os conjuntos de treinamento e de teste apresentam variações aleatórias em relação à variação da potência. Isso significa que o modelo não é tendencioso para um valor de atividade mais alto ou mais baixo.

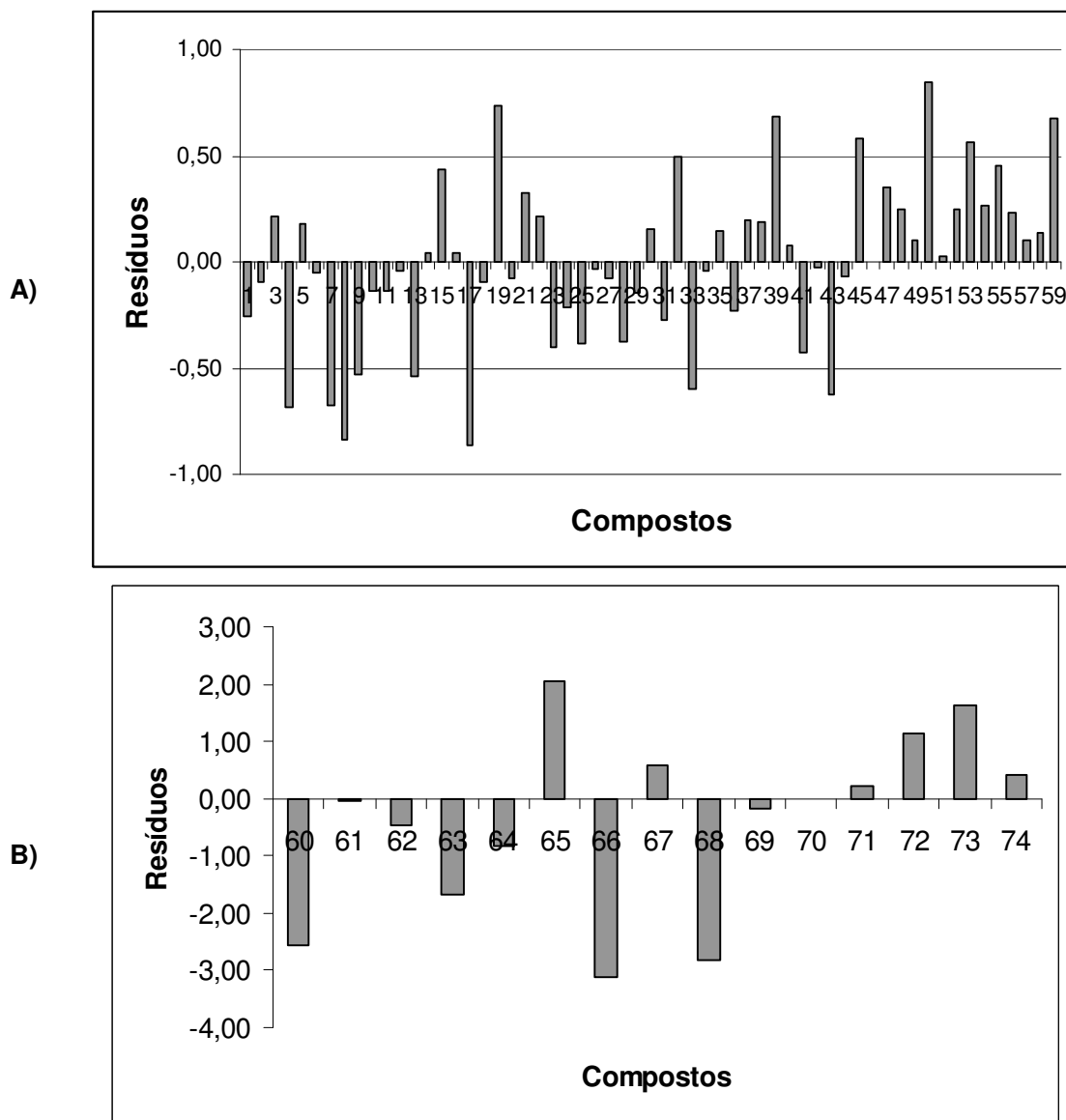


Figura 43. Gráfico de barras dos resíduos ($pIC_{50obs} - pIC_{50Pred}$) dos compostos dos conjuntos de (A) treinamento (1-59) e de (B) teste (60-74) de acordo com a **Eq.Q (BD-IV)**.

D) Análise dos Compostos *Outliers* da Eq.Q (BD-IV)

Na **Eq.Q**, como dito anteriormente, identificou-se seis compostos *outliers*: **60**, **63**, **65**, **66**, **68** e **73**, todos do conjunto de teste, cujas estruturas, valores de atividade biológica (plC_{50}) observados e preditos e respectivos valores residuais, encontram-se na Figura 44, onde se observa que quatro (**60**, **63**, **66** e **68**) apresentam potências preditas superiores às experimentais, enquanto que dois (**65** e **73**) apresentam potências preditas inferiores às experimentais.

A **Eq.Q** foi a que forneceu o maior número de compostos *outliers* comparando com as **Eq.E** (três *outliers*), **Eq.J** (três *outliers*) e **Eq.L** (quatro *outliers*) provenientes a partir dos banco de dados I a III, respectivamente. Dos seis compostos *outliers* da **Eq.Q**, quatro também são *outliers* nas equações anteriores: **60** e **66** aparecem na **Eq.J** (ambos com valores residuais menores), enquanto que **68** e **73** aparecem na **Eq.L** (ambos com valores residuais maiores, porém o resíduo de **68** é positivo). Desta forma, os únicos *outliers* exclusivos desta equação são **63** e **65**. Entretanto, devido ao número excessivo de *outliers* observados nesta equação, a racionalização deste comportamento não será explorada.

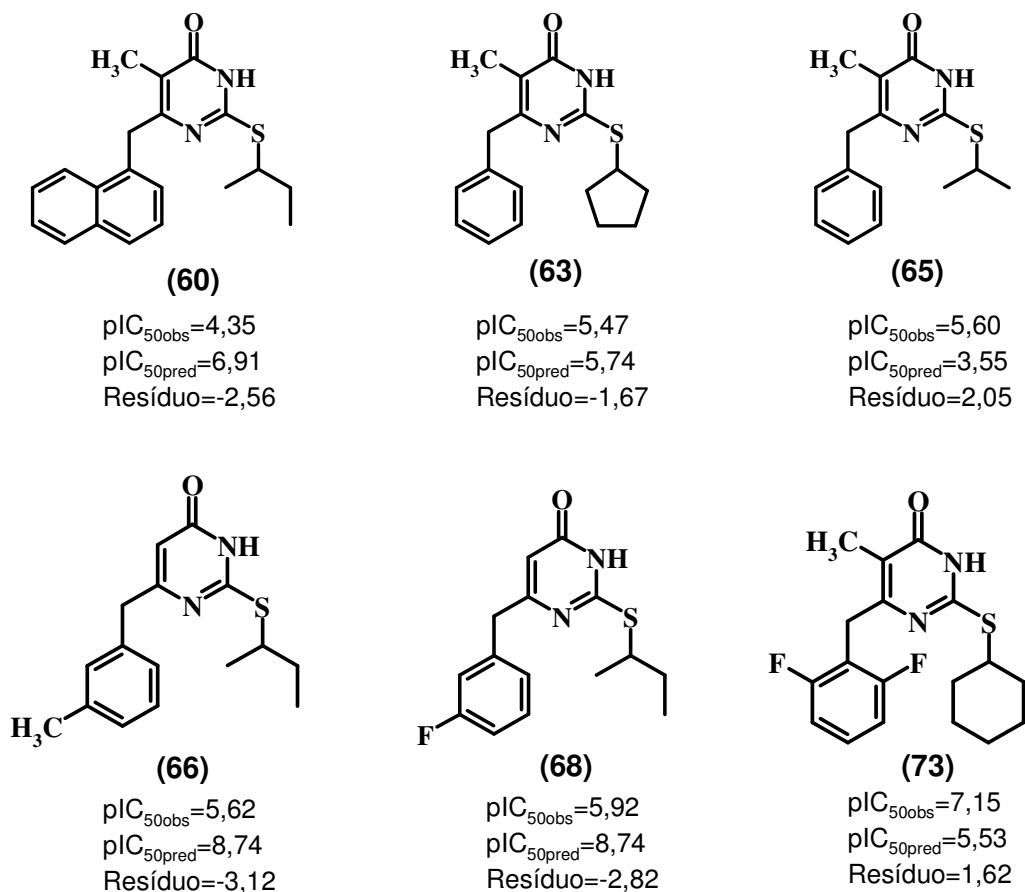


Figura 44. Estruturas químicas, valores de pIC_{50} (M) observados, valores de pIC_{50} (M) preditos e valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) dos compostos *outliers* da **Eq.Q (BD-IV)**.

E) Análise do MKC-442 Segundo a Eq.Q (BD-IV)

De acordo com a **Eq.Q (BD-IV)**, o composto MKC-442 têm potência predita ($pIC_{50} = 10,92$ M) superior à experimental ($pIC_{50Pred} = 7,40$ M) e, conseqüentemente, valor residual (resíduo = 3,52) superior ao dobro do SEE da **Eq.Q**, indicando que o composto é um *outlier*. Este resultado era esperado, visto que para esta equação foi observado o maior número de *outliers*.

F) Análise da Matriz de Correlação Cruzada dos Descritores da Eq.Q (BD-IV)

A análise da matriz de correlação cruzada entre os termos da **Eq.Q** (Tabela 22) mostra que há forte correlação entre os descritores **Pro225C** e **Pro226C** ($r = 0,922$), como citado anteriormente, indicando que a informação contida nestes termos é redundante.

Tabela 22. Matriz de correlação cruzada entre os descritores da **Eq.Q (BD-IV)**.

	Asn136 LJ	Leu187 C	Val179 LJ	Lys103 LJ	Pro225 C	Gly99 C	Pro226 C	Phe227 LJ	Tyr188 C
Asn136 LJ	1,000								
Leu187 C	0,063	1,000							
Val179 LJ	0,220	-0,133	1,000						
Lys103 LJ	-0,238	-0,046	0,005	1,000					
Pro225 C	-0,056	0,119	-0,118	0,333	1,000				
Gly99 C	0,176	-0,233	-0,046	-0,283	-0,280	1,000			
Pro226 C	-0,124	0,190	-0,145	0,301	0,922	-0,351	1,000		
Phe227 LJ	-0,372	0,042	-0,165	0,494	0,375	-0,231	0,335	1,000	
Tyr188 C	0,150	0,033	0,028	0,215	0,292	-0,007	0,252	0,165	1,000

4.2.4. Análise Global das Melhores Equações dos Bancos de Dados I a IV

A Tabela 23 resume as melhores equações dos Bancos de Dados I a IV (equações **E**, **J**, **L** e **Q**), enquanto que os correspondentes índices estatísticos e compostos *outliers* estão resumidos nas Tabelas 24 e 25, respectivamente.

Das quatro equações (Tabela 23), a **Eq.E (BD-I)** é a melhor, não somente em relação a maior capacidade explicativa (maior valor de r^2 e menor valor de SEE), mas também na maior capacidade preditiva, tanto interna (maior valor de q^2 e menor valor de SE_{CV}), onde os compostos do conjunto de treinamento apresentam os menores valores residuais, quanto externa, onde os compostos do conjunto de teste também apresentam os menores valores residuais (Tabela 25).

A segunda melhor equação, **Eq.J (BD-II)**, apresenta o mesmo número de termos (dez) e o mesmo número de compostos *outliers* (três) do que a **Eq.L**, porém, possui menor capacidade explicativa e preditiva (Tabela 24). As equações **L (BD-III)** e **Q (BD-IV)**, apesar de serem mais “econômicas”, porque apresentam apenas sete e nove termos, respectivamente, apresentam maior número de compostos *outliers*, *i.e.* quatro e cinco, respectivamente. Um agravante, no caso da **Eq.L**, são os valores residuais dos quatro compostos *outliers* do conjunto de teste, que são excessivamente elevados, tornando esta equação a pior de todas. Assim, podemos classificar as **Eq.E** e **Eq.J** como as duas melhores equações e as **Eq.L** e **Eq.Q** como as duas piores.

Em relação ao peso das contribuições de Lennard-Jones (**LJ**) e Coulomb (**C**) na relação estrutura-atividade (Tabela 23), pode-se observar uma maior contribuição dos termos de Lennard-Jones, tanto na melhor (**Eq.E**, sete termos de **LJ** e três termos de **C**) quanto na pior equação (**Eq.L**, quatro termos de **LJ** e dois termos de **C**), enquanto que, na **Eq.Q**, houve um ligeiro predomínio dos termos de Coulomb

(quatro termos de **LJ** e cinco termos de **C**). Para a **Eq.J**, este tipo de análise não pode ser realizada, porque as contribuições de **LJ** e **C** não estão individualizadas, lembrando que cada termo, neste caso, corresponde ao somatório das energias de interação estérica e eletrostática.

Considerando os 53 resíduos de aminoácidos contidos no recorte de 10 Å, cerca de 43% (23 resíduos: **Ile94**, **Pro97**, **Gly99**, **Lys101**, **Lys103**, **Val179**, **Tyr181**, **Gln182**, **Tyr183**, **Leu187**, **Tyr188**, **Ser191**, **Glu224**, **Pro225**, **Pro226**, **Phe227**, **Trp229**, **His235**, **Asp237**, **Asn136**, **Asn137**, **Glu138** e **Thr139**) aparecem com maior frequência nas equações **E**, **J**, **L** e **Q** (Tabela 23). Isto indica que, independente da contribuição do termo relacionado a estes aminoácidos ser estérica, eletrostática ou combinada, estes 23 resíduos têm maior importância na relação estrutura-atividade do que os demais.

Destes 23 resíduos, destacamos o resíduo **Tyr188** com três ocorrências (**Eq.E**, **Eq.J** e **Eq.Q**). Em seguida, com duas ocorrências, destacamos os resíduos: **Ile94** (**Eq.E** e **Eq.L**), **Gly99** (**Eq.J** e **Eq.Q**), **Tyr181** (**Eq.E** e **Eq.L**), **Leu187** (**Eq.J** e **Eq.Q**), **Ser191** (**Eq.E** e **Eq.J**), **Glu224** (**Eq.J** e **Eq.L**), **Pro225** (**Eq.L** e **Eq.Q**), **Pro226** (**Eq.E** e **Eq.Q**), **Phe227** (**Eq.J** e **Eq.Q**), **His235** (**Eq.E** e **Eq.L**) e **Asn137** (**Eq.E** e **Eq.L**).

Nas quatro equações (Tabela 23), ao menos um termo está relacionado a um dos quatro resíduos de aminoácidos da subunidade p51: **Asn136** (termo **Asn136LJ** da **Eq.Q**), **Asn137** (termo **Asn137LJ** das **Eq.E** e **Eq.L**), **Glu138** (termo **Glu138C** da **Eq.L**) e **Thr139** (termo **Thr139** da **Eq.J**). Como discutido anteriormente, este fato implica na importância da interação intercadeia (p66-p51) nas equações que melhor descrevem a relação estrutura-atividade para esta classe de compostos.

Adicionalmente, a Tabela 26 mostra a matriz de correlação cruzada entre os valores residuais ($pIC_{50Obs} - pIC_{50Pred}$) calculados para os compostos do conjunto de treinamento das quatro equações (**E**, **J**, **L** e **Q**), com o objetivo de verificar o grau de correlação entre os modelos. Neste tipo de matriz, pares de modelos equivalentes têm resíduos de atividade correlacionados (r próximos ou iguais a 1) e podem representar o conjunto de treinamento de forma similar. Por outro lado, pares de modelos distintos apresentam resíduos de atividade não relacionados entre si ($r < 0,7$) (Hopfinger, 1997). Analisando os dados da Tabela 26, observa-se que, segundo este critério, os modelos não estão fortemente correlacionados, indicando que são distintos, visto que a maior correlação ($r = 0,559$) ocorre entre as equações **E** e **J**, que são, justamente, os dois melhores modelos, enquanto que o modelo **Q** é o que apresenta maior divergência dos demais.

Tabela 23. Melhores equações dos Bancos de Dados I a IV.

Eq.E (BD-I)	$pIC_{50} = 4,853 + 22,417 \text{ Ile94LJ} + 0,231 \text{ Pro97LJ} - 0,153 \text{ Lys101LJ} - 0,110 \text{ Tyr181LJ} - 0,791 \text{ Gln182C} - 0,122 \text{ Tyr188LJ} + 0,323 \text{ Ser191C} + 0,043 \text{ Pro226C} + 0,087 \text{ His235LJ} - 56,813 \text{ Asn137LJ}$
Eq.J (BD-II)	$pIC_{50} = 6,802 + 0,101 \text{ Gly99} - 0,244 \text{ Tyr183} + 1,202 \text{ Leu187} - 0,059 \text{ Tyr188} + 0,360 \text{ Ser191} + 0,822 \text{ Glu224} - 0,028 \text{ Phe227} - 0,026 \text{ Trp229} + 0,061 \text{ Asp237} + 1,437 \text{ Thr139}$
Eq.L (BD-III)	$pIC_{50} = 6,257 + 15,851 \text{ Ile94LJ} - 0,118 \text{ Tyr181LJ} + 0,101 \text{ Pro225C} + 3,525 \text{ Glu224} + 0,062 \text{ His235LJ} - 52,568 \text{ Asn137LJ} - 0,008 \text{ Glu138C}$
Eq.Q (BD-IV)	$pIC_{50} = 7,706 + 0,151 \text{ Gly99C} + 1,162 \text{ Leu187C} - 0,050 \text{ Tyr188C} + 0,403 \text{ Pro225C} - 0,238 \text{ Pro226C} - 0,282 \text{ Val179LJ} + 0,124 \text{ Lys103LJ} - 9,989 \text{ Asn136LJ} - 0,074 \text{ Phe227LJ}$

Tabela 24. Resultados estatísticos das melhores equações dos Bancos de Dados I a IV.

Equação (BD)	N.T.E.^a	q²_{ajus.}^b	q²^c	SE_{cv}^d	PC^e	r²^f	SEE^g	Outliers^h
Eq.E (BD-I)	10	0,660	0,713	0,420	4	0,822	0,500	3
Eq.J (BD-II)	10	0,606	0,667	0,460	6	0,766	0,600	3
Eq.L (BD-III)	7	0,594	0,636	0,480	3	0,723	1,180	4
Eq.Q (BD-IV)	9	0,616	0,669	0,458	6	0,764	0,700	5

a) Número de termos da equação (N.T.E.). b) q² ajustado ao número de termos da equação. c) r² da validação cruzada (q²). d) Desvio padrão da validação cruzada (SE_{cv}). e) Número ótimo de componentes principais (PC). f) Coeficiente de correlação linear quadrático (r²). g) Desvio padrão da estimativa (SEE). h) Número de compostos *outliers* no conjunto de teste.

Tabela 25. Compostos *outliers* (e respectivos valores residuais) identificados nas melhores equações dos Bancos de Dados I a IV.

Equação (BD)	Número do composto <i>outlier</i> (valor residual)
Eq.E (BD-I)	<u>67</u> (1,06), <u>69</u> (-2,30) e <u>70</u> (1,12)
Eq.J (BD-II)	<u>60</u> (-1,22), <u>62</u> (-1,67) e <u>66</u> (-2,57)
Eq.L (BD-III)	<u>67</u> (5,75), <u>68</u> (4,38), <u>70</u> (2,55) e <u>73</u> (4,26)
Eq.Q (BD-IV)	<u>60</u> (-2,56), <u>63</u> (-1,67), <u>65</u> (2,05), <u>66</u> (-3,12), <u>68</u> (-2,82) e <u>73</u> (1,62)

Tabela 26. Matriz de correlação cruzada entre os valores residuais (pIC_{50Obs} – pIC_{50Pred}) das melhores equações dos Bancos de Dados I a IV.

	Eq.E	Eq.J	Eq.L	Eq.Q
Eq.E	1,000			
Eq.J	0,559	1,000		
Eq.L	0,514	0,434	1,000	
Eq.Q	0,289	0,278	0,474	1,000

4.2.5. Predição das Potências Biológicas (pIC_{50}) de DABOs Não Incluídos nos Conjuntos de Treinamento e de Teste

Dois compostos da série original dos DABOs (Mai *et al.*, 2005; Manetti *et al.*, 2005), que neste trabalho recebem os números **76** e **77** (Figura 45), não foram incluídos na composição dos conjuntos de treinamento e de teste. Isto porque, apesar de ambos serem NH-DABOs, assim como os compostos análogos **57** (X=H) e **59** (X=Me) do conjunto de treinamento, eles divergem estruturalmente dos demais.

O NH-DABO **76** (X=Me) é análogo ao NH-DABO **59**, onde o grupo —NH-ciclopentila é substituído por —NH-fenila, sendo o único composto que apresenta um anel aromático (fenila) como substituinte **Y** na posição C2 do anel 4-oxo-pirimidina (Tabela 5, Metodologia), enquanto que todos os demais compostos têm um grupo alquila (ou ciclo-alquila) nesta posição.

O NH-DABO **77** (X=Me) também é análogo ao NH-DABO **59**, onde o grupo benzila (—CH₂Ph ligado na posição C6 do anel 4-oxo-pirimidina) é substituído por um grupo metilbenzila (—CHMePh), sendo o único composto que apresenta um centro estereogênico nesta posição (Tabela 5, Metodologia). É importante ressaltar que o composto **77** foi testado no ensaio biológico na forma de mistura racêmica.

Desta forma, foram feitas as predições teóricas das potências biológicas (pIC_{50}) destes compostos (**76** e **77**), aplicando as melhores equações dos Bancos de Dados I a IV (**Eq.E**, **Eq.J**, **Eq.L** e **Eq.Q**). Estes dois NH-DABOs possuem valores de pIC_{50} experimentais (**76**, $pIC_{50Obs.} = 7,46$ M e **77**-racemato, $pIC_{50Obs.} = 6,82$ M) comparáveis aos dos compostos dos conjuntos de treinamento e de teste, não extrapolando os limites máximo (**59**, $pIC_{50Obs.} = 7,52$ M) e mínimo (**1**, $pIC_{50Obs.} = 4,23$ M) de resposta biológica.

No caso do composto **77**, que possui um centro estereogênico, foram considerados os dois enantiômeros (*R* e *S*) isoladamente e, também, a mistura racêmica como as espécies ativas. Assim, considerando *R* ou *S* como o eutômero, o valor experimental de IC_{50} em (μM) foi multiplicado por dois ($pIC_{50Obs} = 6,52$ M). Por outro lado, considerando que os isômeros *R* e *S* sejam equipotentes na mistura racêmica, o valor experimental de IC_{50} em (μM) não foi alterado ($pIC_{50Obs} = 6,82$ M).

A Tabela 27 mostra os valores de pIC_{50} observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os compostos NH-DABOs **76** e **77** de acordo com as equações **E (BD-I)**, **J (BD-II)**, **L (BD-III)** e **Q (BD-IV)** e, para fins de comparação, são dados também os valores correspondentes dos dois NH-DABOs do conjunto de treinamento (**57** e **59**).

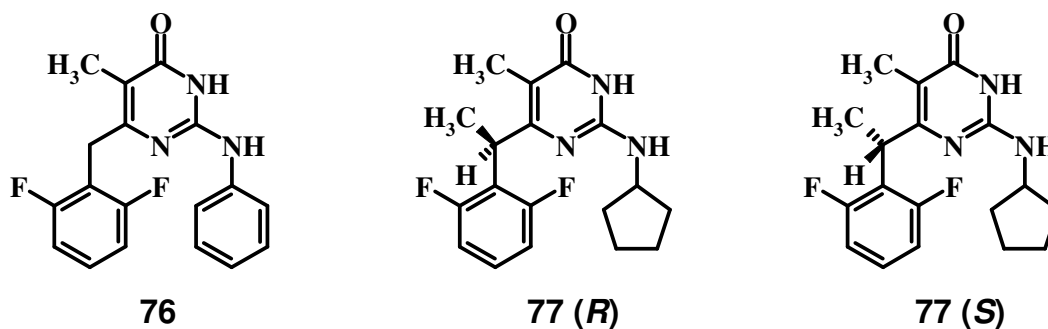


Figura 45. Estruturas dos NH-DABOs **76** e **77** (*R* e *S*) não incluídos nos conjuntos de treinamento e de teste para os quais foram feitas as predições teóricas das potências biológicas (pIC_{50}) aplicando as melhores equações (Bancos de Dados I a IV) de QSAR-3D dependente do receptor.

Tabela 27. Valores de pIC_{50} (M) observados, preditos e resíduos ($pIC_{50Obs} - pIC_{50Pred}$) para os NH-DABOs **57**, **59**, **76** e **77** (*R*, *S* e racemato) de acordo com as melhores equações dos Bancos de Dados I a IV.

#	$pIC_{50Obs.}$	$pIC_{50Pred.}$ (resíduo)			
		Eq.E (BD-I)	Eq.J (BD-II)	Eq.L (BD-III)	Eq.Q (BD-IV)
57	7,15	6,61 (0,54)	6,81 (0,34)	6,42 (0,73)	7,05 (-0,10)
59	7,52	6,68 (0,84)	6,71 (0,81)	6,41 (1,11)	6,85 (-0,67)
76	7,46	8,67 (-1,21) ^d	7,63 (-0,17)	7,87 (-0,41)	7,71 (-0,25)
77-(R)^a	6,52	3,35 (3,17) ^d	5,18 (1,34) ^d	4,92 (1,60)	8,21 (-1,69) ^d
77-(S)^b		8,90 (-2,38) ^d	7,70 (-1,18)	7,46 (-0,94)	10,47 (-3,95) ^d
77 (rac)^c	6,82	6,13 ^e (0,70)	6,44 ^e (0,38)	6,19 ^e (0,63)	9,34 ^e (-2,52) ^d

a) O valor experimental de IC_{50} (μM) foi multiplicado por dois, considerando que o isômero *R* é o eutômero. b) O valor experimental de IC_{50} (μM) foi multiplicado por dois, considerando que o isômero *S* é o eutômero. c) O valor experimental de IC_{50} em (μM) não foi alterado, considerando que os isômeros *R* e *S* sejam equipotentes na mistura racêmica. d) Composto considerado *outlier* de acordo com o SEE da respectiva equação. e) O valor de $pIC_{50Pred.}$ do racemato corresponde a média dos valores dos isômeros calculados isoladamente.

As potências preditas para o NH-DABO **76** (Tabela 27) são superestimadas por todas as equações, mas com valores residuais relativamente pequenos (resíduos < 0,5 em valor modular), com exceção da melhor equação, **Eq.E** (resíduo = -1,21), onde o composto é um *outlier*.

Entretanto, considerando que o análogo NH-DABO **59** (Y=ciclopentila) do conjunto de treinamento apresenta valor residual relativamente alto nesta equação (resíduo = 0,84), podemos dizer que o NH-DABO **76** (Y=fenila) é bem predito pela **Eq.E**. Além disso, a **Eq.E** é a única que prevê a ordem relativa correta de potência entre os dois NH-DABOs do conjunto de treinamento, os análogos **57** (X=H) e **59** (X=Me), onde **59** é mais potente do que **57**. As demais equações prevêm **57** como mais potente (**Eq.J** e **Eq.Q**) ou equipotente (**Eq.L**) ao composto **59**.

Não era esperado que **76** fosse relativamente bem predito pelas melhores equações, visto que os modelos não foram “treinados” para reconhecer um anel aromático na posição **C2** do anel 4-oxo-pirimidina, característica única deste composto.

A energia de interação estérica negativa ($-23,575 \text{ kcal.mol}^{-1}$) de **76** com o resíduo **Lys101**, referente ao termo **Lys101LJ** da **Eq.E**, ajuda a explicar este comportamento *outlier*. Como o coeficiente deste termo também é negativo ($-0,153$), ele contribui de forma significativa para o aumento de potência do composto. Fato semelhante ocorre com o termo **Tyr181LJ** (**Eq.E**), que também apresenta energia de interação ($-9,602 \text{ kcal.mol}^{-1}$) e coeficiente ($-0,110$) negativos.

A Figura 46 mostra o modo de ligação de **76** no NNBS da RT, destacando os resíduos **Lys101** e **Tyr181** relacionados aos termos correspondentes da **Eq.E**. Nesta figura, observa-se que o grupo amino ($-\text{NH-Ph}$) ligado à posição **C2** do anel 4-oxo-pirimidina do composto **76** interage com o átomo de oxigênio do grupo

carbonila da cadeia principal do resíduo **Lys101** a uma distância de 3,65 Å, o que configura uma interação por ligação hidrogênio de intensidade fraca (Jeffrey, 1997). Também é peculiar a distância (2,64 Å) entre o anel aromático do grupo 2,6-di-flúor-fenila ligado à posição C6 (do anel 4-oxo-pirimidina) e o anel aromático da cadeia lateral do resíduo **Tyr181** (Figura 46), o que sugere que esses grupos estejam interagindo por ligação hidrofóbica.

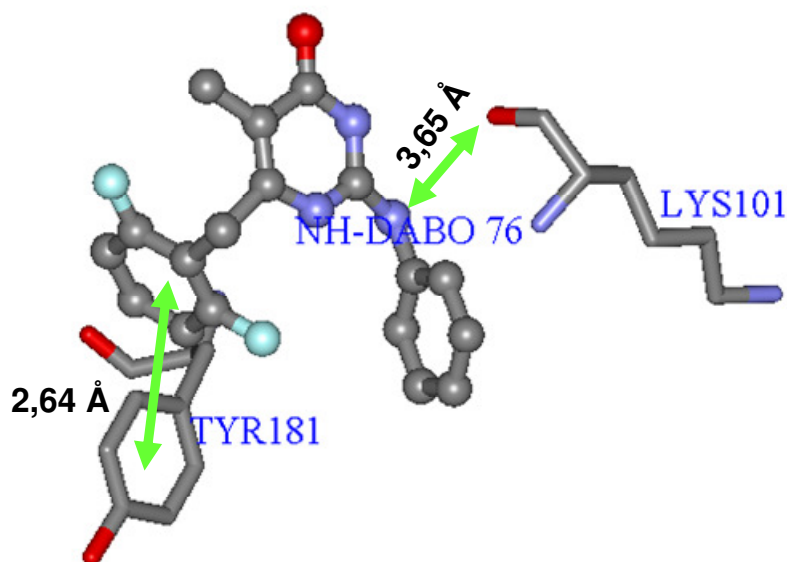


Figura 46. Modo de ligação do NH-DABO **76** (modelo bastão-e-bola, colorido por elemento) no NNBS da RT do HIV-1, destacando os resíduos (modelo bastão, colorido por elemento) **Lys101** e **Tyr181** relacionados aos termos correspondentes da **Eq.E**.

No caso do composto **77**, único que contém um centro estereogênico no substituinte da posição C6 do anel 4-oxo-pirimidina e que foi ensaiado farmacologicamente na forma de racemato, foram preditas as potências considerando três hipóteses: (a) o isômero (*R*) como o eutômero, (b) o isômero (*S*) como o eutômero, e (c) os isômeros (*R*) e (*S*) equipotentes.

Assim, as potências preditas para **77-(R)** (Tabela 27) são subestimadas por todas as equações, com exceção da **Eq.Q** (resíduo = -1,69), onde a potência é

superestimada, e em qualquer caso, com valores residuais relativamente grandes (resíduos $> 1,0$ em valor modular). No caso de **77**-(*S*), as potências preditas são superestimadas por todas as equações e com valores residuais relativamente grandes (resíduos $> 1,0$ em valor modular), com exceção da pior equação, **Eq.L** (resíduo = $-0,94$). No entanto, independente da equação, a ordem relativa de potência predita entre os isômeros (*R*) e (*S*) é sempre a mesma, *i.e.*, o isômero (*S*) é predito ser mais potente do que o isômero (*R*).

No caso da melhor equação, **Eq.E**, independente de se considerar o isômero (*R*) ou (*S*) como o eutômero, o isômero em questão é classificado como *outlier*. No entanto, quando se considera que os isômeros são equipotentes na mistura racêmica, o composto **77** não é mais classificado como *outlier* por esta equação. Este fato sugere que os isômeros sejam equipotentes, ou que o isômero (*S*) seja um pouco mais potente do que (*R*). No entanto, devido a falta dos valores de potência dos isômeros testados isoladamente, esta questão não pode ser esclarecida.

4.3. Planejamento e Proposição de Novos DABOs

A Química Medicinal apresenta diversos conceitos de modificação molecular, com os quais se pode fazer a otimização de compostos protótipos (Thomas, 2003, pp.28-37; Wermuth, 2003, pp.174), tais como a hibridação molecular, a simplificação molecular, a anelação ou restrição conformacional, a extensão de cadeias por homologia e pela formação de vinílogos e benzílogos e o bioisosterismo (Wermuth, 2003, pp. 175-214; Patani & LaVoie, 1996).

Desta forma, utilizando estes conceitos e as informações obtidas no presente estudo de QSAR-3D independente e dependente do receptor, foi possível propor um novo composto com perfil estrutural semelhante ao dos DABOs.

Nesse sentido, foram utilizados como compostos protótipos o NH-DABO **59** (o mais potente da série dos DABOs deste estudo) e a etravirina, um NNRTI da classe dos DAPYs, recentemente aprovado pelo FDA e que tem um esqueleto molecular semelhante ao dos DABOs. Como estratégias de modificação estrutural foram usadas a hibridação molecular e a homologia (Figura 47), estratégias utilizadas com sucesso em diversos casos de planejamento de fármacos (Wermuth, 2003, pp. 175-214).

O anel 4-oxopirimidina e o grupo –NH na posição **C2** dos DABOs serão mantidos, pois são grupos característicos da classe. Os –NH-DABOs são mais potentes do que os S-DABOs, por isso a escolha pelo grupo –NH, que também está presente na etravirina.

Da etravirina foi mantido o grupo aromático da posição **C2**, visto que o composto NH-DABO **76**, testado nas melhores Equações dos **BDs I a IV**, mostrou-se bem predito pelos modelos (Tabela 27), e a única diferença dele para o NH-DABO **59**, composto mais potente da série, é o anel aromático em **C2**. Como a potência de **76** foi aumentada em todas as Equações (Tabela 27), propõe-se que esse anel seja mantido. Porém, o mapa estérico de CoMFA (Figura 22.A) mostra um contorno amarelo próximo aos grupos alquila ligados ao –S ou –NH em **C2**, indicando que substituintes volumosos nessa região acarretariam em uma diminuição da potência dos compostos. Por isso propõe-se o grupo aromático sem o substituinte *para*-nitrila ($p\text{-C}\equiv\text{N}$), presente na etravirina. Bioisósteros do anel aromático fenila, como piridina, imidazola, triazola, tiazola, pirazola, tiofenila, furanila e triazinila, também poderiam ser avaliados.

O grupo metila em **C5** foi trocado por uma etila para maximizar as interações hidrofóbicas no NNBS, principalmente com a cadeia lateral do aminoácido **Val179**, que fica a uma distância de 2,40 Å desse grupo.

Em **C6** seria mantido o espaçador metileno (-CH₂-) dos DABOs ao invés do átomo de oxigênio da etravirina, pois de acordo com os modelos avaliados e as figuras 3D do NH-DABO **59** no NNBS, os resíduos mais próximos a este espaçador são as cadeias laterais de **Val106**, **Tyr181** e **Tyr188**, que não viabilizariam a formação de ligações hidrogênio com os inibidores. Ligado a este espaçador, estaria um anel aromático fenila com substituintes nas posições *orto* (metilas) e *para* (etinila). Este sistema aumentaria os contatos hidrofóbicos com os resíduos **Tyr181** e **Tyr188**, além de diminuir a dependência desses contatos através de uma maior interação com **Trp229**, cuja cadeia lateral está voltada para o grupo etino. Estas escolhas também foram feitas com base nos mapas estérico e eletrostático de CoMFA, que mostraram ser favoráveis substituintes com baixa densidade eletrônica nas posições *orto* do anel aromático em **C6**.

Uma característica importante do composto proposto é que ele não possui centro estereogênico, o que facilita a química sintética e evita efeitos ruins advindos de possíveis enantiômeros tóxicos.

Procurou-se também manter átomos com flexibilidade conformacional em **C2** e **C6**, de forma a possibilitar um bom ajuste do composto no NNBS.

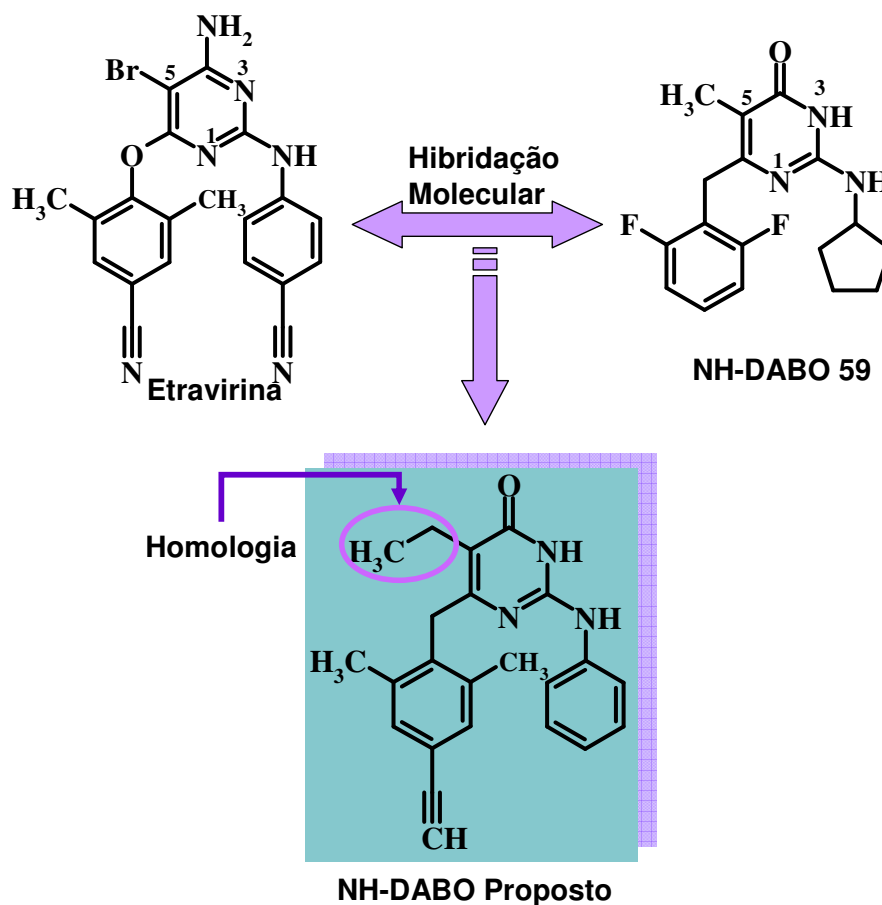


Figura 47. Proposição de novo NH-DABO por hibridação molecular entre o NH-DABO mais potente (**59**) e o fármaco NNRTI recém-lançado (etravirina), e por extensão de cadeia por homologia.

4.3.1. Predição da Potência Biológica (pIC_{50}) do NH-DABO Proposto

O composto proposto foi avaliado pelas **Equações E, J, L e Q** e sua potência foi predita pelos modelos de acordo com a Tabela 28. Ele foi avaliado com melhor potência pela **Eq.Q (BD-IV)**, $pIC_{50pred}=7,33$; seguido pela **Eq.E (BD-I)**, $pIC_{50pred}=7,14$; **Eq.J (BD-II)**, $pIC_{50pred}=6,83$ e, finalmente, **Eq.L (BD-III)**, $pIC_{50pred}=4,39$. De fato, a **Eq.L** foi a que apresentou a menor capacidade preditiva, com $q^2_{ajus}=0,594$. Curiosamente, a **Eq.Q** previu a atividade desse composto como a melhor, uma vez que essa Equação apresentou 5 *outliers* no conjunto de teste.

Tabela 28. Atividades previstas pelas melhores Equações dos **BD-I** a **IV** para o NH-DABO proposto.

#	pIC _{50Pred.} (M)			
	Eq.E (BD-I)	Eq.J (BD-II)	Eq.L (BD-III)	Eq.Q (BD-IV)
NH-DABO Proposto	7,14	6,83	4,39	7,33

A análise das possíveis interações que o composto proposto faz no NNBS pode ser visualizada na Figura 48. O anel fenila na posição **C2** possibilitou contatos hidrofóbicos importantes com os aminoácidos **Leu234** (2,20 Å) e **Tyr318** (3,67 Å). De forma análoga, a presença do grupo etila em **C5** aumentou o contato de van der Waals (3,94 Å) entre esse grupo e a cadeia lateral do aminoácido **Val179**. Esse novo ponto de interação é importante, visto que **Val179** não é um resíduo que sofre mutação na RT.

Mas a principal modificação nessa molécula proposta foi o anel substituído em **C6**, que aumentou as interações hidrofóbicas com quatro resíduos essenciais, **Tyr181** (3,88 Å), **Tyr188** (3,34 Å), **Phe227** (3,14 Å) e **Trp229** (3,05 Å). Aumentar as afinidades via contatos hidrofóbicos com resíduos como **Tyr181** e **Tyr188** é importante, entretanto, conforme enfatizado por Janssen *et al.*, 2005, como existem variantes virais com mutações nesses resíduos, torna-se urgente que se diminua a dependência de interação com esses resíduos e novos pontos de interação sejam encontrados com resíduos menos suscetíveis a mutações, como **Phe227** e **Trp229**. Dessa forma, o composto proposto demonstra boas perspectivas como NNRTI.

Obviamente que características farmacocinéticas e toxicológicas do novo composto precisam ser avaliadas, como LogP, volume molecular, adequação à regra dos cinco de Lipinski, e um *screening* toxicológico *in silico* antes de proceder à sua avaliação farmacológica.

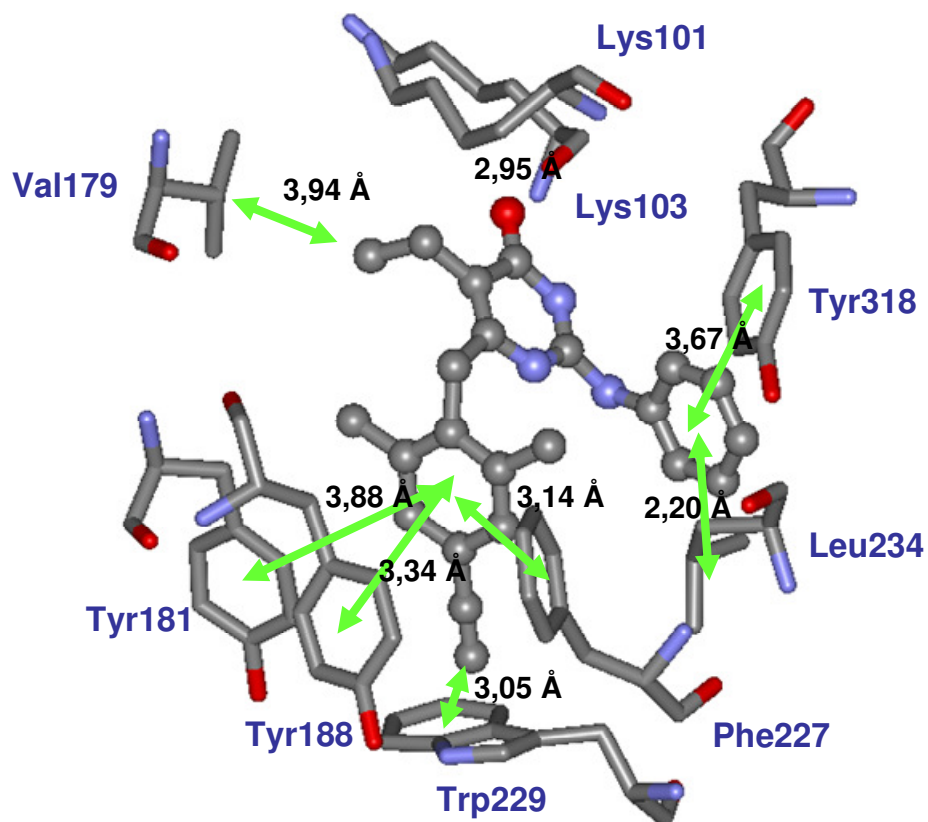


Figura 48. NH-DABO proposto (em bastão-e-bola, colorido por elemento) e possíveis interações com aminoácidos selecionados do NNBS (em bastão, coloridos por elemento).

5. CONCLUSÕES

Neste trabalho foram construídos e avaliados os primeiros modelos da literatura de QSAR-3D Independente (CoMFA) e Dependente da enzima RT do HIV-1 de uma série de 74 inibidores da classe dos S- e NH-DABOs, inibidores não-nucleosídeos ou alostéricos da enzima.

Os modelos independentes da enzima foram construídos com o programa CoMFA, seguindo três estratégias de alinhamento. O melhor deles foi construído com cargas PM3, *cutoff defaults* de 30 kcal.mol⁻¹ para ambos os campos estéricos e eletrostático, como átomo de prova o átomo de carbono sp³ com carga +1 e espaçamento da grade de 2,0 Å. Este modelo apresentou boa consistência interna em termos de r^2 e SEE e boa capacidade preditiva ($q^2=0,691$). Concluiu-se que os alinhamentos testados, as cargas atômicas parciais, o corte nas energias de interação e o átomo de prova, parâmetros variados nesse estudo, tiveram pouca influência sobre os valores estatísticos resultantes e, conseqüentemente, sobre os modelos.

As conclusões estruturais mais importantes do estudo de CoMFA foram a restrição do volume do substituinte na posição **C2** do anel 4-oxo-pirimidina, substituintes mais volumosos na posição **C5** e a presença de grupos ricos em elétrons na posição *para* do anel aromático em **C6**, que aumentariam a atividade biológica, fazendo dessas regiões importantes locais para futuras modificações estruturais.

A combinação das estruturas de raios-X dos complexos inibidor-receptor com estudos de CoMFA permitiu uma interpretação mais detalhada dos mapas de contorno do CoMFA, levando também a uma maior compreensão da interação dos complexos.

Os modelos de QSAR-3D dependentes da enzima foram avaliados em diferentes **BDs** e também apresentaram boa capacidade preditiva, o melhor modelo apresentou $q^2=0,660$. Eles foram consistentes com os dados de cristalografia e apontaram os aminoácidos mais importantes nas interações DABOs-RT, que são **Lys103, Tyr181, Tyr188, Phe227 e Trp229**.

Os termos de interação de van der Waals estiveram mais presentes nas melhores equações de cada **BD**, ratificando a importância da hidrofobicidade para os inibidores. As equações podem ser utilizadas para a predição de outros compostos que pertençam à mesma classe dos DABOs.

A hipótese farmacofórica foi validada pelos bons resultados estatísticos obtidos em ambos os estudos de QSAR-3D independente (CoMFA) e dependente da estrutura do receptor.

Na proposta de novos compostos foram utilizadas informações dos modelos de QSAR-3D Independente e Dependente da RT do HIV-1, usando hibridação molecular e extensão de cadeia por homologia, estratégias muito utilizadas de modificação molecular da Química Medicinal para o planejamento de fármacos. Além disso, a estratégia de bioisosterismo também poderia ser explorada com os anéis substituintes em **C6**.

6. PERSPECTIVAS

A resistência do HIV-1 aos NNRTIs permanece sendo um problema na terapia da AIDS e tem se mostrado um desafio para a comunidade científica, uma vez que vírus resistentes tem se desenvolvido rapidamente e que muitas vezes uma simples mutação pode tornar uma classe inteira de fármacos ineficaz contra o vírus (Mugavero & Hicks, 2004). Dado o número limitado de fármacos antiretrovirais disponíveis, esforços são essenciais no sentido de retardar ou prevenir a emergência de resistência. Dentro dessa perspectiva, pretende-se aplicar o mesmo tipo de análise feita nesse trabalho a enzimas RT mutantes a serem selecionadas e que estejam disponíveis no PDB a fim de aprofundar o conhecimento a respeito das interações entre os NNRTIs e os resíduos mutantes do sítio de ligação do HIV-1.

Adicionalmente, para determinar os requisitos estruturais e termodinâmicos que governam a inibição da RT pelos DABOs, propõe-se usar métodos híbridos que combinam cálculos quantum-mecânicos (QM) e de mecânica molecular (MM). Métodos QM/MM estendem o domínio dos cálculos QM a macromoléculas e permitem que maior rigor teórico seja simulado entre ligante e proteína.

Um conceito relativamente moderno em química de proteínas é que não apenas a estrutura, mas também a dinâmica são fundamentais na função biológica (Huang, 2005). Esse conceito reforça a utilização de simulações de dinâmica molecular para obter informações que são de difícil acesso experimental.

7. REFERÊNCIAS BIBLIOGRÁFICAS

Aboye, T. L.; Sobhia, M. E.; Bharatam, P. V. 3D-QSAR studies of pyruvate dehydrogenase kinase inhibitors based on a divide and conquer strategy. *Bioorg. Med. Chem.*, 12, 2709-2715, 2004.

AIDS Education & Training Centers National Resource Center, www.aids-etc.org/, acesso em 30 de agosto de 2008.

AIDS Education Global Information System, www.aegis.com/, acesso em 30 de agosto de 2008.

AIDS Epidemic Update, Special Report on HIV Prevention. Dezembro de 2007. <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/default.asp>, acesso em 30 de agosto de 2008.

Albany Medical College Division of HIV Medicine, www.amc.edu/patient/hiv/index.htm, acesso em 30 de agosto de 2008.

Albuquerque, M. G. Estudo de QSAR-3D de uma série de inibidores de transcriptase reversa e de uma série de antagonistas de receptor de tromboxana. Tese (Doutorado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 117 p. 1997.

Albuquerque, M. G.; Brito, M. A.; Cunha, E. F. F.; Alencastro, R. B.; Antunes, O. A. C.; Castro, H. C.; Rodrigues, C. R. Multidimensional-QSAR: Beyond the third dimension in drug design. *In: Current Methods in Medicinal Chemistry and Biological Physics*, Ed. Taft, C.A. e Silva, C.H.T.P. 1-10, 2007.

Aly, Y. L.; Pedersen, E. B.; La Colla, P.; Loddo, R. Novel synthesis and anti-HIV-1 activity of 2-arylthio-6-benzyl-2,3-dihydro-1H-pyrimidin-4-ones (aryl S-DABOs). *Synthesis*, 13, 1955-1960, 2007.

- Andries, K.; Azijn, H.; Thielemans, T.; Ludovici, D.; Kukla, M.; Heeres, J.; Janssen, P.; De Corte, B.; Vingerhoets, J.; Pauwels, R.; de Béthune, M. P. TMC125, a novel next-generation nonnucleoside reverse transcriptase inhibitor active against nonnucleoside reverse transcriptase inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, 48, 4680-4686, 2004.
- Arakawa, M.; Hasegawa, K.; Funatsu, K. QSAR study of anti-HIV HEPT analogues based on multi-objective genetic programming and counter-propagation neural network. *Chemom. & Intell. Lab. Syst.*, 3, 91-98, 2006.
- Avery, M. A.; Alvim-Gaston, M.; Rodrigues, C. R.; Barreiro, E. J.; Cohen, F. E.; Sabnis, Y. A.; Woolfrey, J. R. Structure–activity relationships of the antimalarial agent artemisinin 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies. *J. Med. Chem.*, 45, 292–303, 2002.
- Akamatsu, M. Current state and perspectives of 3D-QSAR. *Curr. Top. in Med. Chem.*, 12, 1381-1394, 2002.
- Amaral, A. T.; Montanari, C. A. Química Medicinal: 25 anos de planejamento racional de fármacos. *Quim. Nova*, 25, Supl. 1, 39-44, 2002.
- Amin, E. A.; Welsh, W. J. Highly predictive CoMFA and CoMSIA models for two series of stromelysin-1 (MMP-3) inhibitors elucidate S1' and S1-S2' binding modes. *J. Chem. Inf. Model.*, 46, 1775-1783, 2006.
- Artico, M.; Massa, S.; Mai, A.; Marongiu, M. E.; Piras, G.; Tramontano, E.; La Colla, P. 3,4-Dihydro-2-alkoxy-6-benzyl-4-oxopyrimidines (DABOs): a new class of specific inhibitors of human immunodeficiency virus Type 1. *Antiviral Chem. Chemother.*, 4, 361-368, 1993.

- Baba, M.; Tanaka, H.; De Clercq, E.; Pauwels, R.; Balzarini, J.; Schols, D.; Nakashima, H.; Perno, C. F.; Walker, R. T.; Miyasaka, T. Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclovir derivative. *Biochem. Biophys. Res. Commun.*, 165, 1375-1381, 1989.
- Baba, M.; De Clercq, E.; Tanaka, H.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Nitta, I.; Umezaki, K.; Nakashima, H.; Mori, S. Potent and selective inhibition of human immunodeficiency virus type 1 (HIV-1) by 5-ethyl-6-phenylthiouracil derivatives through their interaction with the HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA*, 88, 2356–2360, 1991.
- Balzarini, J. Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr. Top. Med. Chem.*, 4, 921-944, 2004.
- Bak, A. & Polanski, J. A 4D-QSAR study on anti-HIV HEPT analogues. *Bioorg. & Med. Chem.*, 14, 273-279, 2006.
- Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauter, C.; Axler-Blin, C.; Vézinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for AIDS. *Science*, 220, 868-71, 1983.
- Barreca, M. L.; Rao, A.; De Luca, L.; Zappala, M.; Monforte, A.-M.; Maga, G.; Pannecouque, C.; Balzarini, J.; De Clercq, E.; Chimiri, A.; Monforte, P. Computational strategies in discovering novel NNRTI. *J. Med. Chem.*, 48, 3433-3437, 2005.
- Barreiro, E. J.; Rodrigues, C. R.; Albuquerque, M. G.; Sant'Anna, C. M. R.; Alencastro, R. B. Modelagem molecular: uma ferramenta para o planejamento racional de fármacos em química medicinal. *Quím. Nova*, 20, 300- 310, 1997.

- Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91, 43-56, 1995.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.*, 28, 235-242, 2000.
- Bhongade, B. A.; Gadad, A. K. 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) Inhibitors: a strategic design in novel anticancer agents. *Bioorg. Med. Chem.*, 12, 2797-2805, 2004.
- Bird, L. E.; Philip, P. C.; Guillaume, B. E.; Stewart-Jones, J. R.; Stuart, D. I. Cloning, expression, purification, and crystallisation of HIV-2 reverse transcriptase. *Protein Expres. Purif.*, 27, 12-18, 2003.
- Botta, M.; Artico, M.; S Massa, A Gambacorta, ME Marongiu, A Pani, P La Colla. Synthesis, antimicrobial and antiviral activities of isotrimethoprim and some related derivatives. *Eur. J. Med. Chem.*, 27, 251-257, 1992.
- Brennan, T. M.; Taylor, D. L.; Bridges, C. G.; Leyda, J. P.; Tyms, A. S. The inhibition of human immunodeficiency virus type 1 in vitro by a non-nucleoside reverse transcriptase inhibitor MKC-442, alone and in combination with other anti-HIV compounds. *Antiviral Res.*, 26, 173-187, 1995.
- Brito, M. A. Modelos de CoMFA e CoMSIA de Antagonistas α_1 -Adrenérgicos N-Fenilpiperazínicos. Dissertação (Mestrado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 147 p. 2004.
- Burt, S. K.; Hutchins, C. W.; Greer, J. Predicting receptor-ligand interactions. *Curr. Op. in Struct. Biol.*, 1, 213-218, 1991.

- Camarasa, M. J.; Velazquez, S.; San-Felix, A.; Perez-Perez, M. J.; Gago, F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: A single mode of inhibition for the three HIV enzymes? *Antiviral Res.*, 71, 260-267, 2006.
- Castro, H.C.; Loureiro, N.I.V.; Pujol-Luz, M.; Souza, A.M.T.; Albuquerque, M.G.; Santos D. O.; Cabral, L. M.; Frugulhetti I.C.; Rodrigues, C.R. HIV-1 reverse transcriptase: a therapeutical target in the spotlight. *Curr. Med. Chem.*, 13, 313-324, 2006.
- Cavalla, D. Web Alert-Using the Internet for Medicinal Chemistry. *In The Practice of Medicinal Chemistry*. Ed. Wermuth, C. G. Elsevier, 782p., 2003.
- CDC National Prevention Information Network, www.cdcnpin.org, acesso em 30 de agosto de 2008.
- CDC, 2008a <http://www.cdc.gov/hiv>, acesso em 30 de agosto de 2008.
- CDC, 2008b <http://www.cdc.gov/hiv/pubs/facts/hiv2.htm>, acesso em 30 de agosto de 2008.
- Chen, X.; Rusinko, A., III; Young, S. S. Recursive partitioning analysis of a large structure-activity data set using three-dimensional descriptors. *J. Chem. Inf. Comp. Sci.*, 38, 112-116, 1998.
- Chiu, T. L.; So, S. S. Development of Neural Network QSPR Models for Hansch Substituent Constants. 2. Applications in QSAR Studies of HIV-1 Reverse Transcriptase and Dihydrofolate Reductase Inhibitors. *J. Chem. Inf. Comput. Sci.*, 44, 154-160, 2004.

- Cho, S. J.; Tropsha, A. Cross-validated r^2 -guided region selection for CoMFA: a simple method to achieve consistent results. *J. Med. Chem.*, 38, 1060-1066, 1995.
- Cho, S. J.; Garsia, M. L. S.; Bier, J.; Tropsha, A. Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors. *J. Med. Chem.*, 39, 5064-5071, 1996.
- Coffin, J. M. HIV Population Dynamics In Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy. *Science*, 267, 483-489, 1995.
- Corbett, J. W.; Ko, S. S.; Rodgers, J. D.; Jeffrey, S.; Bachelier, L. T.; Klabe, R. M.; Diamond, S.; Lai, C.-M.; Rabel, S. R.; Saye, J. A.; Adams, S. P.; Trainor, G. L.; Anderson, P. S.; Erickson-Viitanen S. K. Expanded-Spectrum Nonnucleoside Reverse Transcriptase Inhibitors Inhibit Clinically Relevant Mutant Variants of Human Immunodeficiency Virus Type 1. *Antimicrob. Agents Chemother.*, 43, 2893-2897, 1999.
- 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.*, 110, 5959-5967, 1988.
- Cunha, E. F. F. Estudos quantitativos de correlação estrutura-atividade em 3D e 4D de inibidores peptídicos da HIV-1 protease. Tese (Doutorado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 165 p., 2006.
- Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98, 10089-10092, 1993.

- Das, K.; Clark, A. D., Jr.; Lewi, P. J.; Heeres, J.; de Jonge, M. R.; Koymans, L. M. H.; Vinkers, H. M.; Daeyaert, F.; Ludovici, D. W.; Kukla, M. J.; De Corte, B.; Kavash, R. W.; Ho, C. Y.; Ye, H.; Lichtenstein, M. A.; Andries, K.; Pauwels, R.; de Bethune, M.-P.; Boyer, P. L.; Clark, P.; Hughes, S. H.; Janssen, P. A. J.; Arnold, E. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J. Med. Chem.*, 47, 2550–2560, 2004.
- D'Cruz, O.J.; Uckun, F.M. Novel tight binding PETT, HEPT and DABO-based non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Enz. Inhib. Med. Chem.*, 21, 329-350, 2006.
- Daszykowski, M.; Stanimirova, I.; Walczak, B.; Daeyaert, F.; de Jonge, M. R.; Heeres, J.; Koymans, L. M. H.; Lewi, P. J.; Vinkers, H. M.; Janssen, P. A.; Massart, D. L. Improving QSAR models for the biological activity of HIV Reverse Transcriptase inhibitors: Aspects of outlier detection and uninformative variable elimination. *Talanta*, 68, 54-60, 2005.
- De Clercq, E. The role of non-nucleoside reverse transcriptase inhibitors in the therapy of HIV-1 infection. *Antiv. Res.*, 38, 153-179, 1998.
- De Clercq, E. E. Perspectives of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV infection. *II Farmaco*, 54, 26-4, 1999.
- De Clercq, E. New developments in anti-HIV chemotherapy. *II Farmaco*, 56, 3-12, 2001.
- De Clercq, E. New developments in anti-HIV-1 chemotherapy. *Biochim. & Biophys. Acta*, 1587, 258-275, 2002.
- De Clercq, E. HIV-chemotherapy and prophylaxis: new drugs, leads and approaches. *Int. J. Biochem. Cell Biol.*, 36, 1800-1822, 2004.

- De Clercq, E. New Approaches toward Anti-HIV Chemotherapy. *J. Med. Chem.*, 48, 1297-1313, 2005.
- De Corte, B. L. From 4,5,6,7-Tetrahydro-5-methylimidazo[4,5,1-jk](1,4)benzodiazepin-2(1H)-one (TIBO) to Etravirine (TMC125): Fifteen Years of Research on Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase. *J. Med. Chem.*, 48, 1689-1696, 2005.
- De Souza, M. V. N.; Almeida, M. V. Drogas anti-HIV-1: passado, presente e perspectivas futuras. *Quím. Nova*, 26, 366-372, 2003.
- Debnath, A. K. Quantitative Structure-Activity Relationship (QSAR) Paradigm - Hansch Era to New Millennium. *Mini Rev. Med. Chem.*, 1, 187-195, 2001.
- Debyser, Z.; Pauwels, R.; Andries, K.; Desmyter, J.; Kukla, M.; Janssen, P. A. J.; De Clercq, E. An Antiviral Target on Reverse Transcriptase of Human Immunodeficiency Virus Type 1 Revealed by Tetrahydroimidazo [4,5,1-jk] [1,4] Benzodiazepin-2 (1H)-One and -Thione Derivatives. *PNAS*, 88, 1451-1455, 1991.
- Deng, B. L.; Hartman, T. L.; Buckheit, R. W.; Pannecouque, C.; De Clercq, E.; Cushman, M. Replacement of the Metabolically Labile Methyl Esters in the Alkenyl diarylmethane Series of Non-Nucleoside Reverse Transcriptase Inhibitors with Isoxazolone, Isoxazole, Oxazolone, or Cyano Substituents. *J. Med. Chem.*, 49, 5316 -5323, 2006.
- Devillers, J. Genetic Algorithms in Computer-Aided Molecular Design. *In Genetic Algorithms in Molecular Modeling*. Ed. James Devillers. Ed. Academic Press, San Diego. 327 p., 1996.

- Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107, 3902, 1985.
- Do Amaral, A. T.; Montanari, C. A. Química Medicinal: 25 anos de planejamento racional de fármacos. *Quím. Nova*, 25, 39-44, 2002.
- Douali, L.; Villemin, D.; Cherqaoui, D. Neural Networks: Accurate Nonlinear QSAR Model for HEPT Derivatives. *J. Chem. Inf. Comput. Sci.*, 43, 1200-1207, 2003.
- Dunn, W. J.; Rogers, D. Genetic Partial Least Squares in QSAR. In *Genetic Algorithms in Molecular Modeling*. Ed. James Devillers. Ed. Academic Press, San Diego. 327 p., 1996.
- El-Brollosy, N. R.; Jørgensen, P. T.; Dahan, B.; Boel, A. M.; Pedersen, E. B.; Nielsen, C. Synthesis of Novel N-1 (Allyloxymethyl) Analogues of 6-Benzyl-1-(ethoxymethyl)-5-isopropyluracil (MKC-442, Emivirine) with Improved Activity Against HIV-1 and Its Mutants. *J. Med. Chem.*, 45, 5721-5726, 2002.
- Eriksson, M. A. L.; Pitera, J.; Kollman, P. A. Prediction of the Binding Free Energies of New TIBO-like HIV-1 Reverse Transcriptase Inhibitors Using a Combination of PROFEC, PB/SA, CMC/MD, and Free Energy Calculations. *J. Med. Chem.*, 42, 868-881, 1999.
- Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Nat. Struct. Biol.*, 2, 303-308, 1995.

Fattorusso, C.; Gemma, S.; Butini, S.; Huleatt, P.; Catalanotti, B.; Persico, M.; De Angelis, M.; Fiorini, I.; Nacci, V.; Ramunno, A.; Rodriguez, M.; Greco, G.; Novellino, E.; Bergamini, A.; Marini, S.; Coletta, M.; Maga, G.; Spadari, S.; Campiani, G. Specific Targeting Highly Conserved Residues in the HIV-1 Reverse Transcriptase Primer Grip Region. Design, Synthesis, and Biological Evaluation of Novel, Potent, and Broad Spectrum NNRTIs with Antiviral Activity. *J. Med. Chem.*, 48, 7153-7165, 2005.

FDA, 2003 <http://www.fda.gov/bbs/topics/NEWS/2003/NEW00879.html>, acesso em 02 de setembro de 2008.

FDA, 2007 <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01726.html>, acesso em 02 de setembro de 2008.

FDA, 2008 <http://www.fda.gov/oashi/aids/virals.html>, acesso em 02 de setembro de 2008.

FDA, 2008, Intelence Prescribing Information. <http://www.fda.gov/cder/foi/label/2008/022187lbl.pdf>, acesso em 02 de setembro de 2008.

Flexner, C. Antiretroviral Agents and Treatment of HIV Infection. *In Goodman & Gilman: The pharmacological basis of therapeutics*. Ed. Brunton, L. L. Ed. McGraw Hill, 1984 p., 2006.

Flexner, C. HIV drug development: the next 25 years. *Nat. Rev. Drug Disc.*, 6, 959-966, 2007.

Freimoser, F. M.; Jakob, C. A.; Aebi, M.; Tuor, U. The MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide] Assay Is a Fast and Reliable Method for Colorimetric Determination of Fungal Cell Densities. *Applied and Environmental Microbiology*, 65, 3727-3729, 1999.

- Gagnon, A.; Amad, M. H.; Bonneau, P. R.; Coulombe, R.; DeRoy, P. L.; Doyon, L.; Duan, J.; Garneau, M.; Guse, I.; Jakalian, A.; Jolicoeur, E.; Landry, S.; Malenfant, E.; Simoneau, B.; Yoakim, C. Thiotetrazole alkynylacetanilides as potent and bioavailable non-nucleoside inhibitors of the HIV-1 wild type and K103N/Y181C double mutant reverse transcriptases. *Bioorg. & Med. Chem. Lett.*, 17, 4437–4441, 2007.
- Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, 224, 500-503, 1984.
- Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A.K.;Hansch,C. Comparative Quantitative Structure-Activity Relationship Studies on Anti-HIV Drugs. *Chem. Rev.*, 99, 3525-3602, 1999.
- Gaudio, A. C.; Zandonade, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. *Quím. Nova*, 24, 658-671, 2001.
- Gáudio, A. C.; Montanari, C. A. HEPT derivatives as non-nucleoside inhibitors of HIV-1 reverse transcriptase: QSAR studies agree with the crystal structures. *J. Comput.-Aid. Mol. Des.*, 16, 287-295, 2002.
- Gayen, S.; Debnath, B.; Samanta, S.; Jha, T. QSAR study on some anti-HIV HEPT analogues using physicochemical and topological parameters. *Bioorg. & Med. Chem.*, 12, 1493-1503, 2004.
- Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, 186, 1-17a, 1986.

- Geladi, P.; Kowalski, B. R. An example of 2-block predictive partial least squares regression with simulated data. *Anal. Chim. Acta*, 186, 19-32b, 1986.
- Glenn, W.G.; Dunn, I.W.J.; Scott, D.R. Principal components analysis and partial least square. *Tetrahedron Computer Methodology*, 2, 349-376, 1989.
- Golbraikh, A.; Bernard, P.; Chrétien, J. R. Validation of protein-based alignment in 3D quantitative structure-activity relationships with CoMFA models. *Eur. J. Med. Chem.*, 35, 123-136, 2000.
- Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y. D.; Lee, K.-H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comp. Aided Mol. Des.*, 17, 241-253, 2003.
- Güner, O. F. History and evolution of the pharmacophore concept in computeraided drug design. *Curr. Top. Med. Chem.*, 12, 1321-1332, 2002.
- Guo, W.; Hu, X.; Chu, N.; Yin, C. Quantitative structure-activity relationship studies on HEPTs by supervised stochastic resonance. *Bioorg. & Med. Chem. Lett.*, 16, 2855-2859, 2006.
- Hannongbua, S.; Nivesanond, K.; Lawtrakul, L.; Pungpo, P.; Wolschann, P. 3D-Quantitative Structure-Activity Relationships of HEPT Derivatives as HIV-1 Reverse Transcriptase Inhibitors, Based on Ab Initio Calculations. *J. Chem. Inf. Comput. Sci.*, 41, 848-855, 2001.
- Hansch, C.; Fujita, T. Rho-Sigma-Pi Analysis. Method for Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.*, 86, 1616-1626, 1964.
- Hansch, C. A Quantitative Approach to Biochemical Structure-Activity Relationships. *Acc. Chem. Res.*, 2, 232-239, 1969.

- Harms, G.; Feldmeier, H. Review: HIV infection and tropical parasitic diseases - deleterious interactions in both directions? *Trop. Med. & Inter. Health.*, 7, 479-488, 2002.
- Haubrich, R.; Gubernick, S.; Yasothan, U.; Kirkpatrick, P. Fresh from the Pipeline: Etravirine. *Nat. Rev. Drug Disc.*, 7, 287-288, 2008.
- Hayden, F. G. Antimicrobial agents – antiviral agents (nonretroviral). In *Goodman & Gilman's, The pharmacological basis of therapeutics*. Ed. Joel G. Hardman & Lee E. Limbird. Ed. McGraw-Hill. 10a Ed. 2148 p., 2001.
- He, Y.; Chen, F.; Sun, G.; Wang, Y.; De Clercq, E.; Balzarini, J.; Pannecouque, C. 5-Alkyl-2-[(aryl and alkyloxylcarbonylmethyl)thio]-6-(1-naphthylmethyl) pyrimidin 4(3H)-ones as an unique HIV reverse transcriptase inhibitors of S-DABO series. *Bioorg. & Med. Chem. Lett.*, 14, 3173–3176, 2004.
- Hemmateenejad, B.; Miri, R.; Akhond, M.; Shamsipur, M. QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. *Chem. Intell. Lab. Syst.*, 64, 91-99, 2002.
- Hightower, M.; Kallas, E. G. Diagnosis, antiretroviral therapy, and emergence of resistance to antiretroviral agents in HIV-2 infection: a review. *Braz. J. Infect. Dis.*, 7, 7-15, 2003.
- Himmel, D. M.; Das, K.; Clark, A. D.; Hughes, S. H.; Benjahad, A.; Oumouch, S.; Guillemont, G.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C. H.; Grierson, D. S.; Arnold, E. Crystal Structures for HIV-1 Reverse Transcriptase in Complexes with Three Pyridinone Derivatives: A New Class of Non-Nucleoside Inhibitors Effective against a Broad Range of Drug-Resistant Strains. *J. Med. Chem.*, 48, 7582 -7591, 2005.

HIV Clinical Resource, New York State Department of Health AIDS Institute, www.hivguidelines.org, acesso em 02 de setembro de 2008.

Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.*, 39, 1589-1600, 1996.

Hopkins, A. L.; Ren, J.; Tanaka, H.; Baba, M.; Okamoto, M.; Stuart, D. I.; Stammers, D. K. Design of MKC-442 (Emivirine) Analogues with Improved Activity Against Drug-Resistant HIV Mutants. *J. Med. Chem.*, 42, 4500-4505, 1999.

Hopkins, A. L.; Ren, J.; Milton, J.; Hazen, R. J.; Chan, J. H.; Stuart, D. I.; Stammers, D. K. Design of NNRTIs with improved drug resistance properties. *J. Med. Chem.*, 47, 5912-5922, 2004.

Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D.; Hughes, S. H. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure*, 4, 853-860, 1996.

Hsiou, Y.; Das, K.; Ding, J.; Clark, A. D., Jr.; Kleim, J. P. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.*, 284, 313-323, 1998.

Hopfinger, A. J. A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc.*, 102, 7196-7206, 1980.

Hopfinger A. J.; Wang S.; Tokarski J. S.; Jin B. Q.; Albuquerque M. G.; Madhav P. J.; Duraiswami C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.*, 119, 10509-10524, 1997.

- Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, 282, 1669-1675, 1998.
- Huuskonen, J. QSAR modeling with the electrotopological state: TIBO derivatives. *J. Chem. Inf. Comput. Sci.*, 41, 425-429, 2001.
- Hye-Jung, K.; Chae, C. H.; Yi, K. Y.; Park, K. L.; Yoo, S. E. Computational studies of COX-2 inhibitors: 3D-QSAR and docking. *Bioorg. & Med. Chem.*, 12, 1629–1641, 2004.
- Hyperchem 7.0, Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA, 2002.
- Ivanciuc, O. QSPR/QSAR Studies by Molecular Descriptors. *In 3D QSAR Models.*, Ed. Diudea, M. V. Ed. New Science Publishers, 251-253, 2000.
- Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D., Jr.; Lu, X. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 90, 6320-6324, 1993.
- Jalali-Heravi, M.; Parastar, F. Use of Artificial Neural Networks in a QSAR Study of Anti-HIV Activity for a Large Group of HEPT Derivatives. *J. Chem. Inf. Comput. Sci.*, 40, 147-154, 2000.

- Janssen, P. A. J.; Lewi, P. J.; Arnold, E.; Daeyaert, F.; de Jonge, M.; Heeres, J.; Koymans, L.; Vinkers, M.; Guillemont, J.; Pasquier, E.; Kukla, M.; Ludovici, D.; Andries, K.; Béthune, M.-P.; Pauwels, R.; Das, K.; Clark, A. D.; Frenkel, Y. V.; H. Hughes, S. H.; Medaer, B.; De Knaep, F.; Bohets, H.; De Clerck, F.; Lampo, A.; Williams, P.; Stoffels, P. In Search of a Novel Anti-HIV Drug: Multidisciplinary Coordination in the Discovery of 4-[[4-[[4-[(1E)-2-Cyanoethenyl]-2,6-dimethylphenyl] amino]-2- pyrimidinyl]amino]benzonitrile (R278474, Rilpivirine). *J. Med. Chem.*, 48, 1901-1909, 2005.
- Jayatilleke, P. R. N., Nair, A. C., Zauhar, R., Welsh, W. J. Computational studies on HIV-1 protease inhibitors: influence of calculated inhibitor-enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models. *J. Med. Chem.*, 43, 4446–4451, 2000.
- Johns Hopkins AIDS Service, www.hopkins-aids.edu, acesso em 02 de setembro de 2008.
- Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.*, 16, 663-667, 2006.
- Kaufmann, G. R.; Cooper, D. A. Antiretroviral therapy of HIV-1 infection: established treatment strategies and new therapeutic options. *Curr. Op. Microb.*, 3, 508-514, 2000.
- Kireev, D. B.; Chretien, J. R.; Grierson, D. S.; Monneret, C. A. 3D QSAR Study of a Series of HEPT Analogues: The Influence of Conformational Mobility on HIV-1 Reverse Transcriptase Inhibition. *J. Med. Chem.*, 40, 4257-4264, 1997.
- Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.*, 37, 4130-4136, 1994.

- Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 angstrom resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, 256, 1783-1790, 1992.
- Kubinyi, H.; Abraham, U. *In 3D QSAR in Drug Design. Theory, Methods and Applications*. Escom Leiden, p. 717-728, 1993.
- Kubinyi, H. *Hansch Analysis and Related Approaches*. New York: VCH. 240 p., 1993.
- Kubinyi, H. Lock and key in the real world: concluding remarks. *Pharm. Acta Helv.* 69, 259-269, 1995.
- Kubinyi, H. QSAR and 3D-QSAR in drug design. Part 1: methodology. *Drug Disc. Today*, 11, 457-467, 1997.
- Kubinyi, H. QSAR and 3D-QSAR in drug design. Part 2: applications and problems. *Drug Disc. Today*, 12, 538-546, 1997.
- Kulkarni, S. S.; Kulkarni, V. M. Structure Based Prediction of Binding Affinity of Human Immunodeficiency Virus-1 Protease Inhibitors. *J. Chem. Inf. Comput. Sci.*, 39, 1128-1140, 1999.
- Kuno, M.; Hongkrenkai, R.; Hannongbua, S. ONIOM-BSSE scheme for H... π system and applications on HIV-1 reverse transcriptase. *Chem. Phys. Lett.*, 424, 172-177, 2006.
- Kroeger-Smith, M. B.; Rouzer, C. A.; Taneyhill, L. A.; Smith, N. A.; Hughes, S. H.; Boyer, P. L.; Janssen, P. A.; Moereels, H.; Koymans, L.; Arnold, E. Molecular modeling studies of HIV-1 reverse transcriptase nonnucleoside inhibitors: Total energy of complexation as a predictor of drug placement and activity. *Protein Sci.*, 4, 2203-2222, 1995.

- Kroeger-Smith, M. B.; Hose, B. M.; Hawkins, A.; Lipchock, J.; Farnsworth, D. W.; Rizzo, R. C.; Tirado-Rives, J.; Arnold, E.; Zhang, W.; Hughes, S. H.; Jorgensen, W. L.; Michejda, C. J.; Smith, R. H., Jr. Molecular Modeling Calculations of HIV-1 Reverse Transcriptase Nonnucleoside Inhibitors: Correlation of Binding Energy with Biological Activity for Novel 2-Aryl-Substituted Benzimidazole Analogues. *J. Med. Chem.*, 46, 1940-1947, 2003.
- Kroemer, K. T.; Hecht, P.; Liedl, K. R. Different electrostatic descriptors in comparative molecular field analysis: a comparison of molecular electrostatic and Coulomb potentials. *J. Comp. Chem.*, 17, 1296-1308, 1996.
- Lazzarin, A.; Campbell, T.; Clotet, B.; Johnson, M.; Katlama, C.; Moll, A.; Towner, W.; Trottier, B.; Peeters, M.; Vingerhoets, J.; Smedt, G.; Baeten, B.; Beets, G.; Sinha, R.; Woodfall, B. Efficacy and safety of TMC125 (etravirine) in treatment-experienced HIV-1-infected patients in DUET-2: 24-week results from a randomised, double-blind, placebo-controlled trial. *Lancet*, 370, 39-48, 2007.
- Leach, A. R. *Molecular Modelling. Principles and applications*. Ed. Longman, 5th Ed. 595 p., 2001.
- Levine, I. N. *Quantum chemistry*. Ed. Prentice Hall, 739 p., 2000.
- Li, T.; Mei, H.; Cong P. Combining nonlinear PLS with the numeric genetic algorithm for QSAR. *Chem. Intell. Lab.Syst.*, 45, 177-184, 1999.
- Lill, M. A.; Winiger, F.; Vedani, A.; Ernst, B. Impact of Induced Fit on Ligand Binding to the Androgen Receptor: A Multidimensional QSAR Study To Predict Endocrine-Disrupting Effects of Environmental Chemicals. *J. Med. Chem.*, 48, 5666-5674, 2005.

- Lindahl, E.; Hess, B.; van der Spoel, D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7, 306-317, 2001. (<http://www.gromacs.org/>)
- Linden, R. *Algoritmos Genéticos*. Ed. Brasport, Rio de Janeiro, 384 p., 2006.
- Livingstone, D. *Data Analysis for Chemists*. New York, Oxford University Press. 256 p., 1995.
- Loew, G.,H.; Villar, H. O.; Alkorta, I. Strategies for indirect computer-aided drug design. *Pharmac. Res.*, 10, 475-486, 1993.
- Luco, J. M.; Ferretti, F. H. QSAR Based on Multiple Linear Regression and PLS Methods for the Anti-HIV Activity of a Large Group of HEPT Derivatives. *J. Chem. Inf. Comput. Sci.*, 37, 392-401, 1997.
- Magalhães, C. S; Barbosa, H. J. C.; Dardenne, L. E. Métodos de docking receptor-ligante para o desenho racional de compostos bioativos. In *Métodos de Química Teórica e Modelagem Molecular*. Eds. Nelson H. Morgon e Kaline Coutinho. Ed. Livraria da Física, São Paulo, 539 p., 2007.
- Mai, A.; Artico, M.; Sbardella, G.; Massa, S.; Loi, A. G.; Tramontano, E.; Scano, P.; La Colla, P. Synthesis and anti-HIV-1 activity of thio analogues of dihydroalkoxybenzyloxypyrimidines. *J. Med. Chem.*, 38, 3258-3263, 1995.
- Mai, A.; Artico, M.; Sbardella, G.; Quartarone, S.; Massa, S.; Loi, A. G.; De Montis, A.; Scintu, F.; Putzolu, M.; La Colla, P. Dihydro(alkylthio) (naphthylmetil) oxypyrimidines: novel non-nucleoside reverse transcriptase inhibitors of the S-DABO series. *J. Med. Chem.*, 40, 1447-1454, 1997.

- Mai, A.; Artico, M.; Sbardella, G.; Massa, S.; Novellino, E.; Greco, G.; Loi, A. G.; Tramontano, E.; Marongiu, M. E.; La Colla, P. 5-Alkyl-2-(alkylthio)-6-(2,6-dihalofenilmetil)-3,4-dihydropyrimidin-4(3H)-ones: novel potent and selective dihydro-alkoxy-benzyl-oxopyrimidine derivatives. *J. Med. Chem.*, 42, 619-627, 1999.
- Mai, A.; Artico, M.; Ragno, R.; Sbardella, G.; Massa, S.; Maga, C. G.; La Colla, P. 5-Alkyl-2-alkylamino-6-(2,6-difluorophenylalkyl)-3,4-dihydropyrimidin-4(3H)-ones, a new series of potent, broad-spectrum non nucleoside reverse transcriptase inhibitors belonging to the DABO family. *Bioorg. & Med. Chem.*, 13, 2065–2077, 2005.
- Madruga, J. V.; Cahn, P.; Grinsztejn, B.; Haubrich, R.; Lalezari, J.; Mills, A.; Pialoux, G.; Wilkin, T.; Peeters, M.; Vingerhoets, J.; Smedt, G.; Leopold, L.; Trefiglio, R.; Woodfall, B. Efficacy and safety of TMC125 (etravirine) in treatment experienced HIV-1-infected patients in DUET-1: 24-week results from a randomised, double-blind, placebo-controlled trial. *Lancet*, 370, 29–38, 2007.
- Manetti, F.; Este, J. A.; Clotet-Codina, I.; Armand-Ugo'n, M.; Maga, G.; Crespan, E.; Cancio, R.; Mugnaini, C.; Bernardini, C.; Togninelli, A.; Carmi, C.; Alongi, M.; Petricci, E.; Massa, S.; Corelli, F.; Botta, M. Parallel Solution-Phase and Microwave-Assisted Synthesis of New S-DABO Derivatives Endowed with Subnanomolar Anti-HIV-1 Activity. *J. Med. Chem.*, 48, 8000-8008, 2005.
- Mao, C.; Sudbeck, E. A.; Venkatachalam, T. K.; Uckun, F. M. Structure-based drug design of non-nucleoside inhibitors for wild-type and drug-resistant HIV reverse transcriptase. *Biochem. Pharmacol.*, 60, 1251-1265, 2000.
- Marshall, G. R.; Cramer III, R. D. Thre-dimensional structure-activity relationships. *Trends in Pharm. Sci.*, 9, 285-289, 1988.

- Martínez, L.; Borin, I. A.; Skat, M. S. Fundamentos de Simulação por Dinâmica Molecular. In *Métodos de Química Teórica e Modelagem Molecular*. Eds. Nelson H. Morgon e Kaline Coutinho. Editora Livraria da Física, São Paulo. 539 p., 2007.
- Medina-Franco, J. L.; Rodríguez-Morales, S.; Juárez-Gordiano, C.; Hernández-Campos, A.; Jiménez-Barbero, J.; Castillo, R. Flexible docking of pyridinone derivatives into the non-nucleoside inhibitor binding site of HIV-1 reverse transcriptase. *Bioorg. & Med. Chem.*, 12, 6085-6095, 2004.
- Melville, J. L.; Hirst, J. D. On the stability of CoMFA models. *J. Chem. Inf. Comput. Sci.*, 44, 1294–1300, 2004.
- Miyasaka, T.; Tanaka, H.; Baba, M.; Hayakawa, H.; Walker, R. T.; Balzarini, J.; De Clercq, E. A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxyethoxy)methyl]-6-(fenilthio)thymine. *J. Med. Chem.*, 32, 2507-2509, 1989.
- Mota-Neto, J. D.; Zerner, M.; de Alencastro, R. B. A possible mechanism of molecular recognition for the reverse transcriptase of HIV-1. *Int. J. Quant. Chem.*, 44, 225-253, 1992a.
- Mota-Neto, J. D.; Zerner, M.; de Alencastro, R. B. On the implications of the structure of 3'-azido-3'-deoxythymidine and related compounds to antiviral activity. *Int. J. Quant. Chem.*, 44, 743-757, 1992b.
- Mugavero, M. J.; Hicks, C. B. HIV resistance and the effectiveness of combination antiretroviral treatment. *Drug Disc. Today*, 1, 529-535, 2004.

- Mugnaini, C.; Alongi, M.; Togninelli, A.; Gevariya, H.; Brizzi, A.; Manetti, F.; Bernardini, C.; Angeli, L.; Tafi, A.; Bellucci, L.; Corelli, F.; Massa, S.; Maga, G.; Samuele, A.; Facchini, M.; Clotet-Codina, I.; Armand-Ugon, M.; Este, J.A.; Botta, M. Dihydro-alkylthio-benzyl-oxypyrimidines as Inhibitors of Reverse Transcriptase: Synthesis and Rationalization of the Biological Data on Both Wild-Type Enzyme and Relevant Clinical Mutants. *J. Med. Chem.*, 50, 6580-6595, 2007.
- NCBI, 2008 <http://www.ncbi.nlm.nih.gov/retroviruses/>, acesso em 02 de setembro de 2008.
- New York State Department of Health AIDS Institute, <http://www.hivguidelines.org/>, acesso em 02 de setembro de 2008.
- Niculescu, S. P. Artificial neural networks and genetic algorithms in QSAR. *J. Mol. Struct.: THEOCHEM*, 622, 71-83, 2003.
- Oprea, T. I.; Marshall, G. R. Receptor-Based Prediction of Binding Affinities. In *3D QSAR in drug design*. Eds. Kubinyi, H.; Folkers, G.; Martin, Y. C.; Kluwer Academic Publishers, London, Vol. 2, pp 35-61, 2002.
- Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.*, 38, 2681-2691, 1995.
- Parreira, R. L. T.; Júnior, O. A.; Galembeck, S. E. Conformational preferences of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Tetrahedron*, 57, 3243-3253, 2001.
- Patani, G. A.; LaVoie, E. J. Bioisosterism: A Rational Approach in Drug Design. *Chem. Rev.*, 96, 3147-3176, 1996.

- Pauwels, R.; Balzarini, J.; Baba, M.; Snoeck, R.; Schols, D.; Herdewijn, P.; Desmyter, J.; De Clercq, E. Rapid and Automated Tetrazolium-Based Assay for the Detection of anti-HIV Compounds. *J. Virol. Methods*, 20, 309-321, 1988.
- Pauwels, R.; Andries, K.; Desmyter, J.; Schols, D.; Kukla, M. J.; Breslin, H. J.; Raeymaeckers, A.; Gelder, J. V.; Woestenborghs, R.; Heykants, J. Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives. *Nature*, 343, 470-474, 1990.
- Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D. HIV-1 Dynamics In Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science*, 271, 1582-1586, 1996.
- Petersen, L.; Jørgensen, P. T.; Nielsen, C.; Hansen, T. H.; Nielsen, J.; Pedersen, E. B. Synthesis and Evaluation of Double-Prodrugs against HIV. Conjugation of D4T with 6-Benzyl-1-(ethoxymethyl)-5-isopropyluracil (MKC-442, Emivirine) -Type Reverse Transcriptase Inhibitors via the SATE Prodrug Approach. *J. Med. Chem.*, 48, 1211-1220, 2005.
- Peterson, S. D.; Schaal, W.; Karle'n, A. Improved CoMFA Modeling by Optimization of Settings. *J. Chem. Inf. Model.*, 46, 355-364, 2006.
- Pita, S. S. R. Modelos de QSAR-4D dependente do receptor de inibidores peptídicos da tripanotona redutase. Dissertação (Mestrado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 94 p., 2006.
- Polanski, J.; Gieleciak, R.; Magdziarz, T.; Bak, A. GRID Formalism for the Comparative Molecular Surface Analysis: Application to the CoMFA Benchmark Steroids, Azo Dyes, and HEPT Derivatives. *J. Chem. Inf. Comput. Sci.*, 44, 1423-1435, 2004.

- Proudfoot, J. R.; Hargrave, K. D.; Kapadia, S. R.; Patel, U. R.; Grozinger, K. G.; Mcneil, D. W.; Cullen, E.; Cardozo, M.; Tong, L.; Kelly, T. A.; Rose, J.; David, E.; Mauldin, S. C.; Fuchs, V. U.; Vitous, J.; Hoermann, M.; Klunder, J. M.; Raghavan, P.; Skiles, J. W.; Mui, P.; Richman, D. D.; Sullivan, J. L.; Shih, C.; Grob, P. M.; Adams, J. Novel Non-Nucleoside Inhibitors of Human Immunodeficiency Virus Type 1 (HIV-1) Reverse Transcriptase. 4. 2-Substituted Dipyrindodiazepinones as Potent Inhibitors of Both Wild-Type and Cysteine-181 HIV-1 Reverse Transcriptase Enzyme. *J. Med. Chem.*, 38, 4830-4838, 1995.
- Raffanti, S.; Haas, D. W. Agentes Anti-Retrovirais. In Goodman & Gilman, As Bases Farmacológicas da Terapêutica. 11^a ed., 1671 pp., 2006.
- Ragno, R.; Mai, A.; Sbardella, G.; Artico, M.; Massa, S.; Musiu, C.; Mura, M.; Marturana, F.; Cadeddu, A.; La Colla, P. Computer-aided design, synthesis, and anti-HIV-1 activity in vitro of 2-alkylamino-6-[1-(2,6-difluorofenil)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as a novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *J. Med. Chem.*, 47, 928-934, 2004.
- Ragno, R.; Frasca, S.; Manetti, F.; Brizzi, A.; Massa, S. HIV-Reverse Transcriptase Inhibition: Inclusion of Ligand-Induced Fit by Cross-Docking Studies. *J. Med. Chem.*, 48, 200-212, 2005.
- Ranise, A.; Spallarossa, A.; Schenone, S.; Bruno, O.; Bondavalli, F.; Vargiu, L.; Marceddu, T.; Mura, M.; La Colla, P.; Pani, A. Design, Synthesis, SAR, and Molecular Modeling Studies of Acylthiocarbamates: A Novel Series of Potent Non-nucleoside HIV-1 Reverse Transcriptase Inhibitors Structurally Related to Phenethylthiazolylthiourea Derivatives. *J. Med. Chem.*, 46, 768-781, 2003.

- Rao, A.; Balzarini, J.; Carbone, A.; Chimirri, A.; De Clercq, E.; Monforte, A. M.; Monforte, P.; Pannecouque, C.; Zappalà, M. 2-(2,6-Dihalophenyl)-3-(pyrimidin-2-yl)-1,3-thiazolidin-4-ones as non-nucleoside HIV-1 reverse transcriptase inhibitors. *Antiviral Res.*, 63, 79-84, 2004.
- Reid, R. E. Gene Therapy. In *Foye's principles of medicinal chemistry*. Ed. Lemke, T. L. e Williams, D. A. Ed. Lippincott Williams & Wilkins. 5^a Ed., 1114 p., 2002.
- Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High resolution structure of HIV-1 RT: Insights from four RT-inhibitor complexes. *Nat. Struct. Biol.*, 2, 293-302, 1995.
- Ren J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure*, 3, 915-926, 1995.
- Ren, J.; Stammers, D. K. HIV reverse transcriptase structures: designing new inhibitors and understanding mechanisms of drug resistance. *Trends in Pharmacol. Sci.*, 26, 4-7, 2005.
- Ren, J.; Nichols, C. E.; Stamp, A.; Chamberlain, P. P.; Ferris, R.; Weaver, K. L.; Short, S. A.; Stammers, D. K. Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptases mutated at codons 101 or 138. *FEBS Journal*, 273, 3850-3860, 2006.
- Rizzo, R. C.; Wang, D.-P.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Sustiva through Computation of Resistance Profiles. *J. Am. Chem. Soc.*, 122, 12898-12900, 2000.

- Rizzo, R. C.; Udier-Blagovic, M.; Wang, D.-P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H., Jr.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of Activity for Nonnucleoside Inhibitors with HIV-1 Reverse Transcriptase Based on Monte Carlo Simulations *J. Med. Chem.*, 45, 2970-2987, 2002.
- Rodriguez-Barrios, F.; Balzarini, J.; Gago, F. The Molecular Basis of Resilience to the Effect of the Lys103Asn Mutation in Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors Studied by Targeted Molecular Dynamics Simulations. *J. Am. Chem. Soc.*, 127, 7570-7578, 2005.
- Rodgers, D. W.; Gamblin, S. J.; Harris, B. A.; Ray, S.; Culp, J. S. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 1222-1226, 1995.
- Rodrigues, C. R. Modelagem molecular de agentes antiprotozoários utilizando cálculos semi-empíricos AM1, mecânica molecular e métodos de QSAR (CoMFA/HQSAR). Tese de Doutorado, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 139 p., 1999.
- Rogers, D.; Hopfinger, A. J. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.*, 34, 854-866, 1994.
- Rogers, D. Some Theory and Examples of Genetic Function Approximation with Comparison to Evolutionary Techniques. *In Genetic Algorithms in Molecular Modeling*. Ed. James Devillers. Ed. Academic Press, San Diego. 327 p., 1996.
- Romeiro, N. C. Estudos de QSAR independente de dependente do receptor aplicados a inibidores da proteína quinase p38. Tese (Doutorado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 153 p., 2002.

- Romeiro, N. C.; Albuquerque, M. G.; de Alencastro, R. B.; Ravi, M.; Hopfinger, A. J. Free-energy force-field three-dimensional quantitative structure-activity relationship analysis of a set of p38-mitogen activated protein kinase inhibitors. *J. Mol. Model.*, 12, 855-68, 2006.
- Romero, D. L.; Morge, R. A.; Genin, M. J.; Biles, C.; Busso, M.; Resnick, L.; Althaus, I. W.; Reusser, F.; Thomas, R. C.; Tarpley, W. G. Bis(heteroaryl)piperazine (BHAP) reverse transcriptase inhibitors: structure-activity relationships of novel substituted indole analogues and the identification of 1-[(5-methanesulfonamido-1H-indol-2-yl)-carbonyl]-4-[3-[(1-methylethyl)amino]-pyridinyl]piperazine mono methanesulfonate (U-90152S), a second-generation clinical candidate. *J. Med. Chem.*, 36, 1505–1508, 1993.
- Romero, D. L.; Olmsted, R. A.; Poel, T. J.; Morge, R. A.; Biles, C.; Keiser, B. J.; Kopta, L. A.; Friis, J. M.; Hosley, J. D.; Stefanski, K. J.; Wishka, D. G.; Evans, D. B.; Morris, J.; Stehle, R. G.; Sharma, S. K.; Yagi, Y.; Voorman, R. L.; Adams, W. J.; Tarpley, W. G.; Thomas, R. C. Targeting delavirdine/atevirdine resistant HIV-1: identification of (alkylamino)piperidine-containing bis(heteroaryl)piperazines as broad spectrum HIV-1 reverse transcriptase inhibitors. *J. Med. Chem.*, 39, 3769-3789, 1996.
- Roy, P. P.; Leonard, J. T.; Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemom. Intell. Lab. Syst.*, 90, 31-42, 2008.
- Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S. Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. *J. Mol. Graph. Model.*, 26, 1-13, 2007.
- Russell, S; Norvig, P. Busca com Informação e Exploração. *In Inteligência Artificial*. Ed. Campus / Elsevier, 2ª Ed. 1021 pp., 2001.

- Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23, 327–341, 1977.
- Sant'Anna, C. M. R. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). *Quím. Nova*, 25, 505-512, 2002.
- Sarafianos, S. G.; Das, K.; Clark, A. D., Jr.; Ding, J.; Boyer, P. L.; et al. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with α -branched amino acids. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 10027-10032, 1999.
- Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.*, 20, 1449-1461, 2001.
- Schafer, W.; Friebe, W. G.; Leinert, H.; Mertens, A.; Poll, T.; von der Saal, W.; Zilch, H.; Nuber, B.; Ziegler, M. L. Non-nucleoside inhibitors of HIV-1 reverse transcriptase: molecular modeling and X-ray structure investigations. *J. Med. Chem.*, 36, 726–732, 1993.
- Schuettelkopf, A. W.; van Aalten, D. M. F. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D60*, 1355-1363, 2004. (<http://davapc1.bioch.dundee.ac.uk/prodrg/>)
- Scott, C.; Grover, D.; Nelson, M. Is there a role for etravirine in patients with Nonnucleoside reverse transcriptase inhibitor resistance? *AIDS*, 22, 989-990, 2008.
- Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A*, 103, 3596-3607, 1999.

- Selassie, C. D. History of Quantitative Structure-Activity Relationships. *In Burger's Medicinal Chemistry and Drug Discovery*. Ed. Abraham, D. J.; John Wiley & Sons, Inc., Vol. 1, 1-48, 2003.
- Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered Molecular Dynamics Simulation on the Binding of NNRTI to HIV-1 RT. *Biophys. J.*, 84, 3547-3563, 2003.
- Silvestri, R.; Artico, M.; De Martino, G.; Ragno, R.; Massa, S.; Loddò, R.; Murgioni, C.; Loi, A. G.; La Colla, P.; Pani, A. Synthesis, Biological Evaluation, and Binding Mode of Novel 1-[2-(Diarylmethoxy)ethyl]-2-methyl-5-nitroimidazoles Targeted at the HIV-1 Reverse Transcriptase. *J. Med. Chem.*, 45, 1567-1576, 2002.
- Sippl, W.; Holtje, H. D. Structure-based 3D-QSAR – merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods. *J. Mol. Struct. (Theochem)*, 503, 31-50, 2000.
- Sippl, W. Development of biologically active compounds by combining 3D QSAR and structure-based design methods. *J. Comput.-Aided Mol. Des.*, 16, 825-830, 2002.
- Smerdon, S. J.; Jager, J.; Wang, J.; Kohlstaedt, L. A.; Chirino, A. J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA*, 91, 3911-3915, 1994.
- Sobhia M.E.; Bharatam P. V. Comparative molecular similarity indices analysis (CoMSIA) studies of 1,2-naphthoquinone derivatives as PTP1B inhibitors. *Bioorg. & Med. Chem.*, 13, 2331-2338, 2005.
- Sodero, A. C. R. Modelos de QSAR-4D de uma série de análogos do Raloxifeno: um modulador seletivo do receptor de estrogênio., Tese (Mestrado em Química) – Instituto de Química, UFRJ, Rio de Janeiro. 112 p., 2007.

Spartan `06, Irvine, CA 92612 USA, Wavefunction, Inc.

Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. Linear Interaction Energy (LIE) Models for Ligand Binding in Implicit Solvent: Theory and Application to the Binding of NNRTIs to HIV-1 Reverse Transcriptase. *J. Chem. Theory Comput.*, 3, 256-77, 2007.

Sybyl 7.2, Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO 63144.

Szczech, G. M.; Furman, P.; Painter, G. R.; Barry, D. W.; Borroto-Esoda, K.; Grizzle, T. B.; Blum, M. R.; Sommadossi, J. P.; Endoh, R.; Niwa, T.; Yamamoto, M.; Moxham, C. Safety assessment, in vitro and in vivo, and pharmacokinetics of emivirine, a potent and selective nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, 44, 123-130, 2000.

Tanaka, H.; Baba, M.; Hayakawa, H.; Sakamaki, T.; Miyasaka, T.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Nitta, I.; Shigeta, S.; Walker, R. T.; Balzarini, J.; De Clercq, E. A new class of HIV-1-specific 6-substituted acyclouridine derivatives: synthesis and anti-HIV-1 activity of 5- or 6-substituted analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT). *J. Med. Chem.*, 34, 349-357, 1991.

Tang, K.; Li, T. Combining PLS with GA-GP for QSAR. *Chemom. Intel. Lab. Syst.*, 64, 55-64, 2002.

Thomas, G. The SAR and QSAR Approaches to Drug Design. *In Fundamentals of Medicinal Chemistry*, John Wiley & Sons, England, 410 p., 2003.

Todd, M. J.; Freire, E. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins*, 36, 147-156, 1999.

- Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors, Wiley-VCH, 423 p., 2000.
- Togninelli, A.; Carmi, C.; Petricci, E.; Mugnaini, C.; Massa, S.; Corelli, F.; Botta, M. Solution-phase parallel synthesis of S-DABO analogues. *Tetrahedron Lett.*, 47, 65–67, 2006.
- 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 Peptidomimetic Renin Inhibitors. *J. Chem. Inf. Comput. Sci.* 37, 792-811, 1997.
- Toropov, A. A.; Toropova, A. P.; Nesterov, I. V.; Nabiev, O. M. Comparison of QSAR models of anti-HIV-1 potencies based on labeled hydrogen filled graph and graph of atomic orbitals. *J. Mol. Struct.: THEOCHEM*, 640, 175-181, 2003.
- Tranter, D. Monitor: Molecules and Profiles. Antiviral Molecules. *Drug Disc. Today*, 6, 1238, 2001.
- Tronchet, J. M. J.; Grigorov, M.; Dolatshahi, N.; Moriaud, F.; Weber, J. A QSAR study confirming the heterogeneity of the HEPT derivative series regarding their interaction with HIV reverse transcriptase. *Eur. J. Med. Chem.*, 32, 279-299, 1997.
- Tropsha, A. Recent trends in Quantitative Structure-Activity Relationships. *In Burger's Medicinal Chemistry and Drug Discovery*, Abraham, D. J. Ed.; John Wiley & Sons, Inc., Vol. 1, 49-76, 2003.
- Udier-Blagovic, M.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Nucleoside Inhibitor TMC125. *J. Am. Chem. Soc.*, 125, 6016-6017, 2003.

- Udier-Blagovic, M.; Watkins, E. K.; Tirado-Rives, J.; Jorgensen, W. L. Activity predictions for efavirenz analogues with the K103N mutant of HIV reverse transcriptase. *Bioorg. Med. Chem. Lett.*, 13, 3337-3340, 2003.
- van de Waterbeemd H.; Rose, S. Quantitative Approaches to Structure-Activity Relationship. In *The Practice of Medicinal Chemistry*. Ed. Wermuth, C. G. Elsevier, 351-369, 782 p., 2003.
- van Gunsteren, W. F.; Berendsen, H. J. C. GROMOS87. Groningen Molecular Simulation Library Manual. BIOMOS, Groningen, The Netherlands, 1987.
- Vedani, A.; Huhta, D. W. A new force field for modeling metalloproteins. *J. Am. Chem. Soc.*, 112, 4759–4767, 1990.
- Vedani, A.; Huhta, D. W. An algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc.*, 113, 5860–5862, 1991.
- Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.*, 120, 4471–4477, 1998.
- Vedani, A.; McMasters, D.R.; Dobler, M. Genetische Algorithmen im 3D-QSAR: Verwendung multipler Wirkstofforientierungen zur verbesserten Voraussage der Toxizität. *ALTEX*, 16, 9-14, 1999.
- Vedani, A.; Briem, K.; Dobler, M.; Dollinger, H.; McMasters, D. R. Multiple-Conformation and Protonation-State Representation in 4D-QSAR: The Neurokinin-1 Receptor System. *J. Med. Chem.*, 43, 4416-4427, 2000.
- Vedani, A.; Dobler M. Internet laboratory for predicting harmful effects triggered by drugs and chemicals. *ALTEX*, 18, 110–114, 2001.

- Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.*, 45, 2139–2149, 2002.
- Vedani, A.; Dobler, M. Internet laboratory for predicting harmful effects triggered by drugs and chemicals-A progress report. *ALTEX*, 20, 85–91, 2003.
- Vedani, A.; Dollinger, H.; Hasselbach, K.-M.; Dobler, M.; Lill, M. A. Novel ligands for the chemokine receptor-3 (CCR3): A receptor modeling study based on 5D-QSAR. *J. Med. Chem.*, 48, 1515–1527, 2005a.
- Vedani, A.; Dobler, M.; Lill, M.A. Combining protein modeling and 6D-QSAR-simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.*, 48, 3700–3703, 2005b.
- Vedani, A.; Dobler, M.; Lill, M.A. In silico prediction of harmful effects triggered by drugs and chemicals. *Toxicol. Appl. Pharmacol.*, 207, 398–407, 2005c.
- Vedani, A.; Dobler, M.; Lill, M. A. The challenge of predicting drug toxicity in silico. *Pharmacology & Toxicology*, 99, 195–208, 2006.
- Vella, S.; Palmisano, L. Antiretroviral therapy: state of the HAART. *Antiviral Res.*, 45, 1-7, 2000.
- Venezia, C. F.; Howard, K. J.; Ignatov, M. E.; Holladay, L. A.; Barkley, M. D. Effects of Efavirenz Binding on the Subunit Equilibria of HIV-1 Reverse Transcriptase. *Biochemistry*, 45, 2779-2789, 2006.
- Venkatarangan, P.; Hopfinger, A. J. Prediction of Ligand-Receptor Binding Thermodynamics by Free Energy Force Field Three-Dimensional Quantitative Structure-Activity Relationship Analysis: Applications to a Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase. *J. Med. Chem.*, 42, 2169-2179, 1999.

- Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Janssen, P. A. J. Inhibition and substrate recognition: a computational approach applied to HIV protease. *J. Comput.-Aid. Mol. Des.*, 17, 567-581, 2003.
- Wang, Y.; Chen, F.-E.; Balzarini, J.; De Clercq, E.; Pannecouque, C. Non-nucleoside HIV-1 reverse-transcriptase inhibitors. Part 10. Synthesis and anti-HIV activity of 5-alkyl-6-(1-naphthylmethyl)pyrimidin-4(3H)-ones with a mono- or disubstituted 2-amino function as novel dihydro-alkoxy-benzyl-oxypyrimidine (DABO) analogues. *Chemistry & Biodiversity*, 5, 168-176, 2008.
- Weekes, D.; Fogel, G. B. Evolutionary optimization, backpropagation, and data preparation issues in QSAR modeling of HIV inhibition by HEPT derivatives. *Biosystems*, 72, 149-158, 2003.
- Wermuth, C. G. Medicinal Chemistry: Definition and Objectives, the three main phases of drug activity, drugs and disease classifications. In *The Practice of Medicinal Chemistry*. Editor Wermuth, C. G. Elsevier, 29-34, 782 p., 2003.
- WHO/UNAIDS Global Summary of the HIV/AIDS Epidemic. WHO Report, 2007. <http://www.who.int/hiv/pub/epidemiology/epi2007/en/>, acesso em 02 de setembro de 2008.
- Xu, Q.-S.; Daszykowski, M.; Walczak, B.; Daeyaert, F.; de Jonge, M. R.; Heeres, J.; Koymans, L. M. H.; Lewi, P. J.; Vinkers, H. M.; Janssen, P. A.; Massart, D. L. Multivariate adaptive regression splines—studies of HIV reverse transcriptase inhibitors. *Chem. Intell. Lab. Syst.*, 72, 27-34, 2004.
- Yadav, A.; Singh, S. K. Common binding mode for structurally and chemically diverse NNRTIs. *J. Mol. Struct: THEOCHEM*, 723, 205-209, 2005.

- Zheng, M.; Du, L.; Shen, J.; Luo, X.; Zhu, W.; Jiang, H. Towards discovering dual functional inhibitors against both wild type and K103N mutant HIV-1 reverse transcriptases: molecular docking and QSAR studies on 4,1-benzoxazepinone analogues. *J. Comput. Aided Mol. Des.*, 20, 281-289, 2006.
- Zhou, Z.; Madura, J. D. CoMFA 3D-QSAR Analysis of HIV-1 RT Nonnucleoside Inhibitors, TIBO Derivatives Based on Docking Conformation and Alignment. *J. Chem. Inf. Comput. Sci.*, 44, 2167-2178, 2004.
- Zhou, Z.; Madrid, M.; Evanseck, J. D.; Madura, J. D. Effect of a Bound Non-Nucleoside RT Inhibitor on the Dynamics of Wild-Type and Mutant HIV-1 Reverse Transcriptase. *J. Am. Chem. Soc.*, 127, 17253-17260, 2005.

8. ANEXOS

Tabela A. Matriz de correlação cruzada entre os descritores das melhores equações dos **BD-I** a **-IV**.

	Ile94LJ	Pro97LJ	Gly99	Gly99C	Lys101LJ	Lys103LJ	Val179LJ	Tyr181LJ	Gln182C	Tyr183	Leu187	Leu187C	Tyr188	Tyr188LJ	Tyr188C
Ile94LJ	1,000														
Pro97LJ	-0,129	1,000													
Gly99	0,114	0,050	1,000												
Gly99C	0,108	0,056	0,994	1,000											
Lys101LJ	0,105	-0,142	0,255	0,254	1,000										
Lys103LJ	0,215	-0,109	-0,286	-0,291	0,200	1,000									
Val179LJ	-0,028	-0,105	-0,024	-0,020	-0,338	-0,143	1,000								
Tyr181LJ	0,406	0,049	-0,306	-0,310	-0,166	0,369	-0,039	1,000							
Gln182C	0,280	-0,078	-0,054	-0,053	0,108	0,056	0,114	-0,123	1,000						
Tyr183	-0,209	-0,024	0,026	0,027	0,114	-0,087	-0,057	-0,535	0,522	1,000					
Leu187	-0,085	-0,057	-0,355	-0,355	-0,134	-0,048	0,097	-0,381	0,239	0,497	1,000				
Leu187C	-0,085	-0,057	-0,358	-0,359	-0,138	-0,048	0,099	-0,375	0,227	0,482	1,000	1,000			
Tyr188	0,143	-0,040	-0,090	-0,092	0,137	0,305	0,183	0,325	0,071	-0,309	-0,140	-0,136	1,000		
Tyr188LJ	0,177	0,237	-0,137	-0,140	-0,234	0,138	0,200	0,622	-0,371	-0,706	-0,431	-0,421	0,467	1,000	
Tyr188C	0,049	-0,194	-0,015	-0,016	0,301	0,257	0,081	-0,025	0,312	0,095	0,113	0,111	0,832	-0,102	1,000
Ser191	0,058	0,002	-0,148	-0,152	0,314	0,688	-0,013	0,261	0,039	0,004	-0,097	-0,099	0,439	0,180	0,381
Ser191C	0,060	0,005	-0,151	-0,155	0,316	0,694	-0,021	0,264	0,042	0,006	-0,099	-0,100	0,439	0,179	0,382
Glu224	0,327	-0,015	-0,252	-0,254	0,380	0,286	-0,189	0,154	0,197	0,176	0,076	0,072	0,198	-0,145	0,313
Pro225C	0,279	-0,035	-0,315	-0,317	0,330	0,300	-0,190	0,204	0,154	0,139	0,084	0,081	0,198	-0,117	0,296
Pro226C	0,172	-0,072	-0,399	-0,402	0,234	0,284	-0,244	0,260	0,077	0,116	0,157	0,155	0,188	-0,072	0,257
Phe227	0,169	-0,197	-0,446	-0,451	0,208	0,457	-0,256	0,292	0,133	0,084	0,354	0,355	0,267	-0,058	0,336
Phe227LJ	0,244	-0,199	-0,199	-0,204	0,224	0,546	-0,177	0,401	0,234	0,112	0,036	0,032	0,196	0,027	0,203
Trp229	-0,325	0,160	0,179	0,184	-0,177	-0,381	0,221	-0,138	0,018	0,187	-0,002	-0,003	0,012	-0,140	0,101
His235LJ	0,059	0,120	0,041	0,043	-0,121	-0,076	-0,067	-0,015	-0,194	-0,172	0,088	0,094	-0,190	0,004	-0,216
Asp237	0,086	0,131	0,312	0,313	0,190	0,114	-0,372	-0,098	0,104	0,170	-0,145	-0,149	-0,127	-0,101	-0,079
Asn136LJ	0,350	-0,004	0,187	0,186	-0,123	-0,350	0,244	-0,073	0,061	-0,249	0,013	0,018	0,062	0,017	0,060
Asn137LJ	0,489	-0,002	-0,101	-0,100	-0,141	-0,166	0,229	0,061	0,139	-0,174	0,084	0,088	0,028	-0,011	0,038
Glu138C	-0,169	-0,087	0,511	0,513	0,097	-0,468	0,211	-0,226	0,047	0,077	-0,327	-0,330	0,184	0,027	0,190
Thr139	-0,559	0,066	0,138	0,137	0,143	-0,078	0,012	-0,440	-0,193	0,180	0,005	0,003	-0,049	-0,041	-0,029

Tabela A. Matriz de correlação cruzada entre os descritores das melhores equações dos **BD-I** a **-IV** (Continuação).

	Ser191	Ser191C	Glu224	Pro225C	Pro226C	Phe227	Phe227LJ	Trp229	His235LJ	Asp237	Asn136LJ	Asn137LJ	Glu138C	Thr139
Ser191	1,000													
Ser191C	1,000	1,000												
Glu224	0,192	0,195	1,000											
Pro225C	0,254	0,257	0,955	1,000										
Pro226C	0,291	0,294	0,807	0,922	1,000									
Phe227	0,344	0,347	0,583	0,658	0,748	1,000								
Phe227LJ	0,377	0,382	0,400	0,405	0,354	0,668	1,000							
Trp229	-0,165	-0,167	-0,224	-0,247	-0,223	-0,362	-0,322	1,000						
His235LJ	-0,379	-0,376	-0,047	-0,101	-0,233	-0,155	-0,051	0,025	1,000					
Asp237	-0,046	-0,041	0,051	0,014	-0,021	0,066	0,164	-0,205	0,208	1,000				
Asn136LJ	-0,263	-0,265	-0,134	-0,156	-0,233	-0,242	-0,398	0,345	0,177	-0,258	1,000			
Asn137LJ	-0,134	-0,134	0,019	0,008	-0,041	-0,103	-0,242	0,280	0,131	-0,295	0,830	1,000		
Glu138C	-0,028	-0,036	-0,159	-0,165	-0,177	-0,345	-0,225	0,244	-0,224	0,069	-0,040	-0,219	1,000	
Thr139	0,051	0,049	-0,082	-0,096	-0,154	-0,098	-0,022	-0,200	-0,084	0,050	-0,441	-0,687	0,319	1,000

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)